University of Strathclyde Glasgow

Multi-Fidelity Methods for Multi-Objective Optimal Control applied to Aerospace Systems

Doctoral Dissertation of:

Benjamin C D Parsonage

Aerospace Centre of Excellence
Department of Mechanical & Aerospace Engineering

©2025

Declaration

This thesis is submitted to the University of Strathclyde in partial fulfilment of the requirements for the degree of Doctorate of Philosophy in Aerospace Engineering.

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed: Ben Parsonage

Date: 22nd February 2026

Supervisor: Dr Christie Maddock

Second supervisor: Prof Massimiliano Vasile

Aerospace Centre of Excellence
Department of Mechanical & Aerospace Engineering
University of Strathclyde
Glasgow, Scotland, United Kingdom

Abstract

Continuous, super-linear increase in the overall complexity of modern Multidisciplinary Design Optimisation problems inexorably tempers the benefits of advancing computational power. The pursuit of computational efficiency therefore remains of crucial importance, including within all modelling and analysis algorithms employed throughout.

This work presents two methodologies for reducing the computational burden associated with solving complex Multidisciplinary Design Optimisation problems, with a particular focus on Multi-Objective Optimal Control Problems. The first involves adaptively combining separate levels of modelling fidelity into one single approximation. This approach is particularly suited to cases where one or more subsystems can be represented using alternative numerical models/datasets of competing accuracy and/or computational cost. The second involves systematically upgrading an existing level of modelling fidelity defined by the mathematical formulation of the problem itself. This approach is particularly suited to problems where solution fidelity is more dependant on the number of variables (both static and dynamic) used to describe the system/subsystem(s) of interest.

Each methodology is incorporated into separate numerical frameworks intended for application in scenarios where in-depth and exclusively high-fidelity models/analyses are not viable. Several isolated numerical tests are examined in each case to systematically validate the fidelity management steps employed within each framework. Finally, each framework is demonstrated in application to a complex engineering design scenario.

It is found that the proposed approaches offer viable and efficient methods with which to adaptively upgrade modelling fidelity for complex Multidisciplinary Design Optimisation problems. These methods can be formulated in a general manner, without relying on extensive *a priori* information, particularly suiting their application to early-stage preliminary design problems.

With this class of approach, the design space of complex systems can be more thoroughly explored in terms of dimensionality, modelling fidelity and multi-objective performance criteria, at earlier stages in the design process. This helps to quickly identify promising search directions for future development whilst reducing the risk of significant later stage design alterations due to misleading or poorly-modelled physical behaviour/inter-disciplinary interactions.

Executive Summary

The overall aim of this thesis is to contribute techniques for improving the numerical efficiency of multidisciplinary/multi-objective design optimisation. In essence, aiming to increase the attainable accuracy/reliability of results whilst simultaneously reducing the associated computational burden of doing so. Being particularly relevant within the field of aerospace engineering, an emphasis is placed on *Multi-Objective Optimal Control Problems* (MOCPs) and analyses thereof.

The following chapters focus on the exploration and development of two distinct, but fundamentally linked, methodologies enabling the incorporation of multi-fidelity information sources into an *Multidisciplinary Design Optimisation* (MDO)-based framework. This allows the systematic and adaptive upgrading of system/subsystem modelling fidelity for the purpose of solving *Multi-Objective Optimal Control Problems* (MOCPs) in a computationally efficient manner. These methodologies operate according to the following respective strategies according to the primary source of multi-fidelity information utilised therein:

- An **external** strategy, where alternative numerical models of the same subsystem(s), competing in accuracy and computational cost, can be simultaneously incorporated into one single approximation.
- An **internal** strategy, where the numerical complexity of the problem itself (defined by, for example, the mathematical formulation or dimensionality) can be systematically upgraded, concentrating computational resource where it is most needed.

Each is incorporated into a separate numerical framework intended for application in scenarios where in-depth and exclusively high-fidelity models/analyses are not viable. For example, the preliminary design and optimisation phases of complex multidisciplinary engineering systems. This thesis is accordingly structured into two primary parts, addressing each framework in turn.

Part 1 is dedicated to the framework for **external** sources of multi-fidelity information. This framework assumes that any particular subsystem of a complex engineering problem is likely able to be represented by one of several alternative numerical models/formulations. It is assumed that such models conform to an inherent hierarchical trade-off between cost of

evaluation and the relative accuracy to which the underlying physical process is represented. That is, the more accurate the simulation, the more computationally expensive it will be. This framework is therefore primarily concerned with exploiting, in the most effective and efficient manner possible, all available data sources such that the highest numerical accuracy can be achieved at a minimum cost.

Following a review of the relevant literature presented in Chapter 2, Chapter 3 introduces a complete *Multi-Fidelity Model Management* (MFMM) framework in which computationally inexpensive surrogate models of selected subsystems can be autonomously generated and progressively corrected using high-fidelity data. Several data correction stages are formulated and examined. Each is integrated into a modular outer-level structure (Section 3.1.2). Correction stages are designed to operate *cooperatively*, leveraging a particular aspect of the correlation(s) between the available data. As such, each is tested both in isolation and in various sequential combinations using a suite of standard numerical testing functions (Section 3.1.4).

The overall framework is designed with a focus on autonomy, with a level of flexibility appropriate for application to general problems. For instance, any number of competing/alternative models may be specified for a particular subsystem. These models may have non-equal (or overlapping) input domains, and in the case of dynamic systems, may be time-phase specific or shared across multiple phases. In addition to (or instead of) numerical models, pre-existing data may be supplied to supplement the surrogate generation procedure. The initial sample points at which models are to be evaluated may be supplied, or simply determined adaptively. Surrogates are constructed using an *Artificial Neural Network* (ANN)-based modelling procedure (Section 3.1.3), where an autonomous hyper-parameter selection and training routine allows the approximation of any non-linear relationship to a specifiable degree of accuracy. This is achieved without requiring any *a priori* knowledge of the underlying relationship.

The complete MFMM framework is demonstrated in Chapter 4 with an engineering test case representing the preliminary design of an atmospheric sample re-entry vehicle, where multiple levels of fidelity are used to represent the vehicle aerodynamics. It is found that the sequential application of corrective stages allows a greater attainable approximation accuracy relative to a single-fidelity equivalent model for a set number of high-fidelity samples up to some critical sample size, which matches expectation.

Part 2 is dedicated to the framework for **internal** sources of multi-fidelity information. This approach systematically refines the numerical complexity at which the problem solution is represented. This allows computational resource to be more appropriately allocated as part of an adaptive process. For example, within areas that experience high rates of change, or are particularly dominant with respect to the final solution. This approach is designed specifically for direct *Multi-Objective Optimal Control Problems* (MOCPs), where time-dependant state and control profiles are evaluated by successively subdividing the time domain into a distinct grid of evaluation nodes, across which approximating polynomial relationships may be fit.

Chapter 5 presents a review of the relevant literature, with Chapter 6 introducing and

describing the proposed framework. The focus is on adaptively increasing the detail (or, the *resolution*) at which a set of direct Optimal Control solutions may be evaluated. This is achieved by concentrating discretisation nodes/grid-points around areas of high complexity unique to each solution (Section 6.2). Importantly, this means that a population of independent solutions can be processed concurrently, allowing each to represent a particular balance between multiple performance objectives.

A novel aspect of this approach is the simultaneous handling of multiple such unique grid structures within an evolutionary architecture (Section 6.4). This allows individual solutions to independently evolve concurrent with their propagation through objective space, aided via collaborative (social) action with neighbouring solutions. This is enabled by employing a new *social operator*, designed to facilitate such actions between variable-complexity solutions. A complete MOCP may thus be treated as an *open-ended* evolutionary design problem, ultimately allowing a more complete, higher-fidelity representation of the trade-off surface in a more numerically efficient manner (Section 6.5).

The new social operator is tested against several baseline approaches (Section 6.6). A complete *multi-resolution* framework, incorporating the above methods, is then applied to several demonstrative test cases (Section 6.7). The computational cost, achievable solution accuracy and population spread/diversity are measured in each case and compared against a traditional uniform-grid (single-fidelity) approach. An order of magnitude reduction in computational cost required to reach a desired level of accuracy is observed for all cases. Furthermore, the progressive nature of solution refinement (i.e. moving from *coarse* to *fine*, corresponding to a move from global exploration to local exploitation) is shown to facilitate a more complete and diverse representation of the multi-objective trade-off surface. The new social operator is found to successfully promote offspring diversity by allowing a progressively finer level of social action to occur between similar solutions.

Chapter 7 presents a more in-depth engineering test case aimed at demonstrating a practical application of the proposed multi-resolution framework. In this example, the preliminary design and assessment of a high-power interplanetary transportation system using nuclear propulsion is considered. Continuous control profiles are employed to represent the operation of the propulsion system, with the new framework shown to successfully identify and refine a set of optimal throttle-switching structures for the complete outbound/inbound mission with respect to multiple, competing objectives.

It is clear from this work that the potential for introducing higher levels of autonomy into the design and optimisation process of complex engineering systems remains strong, particularly with regards to multi-fidelity information. With respect to both *external* and *internal* information sources, this work has shown that the collaborative incorporation of multiple levels of simulation fidelity represents a significant quantitative improvement over single-fidelity approaches, particularly considering cases (such as early stage design/performance analyses) where extensive high-fidelity simulation is impractical.

Acknowledgements

The research work undertaken during the completion of this thesis was partially funded by the EPSRC Doctoral Training Partnership through the University of Strathclyde.

This research has made extensive use of the ARCHIE-WeSt High Performance Computer (www.archie-west.ac.uk) for the collection and recording of the data presented herein.

The work presented in Chapter 7 was partially funded through a grant by the European Space Agency General Support Technology Programme (ESA 3-16642/20/NL/MG/rk).

Extensive use was made of the MODHOC, DFET and MACS algorithms, developed at the University of Strathclyde and freely available in the SMART repository for Optimisation and Optimal Control (https://github.com/strath-ace/smart-o2c).

List of Publications

Parts of the work presented in this thesis have been published or presented in the following:

Journals

1. B. Parsonage and C. Maddock. A multi-fidelity model management framework for multi-objective aerospace design optimisation. *Frontiers in Aerospace Engineering*, 2 (February):1–21, 2023. doi: 10.3389/fpace.2023.1046177

2. B. Parsonage and C. Maddock. Solving multi-objective optimal control problems using a multiresolution approach. *Journal of Guidance, Control, and Dynamics*, pages 1–14, 2024b. ISSN 1533-3884. doi: 10.2514/1.g008223

Peer-reviewed conference papers and presentations

3. B. Parsonage and C. A. Maddock. Multi-stage Multi-fidelity Information Correction for Artificial Neural Network Based Meta-modelling. *2020 IEEE Symposium Series on Computational Intelligence, SSCI 2020*, pages 950–957, 2020b. doi: 10.1109/SSCI47803.2020.9308255

4. B. Parsonage and C. Maddock. Biased Dyadic Crossover for Variable-Length Multi-Objective Optimal Control Problems. *2024 IEEE Congress on Evolutionary Computation, CEC 2024 - Proceedings*, pages 1–9, 2024a. doi: 10.1109/CEC60901.2024.10611888

5. L. Cao, M. Wang, B. Parsonage, C. Maddock, and G. Dai. Multi-Agent Collaborative Search with Adaptive Heuristics and Weight Vectors for Aerospace Multi-objective Optimal Control Problems. *2024 IEEE Congress on Evolutionary Computation, CEC 2024 - Proceedings*, pages 1–8, 2024. doi: 10.1109/CEC60901.2024.10612203

Conference papers and presentations

6. B. Parsonage and C. Maddock. A multi-fidelity model management framework for multi-objective aerospace design optimisation. In *International Conference on Uncertainty Quantification and Optimisation*, 2020a

7. C. Maddock, B. Parsonage, L. Ricciardi, M. Vasile, and O. Cohen. Design optimisation and analysis of very high power transportation system to Mars. *Proceedings of the International Astronautical Congress, IAC*, A5(October):25–29, 2021. ISSN 00741795

8. B. Parsonage and C. Maddock. A Multi-Fidelity Model Management Framework for Multi-Objective Aerospace Design Optimization. In *ECCOMAS CM3-Transport Workshop 2021: Methods and Tools for Aeronautics Design*, 2021

9. B. Parsonage and C. Maddock. Multi-Objective Design Optimisation and Analysis of a Crewed Earth-Mars Transportation System using Nuclear Thermal Propulsion. *Proceedings of the International Astronautical Congress, IAC*, (October), 2024c

Dedication

My most profound thanks go to my supervisor, Dr Christie Maddock, whose patience and understanding are matched only by her invaluable guidance and exceptional expertise.

Thank you to my colleagues across the Department of Mechanical and Aerospace Engineering, in particular Dr Lorenzo Ricciardi for his advice, technical support and infectious enthusiasm during the earlier days of my PhD.

To my friends, I give my most heartfelt thanks for their (many) years of encouragement and support. You have been a continuing source of strength through waters gentle and fierce.

To my family, who have endured all that fate could throw without ever wavering in their love, thank you. I could not have completed this without you.

A final, special mention to those close to my heart who left us too soon, but whose love and support have been felt no less strongly each day since.

Contents

Abstract ii

Executive Summary iii

Acknowledgements vi

List of Publications vii

List of Acronyms xiv

1 Introduction 1

1.1 Motivation . . . 1

1.1.1 Numerical Efficiency: Optimisation . . . 5

1.1.2 Numerical Efficiency: Approximation . . . 7

1.1.3 Multi-Fidelity Modelling . . . 8

1.2 Research Objectives . . . 10

1.3 Research Methodology . . . 11

1.4 Structure . . . 12

1.5 Contributions . . . 13

I Model-based Fidelity Management 15

2 Multi-Fidelity Surrogate Modelling and Optimisation 16

2.1 Surrogate Modelling . . . 17

2.1.1 Polynomials . . . 18

2.1.2 Radial Basis Functions . . . 20

2.1.3 Artificial Neural Networks . . . 21

2.1.4 Kriging . . . 27

2.1.5 Support Vector Regression . . . 29

2.1.6 Surrogate Validation . . . 30

2.2 Infill Criteria . . . 33

2.2.1 Local Exploitation . . . 38
2.2.2 Global Exploration . . . 40
2.2.3 Balanced Exploitation/Exploration . . . 41
2.2.4 Parallel Infill . . . 50
2.2.5 Multi-Response Sampling . . . 51
2.2.6 Multi-Objective Sampling . . . 56
2.3 Multi-Fidelity Model Management . . . 59
2.3.1 Low-Fidelity Modelling . . . 64
2.3.2 Adaptation . . . 66
2.3.3 Fusion . . . 75
2.3.4 Filtering . . . 80
2.4 Multi-Fidelity Surrogate-Based-Optimisation . . . 81
2.4.1 Classifications and Terminology . . . 84
2.4.2 Global Methods . . . 86
2.4.3 Local Methods . . . 89
2.4.4 Stopping Criteria . . . 97
2.5 Conclusion . . . 98
2.6 Bibliography . . . 100

3 Adaptive Multi-Fidelity Model Management 126
3.1 Multi-Stage Adaptive Correction . . . 128
3.1.1 Formulation . . . 134
3.1.2 Correction Methods . . . 134
3.1.3 Automated Surrogate Modelling using Artificial Neural Networks . . . 138
3.1.4 Numerical Tests . . . 143
3.2 Multi-Fidelity Model Management Framework for Outer-Loop Applications . . . 174
3.2.1 Initial Surrogate Generation . . . 174
3.2.2 Adaptive Sampling . . . 177
3.3 Chapter Summary . . . 178
3.4 Bibliography . . . 183

4 Engineering Test Case: Transatmospheric Vehicle and Mission Design 193
4.1 Multi-Fidelity Modelling . . . 195
4.1.1 Geometry Parametrisation . . . 195
4.1.2 Aerodynamics . . . 196
4.1.3 2D/3D Geometry Meshing . . . 198
4.2 Multi-Fidelity Data Correction . . . 199
4.3 Optimisation . . . 202
4.3.1 System Models . . . 203

4.3.2 Problem Formulation . . . 205
4.3.3 Numerical Settings . . . 207
4.4 Results . . . 208
4.4.1 High/Low-Fidelity Solutions . . . 208
4.4.2 Multi-Fidelity Solutions . . . 210
4.4.3 K-fold Cross-Validation Error . . . 213
4.5 Cost . . . 214
4.6 Conclusion . . . 218
4.7 Bibliography . . . 220

II Resolution-based Fidelity Management 222

5 Background 223
5.1 Optimal Control . . . 224
5.1.1 Indirect Methods . . . 225
5.1.2 Direct Methods . . . 227
5.2 Adaptive Mesh Refinement for Optimal Control . . . 233
5.3 Multi-Objective Optimisation . . . 236
5.3.1 Scalarisation-based . . . 237
5.3.2 Dominance-based . . . 238
5.3.3 Others . . . 240
5.4 Multi-Objective Optimal Control . . . 240
5.5 Adaptive Mesh Refinement for Multi-Objective Optimal Control . . . 242
5.6 Conclusion . . . 244
5.7 Bibliography . . . 246

6 Multi-Objective Optimal Control with Multi-Resolution Grids 255
6.1 Direct Finite Elements in Time . . . 257
6.2 Dyadic Mesh Refinement . . . 261
6.2.1 Vectorised Bézier Curve Refinement . . . 263
6.3 Variable-Length Evolutionary Algorithms . . . 265
6.4 Multi-Resolution Memetics . . . 268
6.4.1 One-Point Crossover . . . 269
6.4.2 Two-Point Crossover . . . 270
6.4.3 Dyadic Crossover . . . 270
6.4.4 Variable-Length Differential Evolution . . . 271
6.5 The VL-MOCP Algorithm . . . 272
6.6 Computational Experiments: Variable-Length Crossover . . . 273
6.6.1 Chemical Reaction . . . 274

6.6.2 Moon Lander . . . 280
6.6.3 Discussion . . . 286
6.7 Computational Experiments: Grid Resolution . . . 287
6.7.1 Drug Displacement . . . 288
6.7.2 Apollo Re-entry . . . 292
6.7.3 Betts Shuttle Re-entry . . . 298
6.7.4 Discussion . . . 302
6.8 Chapter Summary . . . 303
6.9 Bibliography . . . 309

7 Engineering Test Case: Interplanetary Transportation 314
7.1 Background . . . 316
7.1.1 Propulsion . . . 316
7.1.2 Mission Architectures . . . 317
7.2 System Models . . . 318
7.2.1 Vehicle . . . 318
7.2.2 Dynamics . . . 324
7.3 Approach . . . 325
7.3.1 Optimisation Algorithm . . . 326
7.4 Results . . . 327
7.4.1 Earth-Mars Transfer . . . 327
7.4.2 Earth-Mars-Earth Transfer . . . 329
7.5 Conclusion . . . 332
7.6 Bibliography . . . 334

8 Conclusions 337
8.1 Recommendations for Future Work . . . 341

Full List of References 343

List of Acronyms

1PCO 1-Point Crossover 273

2PCO 2-Point Crossover 273

ACAP Accumulated Approximation 60

AMSE Adjusted Mean Square Error 32

AHD Averaged Hausdorff Distance 274

AIC Akaikes Information Criterion 32

AMMO Approximation Model Management Optimisation 90

ANN Artificial Neural Network iv

ANOVA Analysis of Variance 274

ASM Aggressive Space Mapping 72

AWS Adaptive Weighted Sum 239

BIC Bayesian Information Criterion 32

BIMFO Bayesian-Inspired Multi-Fidelity Optimisation 69

BK Blind Kriging 29

CCD Central Composite Design 146

CFD Computational Fluid Dynamics 4

CONLIN CONvex LINearization 60

CP Chemical Propulsion 327

CSSO Concurrent SubSpace Optimization 60

CV Cross-Validation 32

CVT Centroidal Voronoi Tessellation 65

DAE Differential Algebraic system of Equations 226

DARPA Defense Advanced Research Projects Agency

DCO Dyadic Crossover 273

DE Differential Evolution 13

DFET Direct Finite Elements in Time 13

DNS Direct Numerical Simulations 65

DoE Design of Experiment 17

DRA Design Reference Architecture 317

EA Evolutionary Algorithm 239

ECEF Earth Centred Earth Fixed 205

ECLSS Environmental Control and Life Support Systems

EGO Efficient Global Optimisation 86

EI Expected Improvement 33

ESA European Space Agency

EVA Extravehicular Activity

FEA Finite Element Analysis 4

FET Finite Elements in Time 257

FFD Full Factorial Design 146

FOSTRAD Free Open Source Tool for Re-Entry of Asteroids and Debris 197

FVM Finite Volume Method 197

GA Genetic Algorithm 6

GCK Generalised Hierarchical co-Kriging 130

GCR Galactic Cosmic Radiation 318

GEK Gradient-Enhanced Kriging 29

GHBF Generalised Hybrid Bridge Function 68

GIR Grid Intersect Ratio 279

GMSELOO Generalised Mean Square Leave-One-Out Error 77

GP Gaussian Process 43

HEO High Earth Orbit 318

HSCT High Speed Civil Transport 60

IGD Inverted Generational Distance 274

IMSE Integrated Mean Square Error 32

ISS International Space Station

kCVE K-Fold Cross Validation Error 150

KKT Karush-Kuhn-Tucker 231

KBNN Knowledge-Based Neural Network 22

LCB Lower Confidence Bounding 33

LCH4 Liquid Methane

LCSS Longest Common Sub-Sequence 267

LEO Low Earth Orbit 194

LES Large Eddy Simulations 65

LG Legendre-Gauss 231

LGR Legendre-Gauss-Radau 231

LH2 Liquid Hydrogen

LHD Latin Hypercube Design 37

LHS Latin Hypercube Sampling 36

LMO Low Mars Orbit 326

LO2 Liquid Oxygen

LOO Leave-One-Out 32

LSI Local Surface Inclination 197

MACS Multi-Agent Collaborative Search 14

MAE Maximum Absolute Error 32

MCMC Markov-Chain Monte-Carlo 80

ME Maximum Entropy 43

MDO Multidisciplinary Design Optimisation iii

mGA Messy-GA 267

ML Machine Learning 140

MLE Maximum Likelihood Estimator 28

MFM Multi-Fidelity Model 54

MFMM Multi-Fidelity Model Management iv

MLP Multi-Layer Perceptron 23

MM Manifold Mapping 94

MMA Method of Moving Asymptotes 60

MOCP Multi-Objective Optimal Control Problem iii

MODHOC Multi-Objective Direct Hybrid Optimal Control 207

MOGA Multi-Objective Genetic Algorithms 239

MOO Multi-Objective Optimisation 194

MOSA Multi-Objective Simulated Annealing 239

MSE Mean Square Error 24

MTGP Multi-Task Gaussian Process 53

NASA National Air and Space Association

NBI Normal Boundary Intersection 239

NCO No Crossover 273

NEP Nuclear Electric Propulsion 316

NERVA Nuclear Engine for Rocket Vehicle Application 316

NHLF Non-Hierarchical Low-Fidelity 130

NLP Non-Linear Programming 90

NP Nuclear Propulsion 317

NTER Nuclear Thermal Electric Rocket 317

NTR Nuclear Thermal Rocket 316

NTP Nuclear Thermal Propulsion 317

NUCO Near-Uniform Crossover 273

OC Optimal Control 10

ODE Ordinary Differential Equation 224

OK Ordinary Kriging 29

OSM Output Space-Mapping 94

OTV Orbital Test Vehicle 194

PDE Partial Differential Equation 23

PDF Probability Density Function 57

PI Probability of Improvement 33

PINN Physics-Informed Neural Network 23

POD Proper Orthogonal Decomposition 65

PRESS Prediction Error Sum of Squares 32

PSO Particle Swarm Optimisation 6

PVA Photovoltaic Array 318

QBC Query-By-Committee 43

RAE Relative Accuracy Error 32

RANS Reynolds Averaged Navier-Stokes 65

RBF Radial Basis Function 20

RBFNN Radial Basis Function Neural Network 23

RMAE Relative Maximum Absolute Error 148

RMSE Root Mean Square Error 32

RNG Random Number Generator 147

ROM Reduced Order Model 65

RRMSE Relative Root Mean Square Error 148

RSM Response Surface Model 18

SA Simulated Annealing 6

SAGA Species Adaptation Genetic Algorithm 267

SBO Surrogate Based Optimisation 7

SCF Spatial Correlation Function 28

SLP Sequential Linear Programming 60

SM Space Mapping 71

SMF Surrogate Management Framework 87

SOI Sphere of Influence 324

SPE Solar Particle Event 318

SPRP Shape Preserving Response Prediction 72

SQP Sequential Quadratic Programming 207

SSM Searching the Surrogate Model 33

SVLC Synapsing Variable-Length Crossover 267

SVR Support Vector Regression 29

SU2 Stanford University Unstructured Code 197

SUMO MATLAB Surrogate Modelling 50

TMS Thermal Management System

TRL Technology Readiness Level

TRMM Trust Region Model Management 84

UK Universal Kriging 29

UKSA UK Space Agency

UQ Uncertainty Quantification 341

USV Unmanned Space Vehicle 193

VCRSM Variable Complexity Response Surface Modelling 60

VLC Variable-Length Chromosome 269

VLC-GA Variable-Length-Chromosome Genetic Algorithms 268

VV Virtual Virus 267

WRMSE Weighted Root Mean Square Error 32

Introduction 1

1.1 Motivation

Development is driven by a demand for increased performance. Within the context of engineering systems design, performance can be measured by cost, reliability, efficiency, or simply the operational envelope of the primary system state(s) (Rao, 2009). In most cases, the role of the designer broadly becomes the specification of the system parameters or operational decisions that deliver the best possible performance for a given objective or scenario, within such constraints as are relevant to the problem at hand.

A natural consequence of development, itself further facilitated by a commercial incentive to withstand global competition, is a parallel progression in complexity. *Complex* systems in this sense may be defined as comprising of numerous interconnected or interdependent subsystems, each representing one or more particular areas of expertise. In most applications, the primary function(s), or *objective(s)*, of a complex system cannot be expressed in explicit analytical form, as the coupled dependence on the associated design variables is similarly complex and/or implicit (Koziel and Yang, 2011).

The traditional approach to the design of such *multidisciplinary* systems follows a compartmentalised, sequential process structure. Each area of expertise, or discipline, is analysed and developed, either independently or based on the output of another, until each has been locally treated. Should the complete system subsequently fail to meet some predefined global performance requirements, the process is iteratively repeated until convergence to a feasible configuration is achieved.

Unfortunately, this approach is ultimately outpaced by systems development itself. Whilst in premise aiming to satisfy ever increasing performance specifications, high-end complex systems designed using this approach can altogether fail to converge upon an acceptable

solution even after numerous (often costly) iterations (Schlager, 1953). Historically, this resulted in the emergence of Systems Engineering (Schlager, 1953), an integrated approach that considered the complete system (at least during preliminary phases) as more than simply the sum of individual subsystems, but as a complete co-dependant entity.

An enabling technique with great potential to further enhance Systems Engineering was found within the field of mathematical optimisation. Postponing a more rigorous definition, *optimisation* entails determining the set of conditions such that a certain function of decision variables be maximised or minimised with respect to some relevant performance objective. In other words, finding the best result under given circumstances. Up until around the mid-20th century, mathematical optimisation had existed as a largely theoretical endeavour, originally facilitated by the contributions of Newton and Leibnitz to the development of differential calculus in the 1700s and the subsequent foundations of the *calculus of variations* laid by Bernoulli, Euler and Lagrange, amongst others (Martins, 2012; Rao, 2009). Whilst the calculus of variations deals with the minimisation of functionals in a general sense, the extension to constrained problems via the addition of unknown multipliers was introduced by (and subsequently named after) Lagrange. A further significant development was the first application of the steepest-descent method to solve unconstrained minimisation problems by Cauchy.

Whilst the utilisation of mathematical optimisation techniques to solve complex decision making problems can be first attributed to the practice of Operations Research (Mccloskey, 1987) developed during World War II, widespread practical application of numerical approaches only truly became feasible through the emergence of the digital computer. This facilitated huge advancements in applied and theoretical optimisation techniques throughout the latter half of the 20th century. For example, methods to solve constrained optimisation problems stemmed from the simplex method derived by Dantzig et al. (1955), alongside the work of Bellman (1967) on the principle of optimality for dynamic programming. Karush (1939); Kuhn and Tucker (1951) defined a set of necessary and sufficiency conditions, thereafter known as the Karush-Kuhn-Tucker conditions, for the optimal solution of programming problems.

Endeavouring to solve even more complex design problems, a rigorous amalgamation of Systems Engineering and applied numerical optimisation techniques developed during the 1970s and 80s became known as *Multidisciplinary Design Optimisation* (MDO) (Rao, 2009). Initial investigations came from Schmit (1960) and Haftka (1977), who extended their experience in structural optimisation to include other disciplines. Supported by the exponentially increasing capability of computational processing, the simultaneous manipulation of several disciplines and their associated variables (which themselves could be unique or shared cross-discipline) greatly increased the prospect of iterative convergence to a globally optimal design (Martins and Lambe, 2013). That is, one that represents some compromise between constituent

subsystems in order to deliver optimal performance with respect to the system as a whole. Naturally, one of the primary challenges with handling such interdisciplinary coupling is the associated increase in problem complexity. In most cases, increased complexity may simply reflect the additional dimensionality accumulating from the analysis and design variables of all disciplines. In other examples however, systems analyses may require costly non-linear methods, even if each discipline is only linearly modelled. Finally, though each individual discipline may only require optimisation of a single performance objective, a multidisciplinary system is much more likely to require handling of multiple (often competing) objectives, a significantly more complex task numerically. In essence then, MDO represents a far greater computational challenge than treating each discipline independently (Sobieszczanski-Sobieski and Haftka, 1997).

A further hindrance is that the usability of systems-based MDO, at least in a practical sense, is naturally dependant on the extent to which interdisciplinary interactions can themselves be represented. Should individual subsystems be represented in insufficient detail, overall system performance predictions may be consequentially misleading, failing to represent significant design behaviour or interdisciplinary coupling (Christensen and Willcox, 2012; Bryson et al., 2019; Robinson, 2012). Ultimately this may lead to sub-optimal designs and/or increased development costs as a result of later stage alteration (Allaire and Willcox, 2014). Indeed, the consequences of completing a formal MDO process with inaccurate or misleading system models may eradicate any developmental/economic savings gained by including such analyses in the first place (Allaire and Willcox, 2014). Of such importance is this aspect that computational modelling in this respect can be considered an emerging third paradigm in modern science, complementing theory and experiment (Koziel and Yang, 2011). Computational modelling, including mathematical modelling, data analysis, computer simulations, and optimisation, thus represents a significant portion of research activity within science, engineering and industry.

Largely resulting from this development of systems design methods, computational modelling then finds itself subject to the same increasing demand for accuracy, from both an economical and performance-based perspective. Arguably this observation is perhaps trivial, as expanding a particular system's operational envelope naturally implies a proportional advancement in the complexity of the physical processes involved. An appropriate example is that of the progression from subsonic aerodynamics to supersonic and hypersonic flow regimes, requiring super-linear increases in modelling complexity. A related complication arises from the fact that with higher quality subsystem modelling comes design spaces of higher intricacy. For example, exhibiting discontinuities or highly non-linear/multi-modal behaviour in state space and/or objective space. The latter is particularly relevant should the design problem itself comprise several competing objectives.

With increasing complexity comes increased computational cost. Many advanced simulations

methods in engineering, such as *Computational Fluid Dynamics* (CFD) and *Finite Element Analysis* (FEA), still typically require significant computational resource, even for single, isolated analyses. Furthermore, numerical simulations of physical processes are inherently noisy, posing additional searching and convergence difficulties for many optimisation algorithms, especially gradient-based. The overall costs of such methods generally prohibits their use in conjunction with the numerical optimisation algorithms typically employed within any modern simulation based design and analysis paradigm (including MDO), which may require tens, hundreds or even thousands of function evaluations per run (depending on the number of design variables).

In fields where practical simulation is infeasible or prohibitively expensive, or where empirical data is limited and of poor quality, numerical simulation remains vital. This creates a need for flexible and efficient optimisation methodologies that can yield satisfactory solutions even in the presence of analytically intractable (i.e., discontinuous, non-differentiable, or inherently noisy) objectives and/or limited computational resource (Koziel and Yang, 2011).

For modern, complex systems, we therefore have two fundamental aspects inherently reliant on available computational resource; **modelling** (system and subsystem) and **analysis** (including numerical optimisation) (Venkataraman and Haftka, 2004). The first is relevant in a general sense; the ability to simulate physical processes (or indeed any relevant subsystem/interdisciplinary coupling) to an increasingly high degree of *fidelity* (copying accuracy) facilitates designs of higher practical reliability. The second includes the implication that relevant, higher quality detail regarding interdisciplinary interaction should be factored in during earlier stages of the design process, lest the process itself be fundamentally compromised by early convergence to a sub-optimal (or at least only locally optimal) solution. Such a design may fail to reach it's original performance specifications without significant late stage (and therefore very expensive) alterations, if even at all.

We may then conjecture that the continuous, super-linear increase in the overall complexity of MDO problems inexorably tempers the benefits of advancing computational power (most commonly cited in relation to Moore's law (Moore and Others, 1965), see the works by Schaller (1997); Venkataraman and Haftka (2004); Simpson (2008)). Accordingly, the fundamental problem common to most, if not all, areas of engineering design and optimisation can be acutely summarised as a trade off between *cost* and *accuracy*. This is, however, hardly a new observation. Seemingly timeless practical boundaries on computational resource historically necessitated a methodology where modelling fidelity was upgraded at a rate proportional to high-level developmental progress. In a manner not dissimilar to the original sequential approach to systems design, cheaper and faster numerical/empirical methods were used to explore design options and select possible configurations. Further study could subsequently include more detailed analyses, whether discipline specific or multidisciplinary, until even-

tually converging on some acceptable design validated with high-fidelity data. This then by definition implies that more complicated (i.e., representable only through higher-fidelity methods) subsystem interactions, design behaviour or system-level synergy are *not* used to define the initial search directions at arguably the most critical/sensitive time in the design process. This then exacerbates the risk of convergence to a sub-optimal configuration requiring significant (if not infeasible) cost to remedy. Generally speaking then, it is of crucial importance to concentrate research efforts into improving the overall efficiency of systems design/optimisation, including all modelling/analysis algorithms employed throughout.

1.1.1 Numerical Efficiency: Optimisation

Where the objectives and/or subsystems of a particular design problem can be evaluated using explicit formulas or inexpensive functions/models, efforts to increase efficiency are naturally focused on the optimiser itself. The efficiency of the optimiser can often be largely determined by the choice of algorithm given a particular problem. There are many optimisation algorithms in the literature, though as dictated by the *No Free Lunch Theorem* (Wolpert and Macready, 1997), none are suited to all problems in a general sense. The choice of optimiser depends on the type of problem, the nature of the algorithm, the desired quality of the solution(s), available computational resource (including practical limits on execution time), availability of the algorithm implementation, and the expertise of the decision-makers (Yang, 2010).

For example, considering the nature of the optimiser, gradient-based algorithms (such as the classical steepest-descent and Gauss-Newton methods) are not suited to problems in which the objective function is discontinuous. In this case, gradient-free algorithms (such as the Nelder and Mead (1965) downhill simplex method) are more appealing. The type of problem itself may also determine the most suitable choice of algorithm. If the problem is highly non-linear and/or multi-modal, local search methods such as hill-climbing and downhill simplex are largely unsuitable, as they will deterministically converge only to local optima (though they are extremely efficient at doing so), without necessarily reaching the global optimum. For this, global optimisation methods (such as Genetic Algorithms (Goldberg, 1989; Holland, 1992)) are more practical, though often come with greater cost and/or decreased reliability. Where local searches are most often *deterministic*, i.e. will give a predictable, repeatable output for the same input, global searches usually (but not always) incorporate some element of *stochastic* process (i.e. randomness), potentially arriving at a different solution each time. This *exploratory* capability can be introduced via some dependant stochastic mechanism within the algorithm itself[1], such as the randomised mutation/recombination in Genetic Algorithms (specifically, mutation rate and crossover probability), or facilitated by a stochastic modification

[1]Such approaches are also termed *heuristic*, or *meta-heuristic* algorithms. Whilst heuristic in this sense means simply to learn via trial-and-error (Talbi, 2002; Yang, 2010), modern literature increasingly refers to any stochastic algorithm as heuristic, with meta-heuristic reserved for those operating at a higher level.

to an otherwise deterministic process, such as adding a random restart to the hill-climbing algorithm. In this manner, regions within the search space that may be missed by deterministic solvers are more likely to be discovered. These approaches are also examples of Hybrid (or mixed-type) algorithms. That is, algorithms applying some combination of established methods to improve performance or efficiency for specific cases. Genetic Algorithms, in addition to being hybridised with gradient-based local searches, can alternatively employ their genetic operators to modify some component(s) of another algorithm.

Where numerical efficiency is concerned, the choice of particular algorithm could be more practically dependant on the trade-off between solution quality and available computational budget. Whilst high quality solutions cannot typically be guaranteed *a priori*, the aim can still be to obtain *acceptable* solutions (not necessary optimal) within a reasonable and practical time allocation. Where time is the prevalent constraint, *greedy* methods can be employed, or even the aforementioned hill-climbing with an added random restart feature, for example. Where processing power is instead the leading concern, single-agent (or trajectory-based) algorithms that trace out a search path as the optimisation process iteratively advances can be employed, as opposed to multi-agent (population-based) algorithms that exploit the interaction between multiple alternate solutions to facilitate the global search. Examples of trajectory-based approaches include hill-climbing, where the initial and final solutions are linked by a piecewise path through objective space, and the meta-heuristic *Simulated Annealing* (SA) algorithm (Kirkpatrick et al., 1983). Examples of population-based algorithms include *Particle Swarm Optimisations* (PSOs) (Kennedy and Eberhart, 1995) as well as *Genetic Algorithms* (GAs) (Goldberg, 1989; Holland, 1992), both of which will be discussed further in later chapters. The availability of information contained within previous iterations can also be utilised. Most algorithms consider only the current iteration (or the current best solution) with all previous iterations discarded, rather than incorporated for future use within some explicit memory[2]. An example of active memory utilisation is the Tabu search (Glover and Laguna, 1998), in which lists of move-history and recently visited solutions are recorded such that they will not be repeated, thus simultaneously encouraging a more exploratory behaviour without wasting resource.

Even despite this wealth of alternative algorithms available to the designer, straightforward optimisation of a given objective function is not always practical, or even possible. Particularly if relying on computer simulations, the problem objectives and/or subsystem models may be expensive, noisy or non-differentiable/discontinuous, with a complex implicit dependence on the associated design variables (Koziel and Yang, 2011). This then defines a significant bottleneck for the entire optimisation process. Furthermore, the availability of implementation

[2]This should not be confused with simply recording the current state at each iteration (i.e. for no further practical use), as is done in many population-based methods.

packages and/or the expertise of the designer may introduce additional hindrance to determining the 'best' choice for a particular problem. As an example, Newton's method, hill-climbing, Nelder-Mead downhill simplex, trust-region methods (Conn et al., 2000) and interior-point methods (Nesterov and Nemirovskii, 1994) are all implemented in many commercially available software packages, a factor likely contributing to their overall popularity in application. As such, the role of the decision maker may be reduced to simply selecting the best solution, rather than the best algorithm. Even still, in practice the most desirable solution to a given problem may still remain elusive, even with the best possible algorithms and well-crafted implementations. This is because many complex design problems, particularly non-linear optimisation problems, are *NP-hard*, with no efficient (in the polynomial sense) methodology definitively applicable.

1.1.2 Numerical Efficiency: Approximation

To reduce the dependence on the choice of optimiser, an alternative approach is to instead focus on improving efficiency within the problem itself. A traditional approach is to incorporate *approximations* into the design process. If one assumes that such approximations can be used as inexpensive recourse by the optimiser, and yet are nonetheless representative of the problem behaviour of interest to some quantitative degree, this approach can support fast and effective design space exploration without incurring significant cost penalty (Jiang et al., 2020). This then allows the designer to choose from a much less restrictive subset of suitable optimisation algorithms.

Incorporating approximations to accelerate computational tasks has become a well established feature of scalable numerical algorithms (Peherstorfer et al., 2018). For example, quasi-Newton optimisation methods (Broyden, 1965; Fletcher, 2013) construct approximations of system Hessians then apply low-rank updates during the Newton iterations. Solvers based on Krylov subspace methods (Lanczos, 1950) and on Anderson relaxation (Anderson, 1965) perform intermediate computations in low-dimensional subspaces which update as the computation advances.

Approximation-based approaches can be categorised according to both the type of approximation used and its method of incorporation (Sasena, 2002). To begin, if we assume a goal of minimising the number of evaluations of an expensive function and/or process, a useful approach is to employ some approximation technique to estimate the objective response and/or any comparably expensive constituent model. This is most commonly referred to as *surrogate modelling* (or *meta-modelling*) and is often applied in cooperation with one or more of the prescribed optimisation techniques, giving rise to the field of *Surrogate Based Optimisation* (SBO). SBO aims to mitigate the issues related to expensive problems by replacing direct optimisation of the function(s) of interest with inexpensive surrogate models as part

of a continuous, adaptive process. The optimisation process is thus largely guided by the iteratively updated approximation(s), with progressively less-frequent recourse to the 'real' response (Robinson et al., 2008a).

Whilst SBO methods are most commonly applied to approximate the problem objective function, a more indirect fashion is to construct surrogates of expensive subsystems instead. In this case however, the end result is the same (i.e. a reduction in the overall cost). Regardless of which function is targeted, surrogates are typically built upon sequentially sampled data of the expensive function(s). Surrogates themselves are usually assumed computationally inexpensive, continuously differentiable and of a reasonable accuracy so as to produce a fair prediction of the optimal solution (Koziel and Yang, 2011). SBO can therefore offer global insight into the problem design space and perform routine many-query analyses (such as sensitivity analyses or parametric optimisation) at a lower cost (Yondo et al., 2018).

1.1.3 Multi-Fidelity Modelling

A simple and practical SBO paradigm involves constructing a surrogate model of the function/-process of interest prior to the optimisation process (i.e. *offline*). The surrogate remains fixed and acts as a cheap alternative to the original function. Naturally then, the accuracy of the solution is therefore dependant on the accuracy of the surrogate itself. Assuming (as for most practical cases) that the optimal solution is not known *a priori*, the surrogate must demonstrate a high prediction accuracy over it's entire input domain to ensure the estimated solution is as close to the real solution as possible. Consequently, a large number of samples of the real function may be required, which could be impractical, or in the worst case, prohibitive. Given that *efficiency* is the overarching goal, numerous SBO methods have developed further in the direction of adaptively updating their approximations throughout the optimisation process (i.e. *online*). In this case, the information provided by the surrogate is used to explore the problem search space and guide the optimiser towards the optimal solution by continuously supplementing and updating the model samples. An important consideration is thus the selection of samples with which to update the approximations. An efficient algorithm must balance the inherent trade-off between *global exploration* (sampling areas of high prediction uncertainty within the design space) and *local exploitation* (concentrating samples within areas near the current optimal solution, or areas of higher significance). With respect given to this relationship, the point with the highest estimated potential to improve the surrogate quality (and thus indirectly the accuracy/reliability of the optimal solution) is selected and sampled, with all surrogates updated accordingly. Specifically sampling based on the potential for improvement implies an increased overall efficiency given that the algorithm is utilising the available information to a fuller extent. As a consequence, SBO algorithms often require fewer evaluations of both the objective and constraint functions than if applied within meta-heuristic algorithms.

However, a common theme within physics-based design and optimisation (as well as other outer-loop applications, such as inference and uncertainty quantification) is that the computational cost of constructing surrogates of expensive functions through direct sampling can nonetheless remain prohibitive (Zhou et al., 2018). This is especially true if underlying non-linearities, discontinuities and multi-modality necessitate an even higher number of samples to provide a useful representation. A potential solution is to instead incorporate multiple sources of information to supplement, and thus alleviate, the expensive function of interest. In fact, within many fields and applications across science and engineering, any particular system of interest is likely able to be represented by several alternative numerical models/formulations. These methods usually conform to a natural hierarchical trade-off between evaluation cost and relative accuracy. That is, models more representative of the real, underlying physical process(es) induce a more significant computational cost. Adopting the standard nomenclature, an expensive model that describes the system of interest to a required accuracy is referred to as *high-fidelity*, whereas any less accurate but also less demanding model is referred to as *low-fidelity*.

Multi-fidelity (or variable-fidelity) modelling involves leveraging the response trends of one or more of these inexpensive, lower quality numerical models (for example, obtained using a simulation based on a coarsely-discretised structure) to supplement a surrogate construction/refinement process. This can be achieved either in cooperation with a judicious use of the corresponding high-fidelity model, or indeed by entirely replacing the original function altogether. The fundamental premise is that lower-fidelity models can inexpensively supply the functional trends whilst only occasional recourse to the high fidelity model can be used to establish global/local accuracy, apply some correction or adjustment or guarantee outer-loop convergence (e.g. as part of a multi-fidelity optimisation scheme) (Ng and Willcox, 2014; Alexandrov et al., 2001; Koziel and Yang, 2011). In this manner, one may attempt to mitigate the "curse of dimensionality" associated with black-box surrogate modelling by exploiting domain-specific knowledge (Simpson, 2008; Keane and Nair, 2005). Currently active fields of research in this area include the types of surrogate models used, the infill criteria for selecting new sample points and the overall model management framework incorporating all (one or more) available levels of modelling fidelity.

Both single- and multi-fidelity SBO generally assumes that the expensive functions can be treated as 'black-box' models, *external* to the problem formulation itself, by which an output is obtained simply by supplying an appropriate input. These inputs could be design parameters, operational conditions or the output from another subsystem (for example). In some scenarios however, the numerical fidelity of the obtained solution(s) is more defined by the mathematical definition of the problem itself, rather than any external function/process. In which case, the standard SBO techniques as discussed cannot be applied. To explain, real-world

problems require some level of conceptualisation to appropriately formulate and represent the underlying issue as a mathematical problem. A simple example is path-finding problems, where the domain through which a suitable trajectory is to be found/optimised must first be mathematically defined at an appropriate level of abstraction. Particularly within the field of aerospace engineering, many MDO problems[3] are formulated as *Optimal Control* (OC) problems, where *trajectory optimisation and control* is included as a primary discipline. For these problems, the quality of the overall solution (e.g. how well the solution captures complex non-linear/discontinuous behaviour) can depend on such factors as the degree of polynomial used to represent the time-dependant (dynamic) states/control law and (for direct methods) the level of time-domain discretisation, as well as the time-independent (static) system design parameters. Given that the overall fidelity level of such problems is therefore *internal* to the problem formulation, an alternative approach to multi-fidelity modelling and optimisation is required.

1.2 Research Objectives

The aim of this thesis is to explore and develop numerical methods to improve the efficiency of solving complex engineering design problems using multi-fidelity techniques. In essence, increasing the attainable accuracy/reliability whilst simultaneously reducing the associated computational burden of doing so.

A further aim is to include a distinct focus on autonomy, enabling the designer to apply such techniques and methods in a wide variety of scenarios quickly and effectively. This therefore reduces the impact of increasing problem complexity, and indeed the added complexity of including multi-fidelity analyses in the first place. Naturally then, a final aim of this work is to collect the developed tools and methods into a unified design framework or library, to be made available via an appropriate open-source platform.

Being especially relevant within the field of aerospace engineering, a particular emphasis throughout the following is placed on *Multi-Objective Optimal Control Problems* (MOCPs) and analyses thereof. Considering the preceding discussion, this goal may be broken down into a number of distinct research questions:

- How can alternative levels of numerical fidelity be mathematically defined and incorporated into the formulation of a MOCP?
- How can the optimisation process be designed to adaptively manage all available fidelity levels whilst balancing the inherent cost/accuracy trade off?

[3] Including spacecraft launch or orbital manoeuvring/transfer, autonomous robotics and aircraft design (Ricciardi et al., 2016; Shi et al., 2020b; Chai et al., 2017; Vertovec et al., 2021; Bae et al., 2020; Chai et al., 2019).

- What are the operational limits of including the developed methodology with respect to traditional single-fidelity approaches?
- Are the computational savings (if any) sufficient to justify the cost of incorporating such approaches?

1.3 Research Methodology

Following on from the preceding discussion, it is helpful to consider that methods to incorporate multi-fidelity approximations into the design and optimisation process of complex engineering systems can be succinctly categorised by referring to the *source* of the available information. Where the alternative levels of fidelity for a particular physical process/subsystem of interest are defined by numerical models, simulations, experimental data or other sources inherently separate from the employed design and analysis framework, the source can be described as *external*. On the other hand, where fidelity levels are defined by the formulation, discretisation or general mathematical description (e.g. dimensionality) of the problem, the source can instead be described as *internal*.

With this in mind, the research questions presented in Section 1.2 are addressed in this thesis through the exploration and development of two distinct, but fundamentally linked, methodologies enabling the incorporation of multi-fidelity information sources. Both allow the systematic and adaptive upgrading of system/subsystem modelling fidelity for the purpose of solving *Multi-Objective Optimal Control Problems* (MOCPs) in a computationally efficient manner. These methodologies operate according to the following respective strategies according to the primary source of multi-fidelity information utilised therein:

- An **external** strategy, where alternative numerical models of the same subsystem(s), competing in accuracy and computational cost, are simultaneously incorporated into one single approximation.
- An **internal** strategy, where the numerical complexity of the problem itself can be systematically upgraded, concentrating computational resource where it is most needed.

These approaches are in fact largely analogous, with significant overlap in the fundamental issues associated with each, as will be discussed.

Each strategy is incorporated into a separate numerical framework intended for application in scenarios where in-depth and exclusively high-fidelity models/analyses are not viable. For example, the preliminary design and optimisation phases of complex multidisciplinary engineering systems.

1.4 Structure

This thesis is organised into two primary parts, corresponding to the aforementioned approximation approaches for multi-fidelity optimisation.

Part 1 deals with a *model-based* framework for **external** multi-fidelity information. This framework assumes that any particular subsystem of interest of a complex engineering system can be represented by several alternative numerical models/formulations. It is assumed that these methods conform to an inherent hierarchical trade-off between the cost of evaluation and the relative accuracy to which the underlying physical process can be represented. This framework is therefore primarily concerned with utilising, in the most computationally efficient manner possible, the available information from all data sources such that the highest numerical accuracy can be achieved at a minimum cost.

Part 2 deals with a *resolution-based* framework for **internal** multi-fidelity information. Rather than using multiple external sources of information to enhance the design procedure, this approach instead systematically refines the numerical complexity at which the problem solution is represented. This allows computational resource to be more appropriately allocated during the analysis. For example, within areas that experience high rates of change, or are particularly dominant with respect to the final solution. This approach is designed specifically for direct *Multi-Objective Optimal Control Problems* (MOCPs), where the time-dependant state and control profiles representing a feasible solution are evaluated by subdividing the time domain into a distinct grid of control nodes.

Each part follows the same basic structure. After a brief introduction and overview, relevant background information is presented and discussed, including the premise for the proposed contributions. Next, the proposed methods are presented, including theoretical premise, mathematical formulation and demonstration of validity using an appropriate series of test cases. A more complex/thorough engineering test case is included to demonstrate the practical applicability of the proposed methods to systems-level design and optimisation problems, including a brief numerical evaluation. Finally, the overall findings and contributions of the presented work are discussed, including relevant conclusions and recommendations for future study. Chapter-specific contents tables and bibliographies are included for ease of reference.

1.5 Contributions

With reference to the research questions proposed in Section 1.2, key contributions of this thesis include:

Part 1

- The dissemination of a set of generalised, cooperative, data correction methods for model-based systems design (Parsonage and Maddock, 2020).
- The sequential application of these methods to leverage complimentary aspects of the correlation between alternative numerical models.
- An analysis of their effectiveness in individual and sequential application.
- An autonomous ANN-based surrogate model generation technique for general non-linear relationships (Parsonage and Maddock, 2020, 2023).
- An adaptive sampling routine to generate and update multi-fidelity approximations based on the current solution(s) of the overarching design problem.
- A high-level framework for autonomous and adaptive multi-fidelity SBO (Parsonage and Maddock, 2023).
- A demonstration of the above methodologies in application to an aerospace engineering design problem.

Part 2

- The combination of the *Direct Finite Elements in Time* (DFET) transcription process with dyadic grid structures to facilitate adaptive solution refinement.
- A demonstration of the use of Bezier curves to enable non-destructive refinement of polynomial-based discretisation for direct Optimal Control problems.
- The use of dyadic grid structures to enable cooperative, social action between multi-objective populations of variable-complexity solutions.
- A new evolutionary crossover operator for variable length MOCPs (Parsonage and Maddock, 2024c,b).
- A new version of the *Differential Evolution* (DE) algorithm allowing the combination of variable-length solution vectors (Parsonage and Maddock, 2024b).

- A new framework based on the *Multi-Agent Collaborative Search* (MACS) algorithm incorporating the above methods for solving MOCPs (Parsonage and Maddock, 2024b).
- A demonstration of the above framework in application to a complex, computationally intensive aerospace design problem (Parsonage and Maddock, 2024c).

Software Repositories The methods and frameworks developed during completion of this thesis have been made freely available in the form of open-source software libraries. Please contact ben.parsonage@strath.ac.uk for more information.

Part I

Model-based Fidelity Management

Multi-Fidelity Surrogate Modelling and Optimisation 2

2.1 Surrogate Modelling . . . 17
2.1.1 Polynomials . . . 18
2.1.2 Radial Basis Functions . . . 20
2.1.3 Artificial Neural Networks . . . 21
2.1.4 Kriging . . . 27
2.1.5 Support Vector Regression . . . 29
2.1.6 Surrogate Validation . . . 30
2.2 Infill Criteria . . . 33
2.2.1 Local Exploitation . . . 38
2.2.2 Global Exploration . . . 40
2.2.3 Balanced Exploitation/Exploration . . . 41
2.2.4 Parallel Infill . . . 50
2.2.5 Multi-Response Sampling . . . 51
2.2.6 Multi-Objective Sampling . . . 56
2.3 Multi-Fidelity Model Management . . . 59
2.3.1 Low-Fidelity Modelling . . . 64
2.3.2 Adaptation . . . 66
2.3.3 Fusion . . . 75
2.3.4 Filtering . . . 80
2.4 Multi-Fidelity Surrogate-Based-Optimisation . . . 81
2.4.1 Classifications and Terminology . . . 84
2.4.2 Global Methods . . . 86
2.4.3 Local Methods . . . 89
2.4.4 Stopping Criteria . . . 97
2.5 Conclusion . . . 98
2.6 Bibliography . . . 100

This chapter compiles selected background information from the associated literature relevant to multi-fidelity surrogate modelling. Section 2.1 introduces several prevalent methods for generating surrogate approximations of expensive functions/processes. Included is a discussion of methods with which to numerically validate such approaches. Section 2.2 overviews several static and dynamic sampling methodologies with which a surrogate modelling process may be informed. Section 2.3 introduces the fundamental premises behind common methods of simultaneously managing alternate levels of modelling fidelity. Finally, Section 2.4 introduces a number of approaches with which the above methodologies may be seamlessly integrated into high-level global/local frameworks for numerical optimisation.

2.1 Surrogate Modelling

Surrogate models (also commonly referred to as *meta-models*) can be generally defined as functional abstractions (approximations or interpolations) of any real process that can be represented by a dataset of deterministic input/output correspondence (Simpson, 2008). The strength of the model is given by the ability to simulate unknown quantities of interest based only on limited observations. Surrogate models are often constructed only to represent specific *local* regions of the target function domain (though as will be seen later, *global* methods can be effective in some scenarios), and can be adaptively updated throughout an iterative optimisation process as new information becomes available. A generalisation of the global surrogate modelling process is visualised in Figure 2.1.

This simple conjecture has given rise to a wide range of developmental areas across many fields of science and engineering. Consequentially, in defence of Moore's law[1] (Moore and Others, 1965), it can be argued that the decreasing cost ratio of surrogate generation relative to the cost of the required expensive analyses has promoted the development and application of increasingly sophisticated surrogate modelling techniques (Simpson, 2008).

Surrogate accuracy is largely determined by three significant choices given to the designer: (1) the *sampling* of data points via a suitable *Design of Experiment* (DoE) procedure, (2) the *selection* of an appropriate meta-modelling technique and (3) the *training* of the model to fit the available data (Simpson, 2008). This chapter will first address point (2), arguably the most practically significant choice of the three, postponing discussion of points (1) and (3) to Sections 2.2 and 2.1.6 respectively.

Currently, there is no universally accepted surrogate modelling process applicable to general problems. This is largely due to the fact that prediction performance is heavily influenced by the geometric characteristics of the underlying function. For example, non-linear or multi-modal behaviour, problem dimensionality, the existence of noise or the approach and quality

[1] see Schaller (1997).

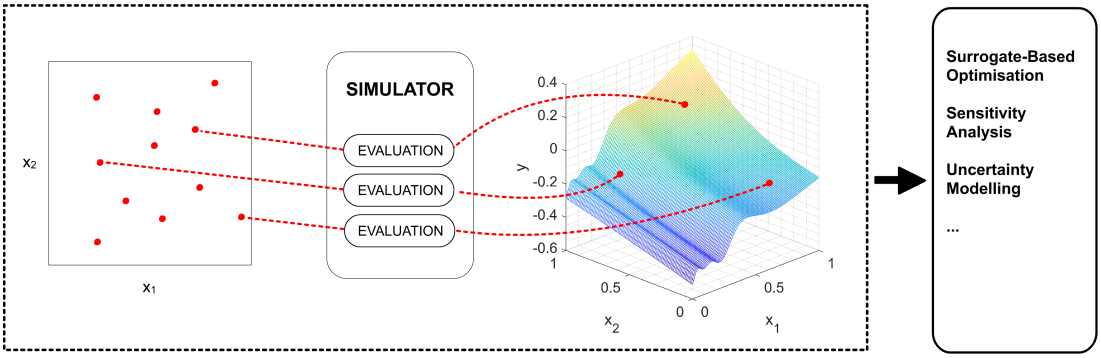


FIGURE 2.1: A visualisation of a global surrogate modelling process and it's resulting applications. First, a small number of sample points within the target function input domain are selected for evaluation using the "expensive" simulator. The resulting output data is then used to generate/train an approximating, but cheap to evaluate, surrogate model.

of the training procedure may all cause the resulting meta-model to significantly deviate from the true response (Forrester and Keane, 2009). In this context, 'true' is used in a relative sense, meaning only the response obtainable with highest available simulation fidelity, as it is acknowledged that even a high-fidelity simulation is in itself an approximation, with or without numerical noise, and may be inadequate for the task in hand (Kennedy et al., 2002). This definition naturally extends to include such sources as expert-driven knowledge, empirical or historical data, though in this work discussion is predominantly limited to that of numerical models.

A large number of surrogate modelling techniques exist in the current literature. The purpose of this chapter is not to investigate each to a particularly fine degree (with selected exceptions in future chapters), but more to introduce the reader to the wide variety of available options. For more information on surrogate modelling techniques, refer to Bhosekar and Ierapetritou (2018); Forrester and Keane (2009); Yondo et al. (2018).

2.1.1 Polynomials

A simple and widely used surrogate approximation method is polynomial regression (Forrester and Keane, 2009; Box and Draper, 1987). *Response Surface Models* (RSMs) (Myers and Montgomery, 1995; Yondo et al., 2018; Venter et al., 1998) are constructed via a linear regression of a polynomial relationship of any order across an existing dataset. Simple construction and reliable prediction of low-order responses has prompted the extensive use of RSMs as low-fidelity approximations of noisy data within multi-fidelity frameworks (Sun et al., 2010; Chen et al., 2015). A m-th order polynomial approximation of a 1D function f can be given by:

$$\hat{y}(x, m, \mathbf{a}) = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m = \sum_{i=0}^{m} a_i x^i \tag{2.1}$$

where the regression coefficients $\mathbf{a} = \{a_0, a_1, \ldots, a_m\}^T$ can be estimated through the solution of $\boldsymbol{\phi}\mathbf{a} = \mathbf{y}$, where $\boldsymbol{\phi}$ is the Vandermonde matrix, defined for approximating n-dimensional functions as (Weisstein, 2002):

$$\boldsymbol{\phi} = \begin{bmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^m \\ 1 & x_2 & x_2^2 & \ldots & x_2^m \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_n & x_n^2 & \ldots & x_n^m \end{bmatrix} \tag{2.2}$$

and $\mathbf{y}$ is the vector of observed responses. The estimation of $\mathbf{a}$ can thus be calculated as:

$$\mathbf{a} = (\boldsymbol{\phi}^T\boldsymbol{\phi})^{-1}\boldsymbol{\phi}^T\mathbf{y} \tag{2.3}$$

Equation (2.1) is similar to a Taylor series expansion of the underlying function f truncated after $m + 1$ terms, where m is the order of the resulting approximation (Box and Draper, 1987). One may expect then that a greater value of m (thus more Taylor expansion terms) will yield a more accurate approximation. However, increasing the number of higher order terms increases the risk of over-fitting to any noise that may be corrupting the underlying response. This ultimately may result in poor generalisation characteristics. This can be partially seen in Figure 2.2, where the one-dimensional numerical test function presented by Forrester et al. (2007) is approximated by several polynomial regression functions of increasing order. The highest order approximation is seen to generally fit the target function to the highest degree, but still exhibits significant error at the extremes of the input domain. Over-fitting can be prevented by employing one of a number of techniques to estimate the appropriate value of m according to different criteria (Cherkassky and Mulier, 2007). A method for dealing with numerical noise present in otherwise deterministic computer simulations is to assume to some degree that the error between the polynomial model and the data is independently randomly distributed (Forrester and Keane, 2009). The extent to which this assumption is integrated can be specified by controlling the degree of regression employed within the approximation model. Nonetheless, due to this risk of over-fitting, polynomial RSMs are considered unsuitable for highly non-linear, multi-modal or multi-dimensional function approximations (Barton, 1992; Forrester and Keane, 2009; Simpson, 2008), unless supported by a relatively large dataset of samples (Yondo et al., 2018). Particularly for higher-dimensional responses though, it may not be possible to supply the necessary amount of data for anything other than a low-order estimation (Forrester and Keane, 2009). Alternatively, one may reduce the ranges of the variables under consideration, such as is done in many trust-region methods (see Section 2.4.3.1) (Forrester et al., 2008), though this is in fact a general mitigation strategy potentially applicable to all approximation approaches.

An extension to this method is to instead define a series of polynomials in a piecewise manner, rather than employing only one over the entire functional range. This allows a more reliable

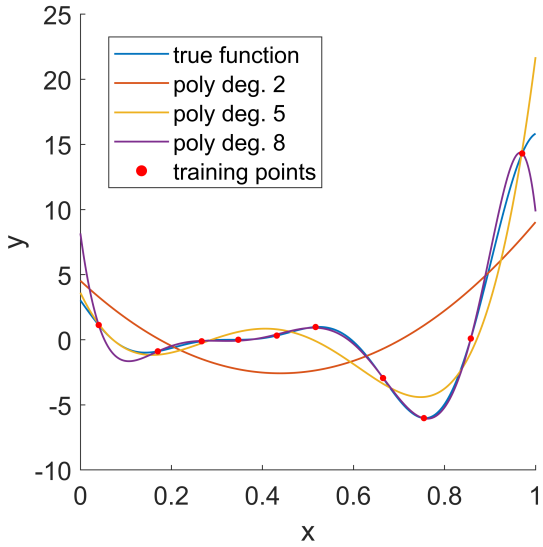


FIGURE 2.2: An example of several polynomial regressions of increasing order, trained using a small number of samples of the 'true' function (in this case, the one-dimensional test function in (Forrester et al., 2007)).

approximation of non-linear responses with less risk of over-fitting. The available data is split into separate ranges, with a polynomial approximation constructed over each. Boundary conditions are defined to ensure each section meets a prescribed order of continuity. For example, it is common to define C2 continuity such that consecutive polynomials match value, slope and curvature at the interface point between them. *Interpolating splines* assume that each data point is itself a boundary locus, such that each lies exactly on the approximating polynomials. While this means interpolating splines tend to match the available data to a stronger degree than polynomials tuned with a weighted least-squares approach, the nature of the response between data points can still exhibit higher variability (Sasena, 2002). A variation on this approach, known as the *smoothing spline*, adjusts a weight factor (the smoothing parameter) such that a compromise between the goodness of fit seen in interpolating splines and the smoothness of the approximating functions seen in least squares models can be achieved (Craven and Wahba, 1978; Wahba, 1990).

2.1.2 Radial Basis Functions

Related to polynomial RSMs, and indeed largely replacing them in a number of fields (Forrester and Keane, 2009), *Radial Basis Functions* (RBFs) attempt to use a weighted combination of simple functions to approximate complex responses (Broomhead and Lowe, 1988). In fact, the standard formulation of RBF is equivalent to an ANN with one hidden layer and radial coordinate neurons (see Section 2.1.3). To demonstrate, consider a function f observed in absence of error according to the sampling plan $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}^T$ which yields the

responses $\mathbf{y} = \{y_1, y_2, \ldots, y_n\}^T$. An RBF approximation $\hat{\mathbf{y}}$ may then be constructed according to the fixed form:

$$\hat{\mathbf{y}} = \hat{f}(\mathbf{x}) = \mathbf{w}^T \boldsymbol{\psi} = \sum_{i=1}^{n_c} w_i \psi(\|\mathbf{x} - \mathbf{c}_i\|) \tag{2.4}$$

where $\mathbf{c}_i$ is the i-th of n_c basis function centres and $\boldsymbol{\psi}$ is an n_c vector containing the values of the basis functions ψ themselves, evaluated at the Euclidean distances between the prediction site $\mathbf{x}$ and the centres of the basis functions.

Basis functions may be, for example, linear ($\psi(r) = r$), cubic ($\psi(r) = r^3$) or thin plate splines ($\psi(r) = r^2 \ln r$). More complex, parametric bases (Gaussian: $\psi(r) = e^{-r^2/2\sigma^2}$, multi-quadric: $\psi(r) = (r^2 + \sigma^2)^{1/2}$, inverse multi-quadric: $\psi(r) = (r^2 + \sigma^2)^{-1/2}$ etc.) may be used to improve generalisation. If fixed bases are used, there is one undetermined weight per base. Furthermore, if based on the Gaussian density function, RBFs typically lead to numerically stable computations of coefficients of the linear combination of the data-fit model (Peherstorfer et al., 2018).

Weights can be estimated using the interpolation condition:

$$\hat{f}(\mathbf{x}_j) = \sum_{i=1}^{n_c} w_i \psi \|\mathbf{x}_j - \mathbf{c}_i\| = y_j, \quad j = 1, \ldots n \tag{2.5}$$

However, RBF additionally requires the specification of a number of other hyperparameters, e.g. the bandwidth of the Gaussian density function. Whilst an appropriate choice of weights will ensure the resulting approximation accurately represents the training data, a similarly appropriate choice of hyperparameters is needed to minimise the (estimated) generalisation error of the data-fit model (Forrester et al., 2008). Selection and optimisation of these hyperparameters can often be computationally expensive (Peherstorfer et al., 2018).

2.1.3 Artificial Neural Networks

The preceding methods generally incorporate certain fundamental assumptions regarding the sampled data. For example, whilst using interpolating splines (Section 2.1.1) does not require the fit and analyse approach commonly employed by weighted least squares models, they still assume that the sample data can be locally approximated using at least a cubic polynomial (Sasena, 2002).

A method becoming particularly popular, in part due to its ability to fit data outwith this kind of assumption, is the use of *Artificial Neural Network* (ANN). An ANN is a collection of simple computational units arranged into an inter-connecting parallel processing structure. A network may be trained according to a finite set of input vectors and corresponding target

vectors. A suitably trained ANN may perform tasks such as pattern recognition, identification, classification, system control and function approximation (non-linear regression) (Beale et al., 2017; Cheng and Titterington, 1994). ANNs offer a number of advantages for surrogate modelling applications, including dimensionality reduction, implicit sensitivity analysis and the availability of multiple transfer functions and training algorithms (Yondo et al., 2018).

McCulloch and Pitts (1943) originally theorized the ANN in 1943 as an attempt to mimic the processing structure of biological neural pathways and thus gain a deeper insight into brain function. Despite an initially fruitful investigation into the underlying mathematics by the scientific community, including development of the first learning rule by Hebb (1949), the first perceptron by Rosenblatt (1958) and the adaptive linear element in Widrow and Hoff (1960), the field was ultimately stagnated by the relative juvenility of computational processing technology (Suykens et al., 1996; Paiva et al., 2010). The widespread emergence of the integrated digital microprocessor in the 1980s initiated a resurgence in ANN research. Notable contributions from this time include the development of the associative memory network by Hopfield (1982) and the self-organizing map by Kohonen (1982).

ANNs have found extensive use in many fields. Hajela and Berke (1992) review the use of neural networks in structural analysis and design. Cheng and Titterington (1994) review the uses of ANNs from a statistical perspective. In aerodynamics, ANNs have been used for the estimation of aerodynamic coefficients (Santos et al., 2008; Linse and Stengel, 1993; Wallach et al., 2006), space vehicle design and trajectory optimisation (Roshanian et al., 2019; Minisci and Vasile, 2013), design optimisation of turbo-machinery blades (Mengistu and Ghaly, 2008), wing design (Rai and Madavan, 2000; Paiva et al., 2010), flow control, aeroelasticity, interpolation of wind tunnel data and classification problems (Steck et al., 1997; Yondo et al., 2018). Eason and Cremaschi (2014) applies an ANN to predict CO_2 absorption/ desorption rates using aqueous amines.

ANNs have also found particular application in multi-fidelity modelling, albeit with an initially slower uptake (Fernández-Godino et al., 2016). Bakr et al. (2000) present a space-mapping based neuro-modelling optimisation technique in which the high-fidelity optimum is found by replacing the coarse model at each iteration with an adaptively-sized ANN trained using the high-fidelity samples from the previous iteration. Leary et al. (2003) introduced the *Knowledge-Based Neural Network* (KBNN) technique as a means of incorporating low-fidelity data into the training procedure of a limited set of high-fidelity samples via scaling/translation based parameter mapping. Minisci and Vasile (2013) estimate aerodynamic coefficients using a multi-fidelity fusion/hierarchical ANN surrogate first trained on a global set of samples given by a simplified analytical model, then iteratively refined with CFD data scheduled via evolution control. Low fidelity samples are sequentially discarded should they fall outwith a predefined applicability region relative to new high-fidelity points.

More recently, multi-fidelity methods have been applied to enhance the training of *Physics-Informed Neural Networks* (PINNs) (Toscano et al., 2025). For example, Meng and Karniadakis (2020) employ several networks to 1) approximate the low-fidelity input data, 2) learn the linear/non-linear correlations between low and high-fidelity data and 3) encode the underlying *Partial Differential Equations* (PDEs). Regazzoni et al. (2021) extended this methodology to Bayesian networks to help quantify the uncertainty in the resulting prediction. Wang and Lai (2024) introduced a multi-stage ANN that divided the training process into sequential stages, where each involved a new network fit to the residuals of the previous. Penwarden et al. (2022) offer an approach to combine different ANNs of varying degrees of fidelity by exploiting their low-rank structure.

The use of ANNs specifically in multi-fidelity surrogate modelling is discussed in more detail later in chapter 3. Here, the premise and formulation of a standard ANN is briefly introduced in the context of generalised surrogate modelling.

The standard computational unit used to construct an ANN is the single-input *neuron* (also referred to as a *single-unit perceptron*). A neuron processes an input according to the general form:

$$a = f(w * p + \beta) \tag{2.6}$$

where p, w and β are the scalar input, weight and bias respectively, and f is a monotonic transfer function, typically either a step, linear or log-sigmoid function. More complex neuron formulations exist in the literature, including the radial basis perceptron used to construct *Radial Basis Function Neural Networks* (RBFNNs). However, this work concentrates only on the regular classification, given the improved generalisation at input sites far from the training set (Minisci and Vasile, 2013).

A single neuron on its own is of course limited in computational functionality. To form a network, it is required to group neurons into an appropriate architecture. One of the more widely used architectures is the *Multi-Layer Perceptron* (MLP), part of the original set of classifications proposed by McCulloch and Pitts (1943). A *layer* is defined as a distinct set of neurons accepting a common input with no intra-layer connections between neurons (Paiva et al., 2010). A feed-forward MLP architecture comprises of three distinct regions of sequentially inter-connecting layers of neurons (see Figure 2.3); the input layer, one or more hidden layers and the output layer (Suykens et al., 1996).

A multi-layer network may contain any number of hidden layers. However, it has been widely reported that an ANN with biases, one sigmoidal hidden layer (chosen due to being easily differentiable) and a linear output layer (to allow a non-bounded output) is capable of approximating any linear/non-linear function with a finite number of discontinuities to an arbitrary accuracy, provided that associated conditions are satisfied (Razavi et al., 2012; Beale

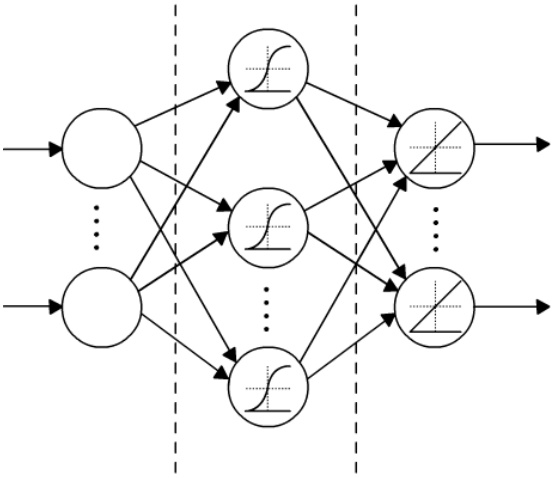


FIGURE 2.3: A feed-forward MLP architecture, including an input layer, one hidden layer and an output layer.

et al., 2017; Rumelhart et al., 1994). An additional implication here is that one may employ an ANN based surrogate modelling scheme without any prior knowledge of the underlying function to be approximated (Minisci and Vasile, 2013), a particularly useful property for SBO applications.

For clarity, the simplified notation presented by Beale et al. (2017) is adopted:

$$\mathbf{a}^1 = f^1(\mathbf{W}^{1,1} * \mathbf{p} + \mathbf{b}^1) \tag{2.7}$$

where each layer has a $S \times R$ weight matrix $\mathbf{W}$, a S-length bias vector $\mathbf{b}$, and an S-length output vector $\mathbf{a}$. The vector $\mathbf{p}$ is an R-length input vector. From this notation, a 3-layer feed-forward network with $R = 2$ inputs and $S = 4$ hidden layer neurons may be graphically depicted as in Figure 2.4. In this case, the weights connected to the input vector are denoted $\mathbf{IW}$ and the weights connected to a preceding layer are denoted $\mathbf{LW}$.

The process by which a network may be trained is known as back-propagation, itself a generalisation of the Widrow and Hoff (1960) learning rule applied to multi-layer networks and non-linear differentiable transfer functions (Beale et al., 2017). Back-propagation updates the network weights and biases proportional to the negative direction of the gradient (i.e. the direction in which the performance function decreases most rapidly):

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \frac{\partial J(\mathbf{x})}{\partial \mathbf{x}_k} \tag{2.8}$$

where $\mathbf{x}_k$ is the current vector of weights and biases, $\partial J(\mathbf{x})/\partial \mathbf{x}_k$ is the current gradient of the cost function J (e.g. most commonly the *Mean Square Error* (MSE)) with respect to $\mathbf{x}_k$ and α_k is the learning rate.

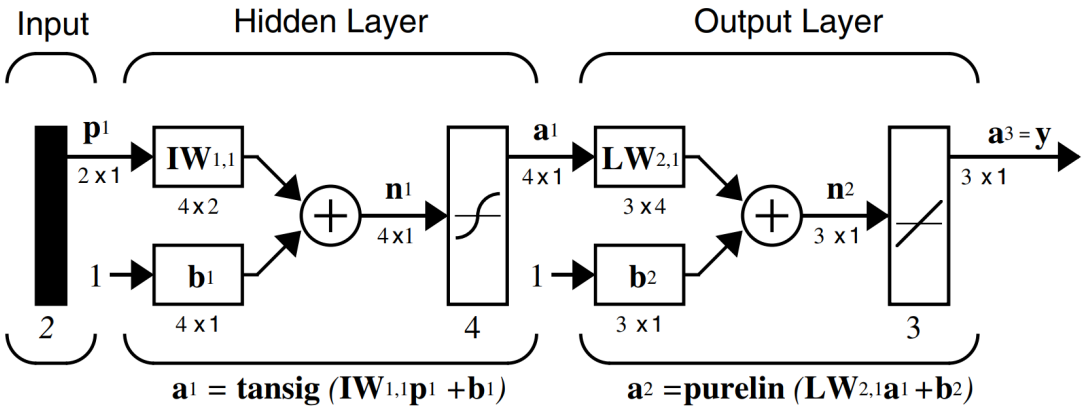


FIGURE 2.4: Simplified notation of a 3-layer MLP (Beale et al., 2017). The weights connected directly to the input layer are denoted **IW**, with weights connected only to the immediate preceding layer denoted as **LW**. Layer outputs and biases are denoted by **a** and **b**respectively. Superscripts are used to identify the destination (first index) and source (second index) for the various elements of the network. For example, Elements with superscript 1 are associated with the first layer.

However, non-linear transfer functions in a multi-layer network can introduce many local minima in the error surface (Beale et al., 2017). Thus the use of such a gradient-descent method described by (2.8) risks the entrapment of the network solution in one of these local minima for a given initial condition. While not necessarily undesirable, especially if the local minima is close to the global minimum, the result is that a single back-propagation training procedure on any multilayer network, even with sufficient neurons in the hidden layer, cannot guarantee the convergence of the weights to the optimum solution.

The training procedure may be described as either supervised or unsupervised. In supervised learning, data is presented for training in a coupled input/target format as follows:

$$\{\mathbf{p}_1, \mathbf{t}_1\}, \{\mathbf{p}_2, \mathbf{t}_2\}, \ldots, \{\mathbf{p}_n, \mathbf{t}_n\} \tag{2.9}$$

The network processes the inputs, then compares the resulting outputs to the target values. The network weights and biases are then updated via a training algorithm to iteratively reduce the error between the outputs and the targets. This is opposed to unsupervised learning, in which the network is not supplied with targets. The reader is directed to Beale et al. (2017) for a more in-depth discussion on supervised vs unsupervised learning.

A further classification may be made with respect to the manner in which input data is presented for training. Input/target pairs may be presented concurrently (batch training) or sequentially (incremental training), the latter to represent temporal variations or include new data as it is simulated. Batch training only allows the update of all weights and biases after the

entire training set has been applied to the network, and thus is notably more computationally efficient than incremental training (Beale et al., 2017).

Multi-layer networks are extremely sensitive to the number of neurons included within their hidden layer(s). Too few neurons and the network cannot represent non-linear responses in appropriate detail, and thus converges only to sub-optimal solutions, if at all. Increasing the number of neurons increases the complexity of the functions that can be approximated at a further computational cost. However, too many neurons can result in over-fitting, resulting in poor generalisation and often incurring wild oscillations in function response, despite negligible errors at the training points. Kolmogorov (1957) states that this behaviour implies the existence of an optimal set of parameters that can allow a neural network to approximate any function for any application.

However, it can often be very difficult to estimate *a priori* the number of neurons that should be included in each hidden layer for a particular application. Some authors typically employ simple trial and error schemes, the continuous re-training of multiple networks or discrete-optimisation-capable methods (e.g., Genetic Algorithms) (Paiva et al., 2010). A recent development in this area is the emergence of software solutions for solving Conditional Search Space Problems (Lévesque et al., 2017). That is, problems in which the structure, dimensionality, or available variables/parameters change (are conditional) based on decisions made earlier. Considering then the problem of automated hyperparameter optimisation/tuning, an example is the Optuna framework for Python-based applications Akiba et al. (2019).

To help mitigate this issue, focus can instead be shifted to improving the generalisation of the approximate response. The two most commonly applied methods here are early-stopping and regularisation. It should be noted that the likelihood of over-fitting is proportional to the ratio of network parameters to available training data. If obtaining more data is computationally viable, there is little need to employ generalisation improvement techniques, thus the following discussion is of primary concern to computationally expensive problems with a limited available budget of training data (Beale et al., 2017).

To implement an early-stopping procedure, the available training data must first be divided (according to an appropriate data division algorithm e.g. indexed, random or block data division) into three subsets: the *training* set, the *validation* set and the *testing* set. The *training* set is used to update the network weights and biases via back-propagation. The *validation* set prediction error is monitored during the training process. Typically the validation error will decrease in correspondence with the training set error as the network is fit to the available data. However, at the onset of over-fitting, the validation error will begin to rise. If this increase persists for a certain specified number of iterations, the training procedure is stopped and the weights and biases at the minimum validation error are selected. The *testing* set simply

provides an independent sample set not used during the training process.

Regularisation involves modifying the network performance function to minimise network complexity. The typical performance function for feed-forward networks is the the mean sum of squares of the network errors (MSE):

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(e_i)^2 = \frac{1}{N}\sum_{i=1}^{N}(t_i - \alpha_i)^2 \tag{2.10}$$

The magnitude of the network weights and biases is inversely proportional to the overall smoothness of the response. Therefore, it is of interest to the designer to minimise these terms by including the mean sum of squares of the network weights and biases as an additional term in the performance function:

$$\text{MSE}_{reg} = \gamma\,\text{MSW} + (1-\gamma)\text{MSE} \tag{2.11}$$

where γ is the performance ratio and

$$\text{MSW} = \frac{1}{n}\sum_{j=1}^{n} w_j^2 \tag{2.12}$$

The Bayesian framework presented by MacKay (1992) assumes the weights and biases of the network are random variables with specified distributions. The regularisation parameters are related to the unknown variances associated with these distributions and therefore can be determined using statistical techniques. The reader is referred to Foresee and Hagan (1997) for a detailed discussion of the use of Bayesian regularisation, in combination with Levenberg-Marquardt training.

2.1.4 Kriging

A common assumption in surrogate modelling is that the true response can be represented by the approximate response plus some residual error term:

$$y(x) = \hat{y}(x) + \epsilon \tag{2.13}$$

where the residuals $\epsilon(x)$ are independent, identically distributed normal random variables, $\epsilon\, N(0, \sigma^2)$. Kriging (Sacks et al., 1989) is a form of Gaussian Process Regression that assumes that the errors in the predicted values are *not* independent, but instead a systematic function of x. The approximation of the true function response is constructed as a weighted combination of:

1. a linear regression $f(\mathbf{x})$ via a set of basis functions modelling the drift of the process mean (representing low-frequency variations), and

2. a realisation of a stochastic process $Z(x)$ to model high-frequency variation by quantifying correlation between data points.

This can be written as:

$$Y(\mathbf{x}) = \sum_{j=1}^{k} \beta_j f_j(\mathbf{x}) + Z(\mathbf{x}) \tag{2.14}$$

where $f_j(\mathbf{x})$ are the basis functions (i.e., the polynomial terms), β_j are the corresponding coefficients, and $Z(\mathbf{x})$ is a Gaussian process. Note, $Y(x)$ is capitalised to specify it as being a random variable, as opposed to any particular realisation of such, $y(x)$. Similarly, the random variable $Z(\mathbf{x})$ represents uncertainty in the mean of $Y(\mathbf{x})$ with $E(Z(\mathbf{x})) = 0$ and

$$\operatorname{cov}\left(Z(\mathbf{x}^{(i)}), Z\left(\mathbf{x}\right)\right) = \sigma_z^2 \mathrm{R}\left(\mathbf{w}, \mathbf{x}\right) \tag{2.15}$$

Here, σ_z^2 is a scale factor (known as the *process variance*) that can be tuned to the data and $\mathrm{R}(\mathbf{w}, \mathbf{x})$ is the *Spatial Correlation Function* (SCF). There are numerous choices of $\mathrm{R}(\mathbf{w}, \mathbf{x})$ to quantify the manner by which the approximate function moves between observed data points. One of the most common may be written in k-dimensional form as:

$$\mathrm{R}(\mathbf{x}^{(i)}, \mathbf{x}) = \operatorname{cor}\left[Y(\mathbf{x}^{(i)}, Y(\mathbf{x}))\right] = \exp\left(-\sum_{j=1}^{k} \theta_j \left|x_j^{(i)} - x_j\right|^{p_j}\right) \tag{2.16}$$

The variance of the basis function can be controlled by $\theta_j > 0$ and, instead of being fixed at 2, the exponent can be varied by p_j. Where the influence of a sampled data point on the point to be predicted reduces as $|x^{(i)} - x|$ increases, the deterioration of this influence is determined by θ. On the other hand, the exponent control p determines the smoothness of the resulting approximation. Kriging allows the specification of different basis functions and correlation parameters for each dimension, though this is perhaps unnecessary for most problems (Sacks et al., 1989). Of note, if the model parameters are chosen such that $\mathrm{R}(\mathbf{x}, \mathbf{x}) = 1$, the resulting trained predictor will pass exactly through any measured data point. More advanced techniques, such as Blind Kriging (Joseph et al., 2008; Couckuyt et al., 2012), exist to allow a non-interpolating prediction.

The unknown parameters θ_j and p_j are often estimated as *Maximum Likelihood Estimators* (MLEs), maximising the natural logarithm of the likelihood with constant terms removed. The concentrated *ln-likelihood function* is written as:

$$\ln(L) \approx -\frac{n}{2} \ln \hat{\sigma}^2 - \frac{1}{2} \ln |\Psi| \tag{2.17}$$

where

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - 1\mu)^T \mathbf{\Psi}^{-1} (\mathbf{y} - 1\mu)}{n} \tag{2.18}$$

and $\mathbf{\Psi}$ is an $n \times n$ matrix of correlations between the sample data, with each element given by (2.16). The maximisation of (2.17) represents the primary computational expense of the Kriging technique. The estimation of θ_j and p_j can often prove costly, resulting in overall lengthier model training times, thus Kriging is of most use when the true function is particularly expensive computationally. With parameters estimated, approximate function values at an unknown point $\mathbf{x}$ can be predicted as:

$$\hat{y}(\mathbf{x}) = \hat{\mu} + \boldsymbol{\psi}^T \mathbf{\Psi}^{-1}(\mathbf{y} - \mathbf{1}\hat{\mu}) \tag{2.19}$$

where

$$\hat{\mu} = \frac{\mathbf{1}^T \mathbf{\Psi}^{-1} \mathbf{y}}{\mathbf{1}^T \mathbf{\Psi}^{-1} \mathbf{1}} \tag{2.20}$$

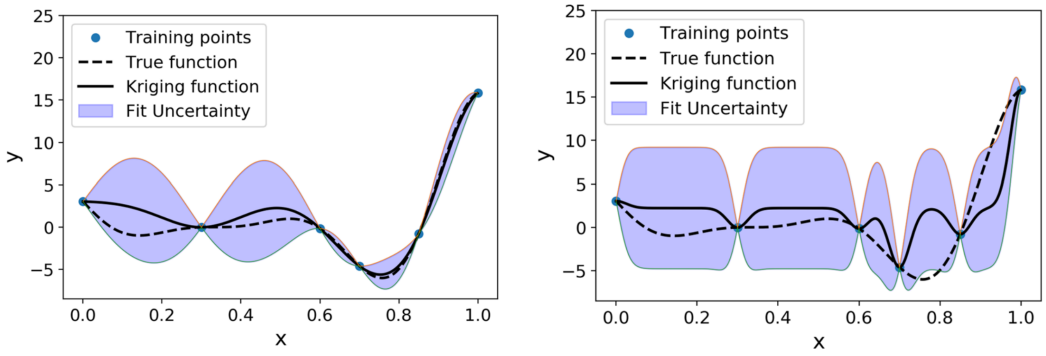
A further strength of Kriging (as with other models based on a Gaussian process), widely supporting its use in surrogate-based optimisation, is the provision of an estimated error. The MSE for a Kriging model can be given by (Sacks et al., 1989):

$$s^2(\mathbf{x}) = \sigma^2 \left[1 - \boldsymbol{\psi}^T \mathbf{\Psi} \boldsymbol{\psi} + \frac{1 - \mathbf{1}^T \mathbf{\Psi}^{-1} \psi}{\mathbf{1}^T \mathbf{\Psi}^{-1} \mathbf{1}} \right] \tag{2.21}$$

The wide applicability of the Kriging approach has given rise to numerous modifications and extensions to the fundamental process, in addition to a wide range of hyper-parameter optimisation methods (the importance of which being visualised in Figure 2.5). The above formulation is classified as *Ordinary Kriging* (OK) (Kennedy and O'Hagan, 2000; Grabe et al., 2012), though alternative formulations include, among others, *Universal Kriging* (UK) (Zimmerman et al., 1999; Bouhlel et al., 2016), *Blind Kriging* (BK) (Joseph et al., 2008; Couckuyt et al., 2012), Non-Stationary (covariance) Kriging (Xiong et al., 2006; Toal and Keane, 2012), *Gradient-Enhanced Kriging* (GEK) (Han et al., 2013; Chung and Alonso, 2002) and Co-Kriging (Forrester et al., 2007; March et al., 2011). The reader is referred to Jones (2001) for a complete derivation of the OK method for surrogate approximation.

2.1.5 Support Vector Regression

In a similar fashion to Kriging (Section 2.1.4) and RBFs (Section 2.1.2), *Support Vector Regression* (SVR) involves the linear combination of a weighted series of basis functions. SVR distinguishes itself from alternative approaches through the principle of risk minimisation; the ability to specify or calculate an acceptable error margin (defined as a region of inapplicability with regards to a specified penalty function) to mitigate the effects of erroneous samples on the response surface (Forrester and Keane, 2009). This is useful if the supplied sample data contains some element of randomised error. For example, due to a finite mesh discretisation. In departure from previously discussed methods, the objective is then to minimise the upper bound of the generalised error, rather than the error itself. SVR is attractive for surrogate



(a) Optimised hyper-parameters

(b) Overestimated hyper-parameters

FIGURE 2.5: Example of Kriging applied to approximate a one-dimensional function (Forrester et al., 2007). The blue space is the uncertainty of the kriging model, where (a) shows the prediction obtained using optimised hyper-parameters, and (b) shows the prediction obtained instead using overestimated hyper-parameters.

modelling approaches given its reliable handling of multidimensional problems, even with limited datasets. Additional flexibility is encouraged through the user-specification of different Kernel functions. However, SVR requires the non-trivial optimisation of hyperparameters in addition to the learning phase (Yondo et al., 2018).

The basic form of the SVR prediction is the familiar sum of basis functions $\psi^{(i)}$,with weightings $\mathbf{w}^{(i)}$, added to a base term μ; all calculated in different ways to their counterparts in the Kriging and RBF literature, yet contributing to the prediction in the same way:

$$\hat{f}(\mathbf{x}) = \mu + \sum_{i=1}^{n} w^{(i)} \psi(\mathbf{x}, \mathbf{x}^{(i)}) \tag{2.22}$$

The theory behind SVR stems from the work completed during the 1990s regarding Support Vector Mechanics (Vapnik, 1999). The reader is referred to (Forrester and Keane, 2009) for a complete derivation of the SVR predictor.

2.1.6 Surrogate Validation

The effective quality of a particular surrogate model is often evaluated relative to a number of metrics assessing both local and global performance (Yondo et al., 2018; Forrester et al., 2008; Viana et al., 2014; Goel et al., 2007). The most common approaches involve estimating the approximation error. That is, the expected error between the generated model and an unseen sample. Such methods are helpful in establishing whether a particular surrogate modelling approach should be used over another, or even assessing its own applicability to a particular problem.

Some of the methodologies described in the previous sections construct the surrogate model together with some estimation of the attendant approximation error (e.g. Kriging, see Section 2.1.4). Alternatively, one may employ a number of procedures in a stand-alone manner to validate the prediction capability of a given model beyond the set of n_s training points.

First, recall the difference between the training error and the testing error. If $\tilde{f}$ denotes a surrogate of the expensive function f, and if $\mathbf{x} \in \mathcal{R}$ denotes the sampling plan such that:

$$\mathbf{X}_i \rightarrow y_i = f(\mathbf{x}^{(i)}) | i = 1, 2, \ldots, n_s \tag{2.23}$$

where

$$\mathbf{x} = \left[\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n_s)}\right]^T$$
$$\mathbf{y} = \left[y(\mathbf{x}^{(1)}), \ldots, y(\mathbf{x}^{(n_s)})\right]^T$$

then the training error is defined as the average loss over the trained samples:

$$e_{tr} = \frac{1}{n_s} \sum_{i=1}^{n_s} \mathbf{L}(y_i, \tilde{f}(\mathbf{x}_i)) \tag{2.24}$$

The testing error is the expected prediction error over an independent sample set:

$$e_{ts} = \mathbf{E}_{\mathbf{x},\mathbf{y},ts}[\mathbf{L}(\mathbf{y}, \tilde{f}(\mathbf{x})] \tag{2.25}$$

where $\mathbf{L}$ is the loss function(s). Commonly used loss functions include the square and absolute errors (for continuous responses) or the 0-1 loss and log-likelihood (for responses of discrete values).

Training errors may offer only poor estimates of the testing errors, dropping towards zero as model complexity increases (Yondo et al., 2018). The surrogate construction process typically requires the (random) splitting of the available data into three sets: the *training* set (the points to which the surrogate should be fit to), the *validation* set (to evaluate the unbiased prediction error), and the *testing* set (used to assess the generalised error of the final chosen model). Deciding the proportions of samples to assign to each of these sets remains an open question. Common practice is to allocate 50-60% to the training set and 20-30% to both the testing and validation sets. An initial temptation may be to utilise 100% of the available data to train the surrogate. However, doing so greatly increases the risk of over-fitting and poor generalisation, with all error estimates being lower than the true errors. This effect becomes more pronounced with models that have a larger number of parameters. The validation of the testing error estimate is largely dependent on the type and amount of trained samples contained in both the training and validation sets. A disadvantage here is that the validation set error most often tends to overestimate the test error of the model fit on the entire data set.

Given that the training error of an ideal well-fit surrogate will tend to zero, surrogate validation is most often performed using the estimation of the testing set error (also referred to as the *prediction* error) as a more reliable metric. Numerous methods for estimating the testing set error exist in the literature. Popular examples include the Hold-Out method, the Split Sample method, Cross-Validation (inc. Random sub-sampling, Leave-One-Out and K-Fold), Bootstrapping, Schwartz's *Bayesian Information Criterion* (BIC) and *Akaikes Information Criterion* (AIC). In the Split-Sample algorithm (McCulloch and Pitts, 1943), the set of available samples is divided into training and testing subsets. The value of error estimated using this approach depends strongly on how the set of samples is partitioned. Furthermore, the binary splitting of data suggests that a rather large proportion of data is not used in surrogate construction, which may affect the overall quality to a greater detriment relative to the benefit of inclusion within the testing set. *Cross-Validation* (CV) (McCulloch and Pitts, 1943; Golinski, 1973) is a particularly popular methodology. Input data is divided into L subsets, where each is sequentially employed as a testing set for a separate surrogate constructed on the other $L \breve{} 1$ subsets. The case in which the number of subsets L is equal to the sample size p is equivalently known as *Leave-One-Out* (LOO) CV (McCulloch and Pitts, 1943). The prediction error can be estimated from all L error measurements obtained using this process (for example, as an average value). The CV error estimation is generally less biased than that of the split-sample method (McCulloch and Pitts, 1943). A disadvantage is the requirement of multiple separate surrogates, which may be computationally infeasible depending on the application, size of sample set or surrogate modelling methodology employed. However, it can be counter-argued that the inclusion of multiple approximations may improve the robustness of the entire surrogate generation and validation approach, since all of the available data is utilised for both training and testing purposes.

The designer is thus presented with yet another choice, as certain error-estimation methods may be unsuitable or erroneous in themselves for a given application. For example, using CV to validate a Kriging model may demand an excessive computational cost and may lead to erroneous model validation (Wang and Shan, 2007).

More standardised metrics, generally more affected by the domain-level distribution of training points, include the *Root Mean Square Error* (RMSE), the *Adjusted Mean Square Error* (AMSE), the *Weighted Root Mean Square Error* (WRMSE), the *Integrated Mean Square Error* (IMSE), the *Maximum Absolute Error* (MAE), the *Prediction Error Sum of Squares* (PRESS), the *Relative Accuracy Error* (RAE) or Pearson's correlation coefficient (Queipo et al., 2005; Jin et al., 2001). Given an inherent bias relative to sample point distribution, many of these methods may in fact offer only a poor representation of the actual error.

2.2 Infill Criteria

The performance of any surrogate modelling technique can be extremely dependant on the number of, location and distribution of the training samples (Liu et al., 2018a; Yondo et al., 2018). The initial selection and subsequent update of the sample set plays a particularly important role in a surrogate modelling and optimisation process. A high quality surrogate increases the reliability of the final obtained optimal solution or may allow a more accurate identification of important inputs for a sensitivity analysis (Liu et al., 2018a), though typically requires a high number of samples, which often proves computationally prohibitive. The speed of convergence is directly correlated to the efficiency of the modelling process. An additional goal then, particularly when considering expensive functions, is naturally to sample the design domain in as few function evaluations as possible whilst still achieving a suitably accurate representation of the underlying function(s).

The judicious selection of new samples, referred to as *infill* points, represents the heart of most surrogate-based optimisation methods (Forrester et al., 2008). Newly selected points are used to both validate the current approximation(s), then enhance the prediction quality even further. The determination of the most suitable infill points is accomplished by defining certain criteria (the *infill criteria*) for selection. This is alternatively known as adaptive sampling, sequential sampling, or active learning and refers to the iterative (potentially parallel) sampling of the target function in promising areas of the input space based on information gathered from the existing and/or newly updated surrogate (Forrester et al., 2008; Yondo et al., 2018).

The premise of most infill criteria, especially when considering expensive functions, is to balance the trade off between the global surrogate accuracy and the local accuracy of 'interesting' (e.g. highly non-linear or multi-modal) regions, or within region(s) most likely to enclose (or largely influence) the problem optimum (Yondo et al., 2018). This is most often referred to as the balance between *global exploration* and *local exploitation*.

Numerous infill sampling criteria (Forrester et al., 2008; van Beers and Kleijnen, 2008; Da Ronch et al., 2017; Martínez-Frutos and Herrero-Pérez, 2016; Parr et al., 2012b,a) exist to guide the process during the exploitation phase. Examples include; *Searching the Surrogate Model* (SSM), minimising the predictor and/or maximising the predicted error, the *Lower Confidence Bounding* (LCB), the *Expected Improvement* (EI), the *Probability of Improvement* (PI), likelihood estimation or the Kushner criterion. A taxonomy on data-fit surrogates-based global optimisation methods has been published by Jones (2001), introducing several infill criterion methods. Forrester and Keane (2009) and Liu et al. (2014) compared methods for selecting search points within Kriging-based optimisation frameworks. MacKman et al. (2013) compared sequential sampling methods for generation of aerodynamic surrogate models.

The generation of a sample set within a specific domain is achieved via a suitable *Design of Experiment* (DoE)[2] technique. Originally, the conditions under which to run experiments (i.e. the input conditions) were planned *a priori* using one of the following simplistic methods (Montgomery, 2001):

- Varying one factor or variable at a time.
- Indefinite trial and error, aiming to ascertain the best guesses or sample point locations.

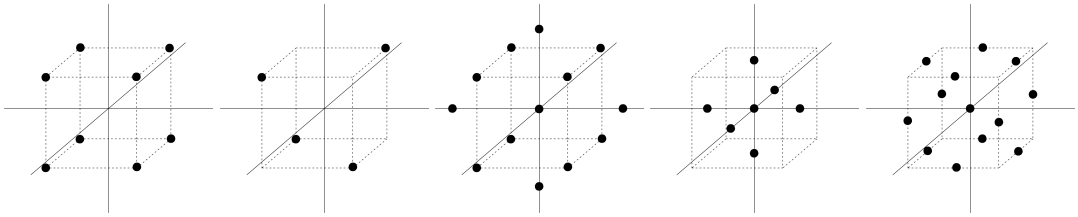
These methods often not only failed to consider the interaction between variables (or factors), but also required significant expense given the number of individual tests, with no guarantee that a suitable sample location could even be found (Yondo et al., 2018). The term 'Design of Experiments' was originally proposed by Fisher (1926) as a means to bypass these limitations, offering a systematic method to plan experiments based on an understanding of the probabilistic behaviour of agricultural crops systems. Originally concerned with only physical experiments, work from this period established the basic principles of DoE, namely; randomisation, replication, and blocking, for controlling random data-corrupting errors (noise) and bias.

With the introduction of numerical methods and computing facilities, simulation-based experiments were quickly established with design space sampling a key issue. Similarly to physical experiments, the prevailing goal was maximising the information gained from a minimum/limited set of individual sample points. A detailed discussion on physical experiments versus simulation-based is available in Kleijnen and Van Beers (2004).

DoE methods can generally be classified as either *static* (one-shot) or *dynamic* (sequential). Static DoE determines both the sample size and locations in a single step, whereas dynamic methods iteratively cluster points in regions of interest selected via previous model evaluations/data in order to reduce the overall number of samples needed to represent a given functional response, or guide a searching algorithm to a local/global optimum (Tyan and Lee, 2019; Liu et al., 2018a).

Static approaches for deterministic computer simulations are usually concerned with satisfying the *space filling* criterion. That is, exploring the nature of the underlying function by spreading samples evenly across the entire input domain rather than concentrating on a particular area or feature e.g. domain boundaries (as is the case for many classical sampling methodologies) (Liu et al., 2018a; Sacks et al., 1989; Jin et al., 2001). Consequentially, the surrogate model

[2]'Experiment' in this sense loosely refers to the testing/evaluation of some process, physical or otherwise, accepting certain input conditions such that a corresponding output may be observed and recorded. E.g., Wind Tunnel Testing (Yondo et al., 2018).



(a) (b) (c) (d) (e)

FIGURE 2.6: Factorial DoE methods for three design variables ($n = 3$): (a) full factorial design, (b) fractional factorial design, (c) central composite design, (d) star design, and (e) Box- Behnken design.

approximation built using these samples can more reliably represent the underlying functional behaviour and guide a searching algorithm towards interesting regions (Liu et al., 2018a). Classical static sampling methods include Full Factorial Design, Fractional Factorial Design, Central Composite Design, Box-Behnken Design and Plackett-Burman Design. Some examples of classical sampling techniques are shown in Figure 2.6. Factorial designs (Giunta et al., 2003), when applied to discrete design variables (or discretised continuous variables), explore large regions of the search space. For example, full-factorial sampling involves the sampling of all possible configurations of the input vector. Given the apparent inefficiency in this approach (the number of samples increases exponentially with the number of design variables), fractional-factorial can be used as a less expensive alternative. Other factorial designs, including central composite design, star design (frequently used in combination with space mapping (Cheng et al., 2006), see Section 2.4.3.2), or Box-Behnken design can be seen in Figure 2.6 (Koziel and Yang, 2011).

Classical static sampling techniques can be summarised by a few relevant features (Liu et al., 2018a):

- Sample locations are chosen to reduce the impact of random errors on physical experiments.
- Samples are disproportionally gathered towards boundary regions, rather than internal regions.
- Samples are chosen based on a predetermined regression model. That is, they are not influenced by the resulting model output characteristics.

Another salient feature is that most often require a relatively large number of samples to represent a target function of a given dimensionality. This computational burden makes classical methods less practical for more complex scenarios.

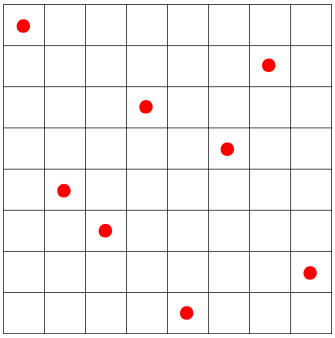


FIGURE 2.7: A *Latin Hypercube Sampling* (LHS) realisation of 8 samples in a two-dimensional design space.

Several classical methods and modifications exist aimed at reducing this burden. One of the most widely used methods in modern research is *Latin Hypercube Sampling* (LHS). In this approach, the range of each input parameter is divided into p bins. Assuming n design variables, this yields a total of $p \times n$ bins within the design space. Samples are thus randomly selected in the design space so that (i) each sample is randomly placed inside a bin, and (ii) for all one-dimensional projections of p samples and bins, there is exactly one sample in each bin (Koziel and Yang, 2011). A graphical example of this can be seen in Figure 2.7. LHS has become extremely popular in the surrogate modelling community largely due to two significant properties; (1) space-filling criteria can be met across the entire input domain, allowing the representation of underlying function structure with a minimum number of function evaluations (Yondo et al., 2018), and (2) the distribution structure exhibits a non-collapsing property, i.e. the latin-hypercube definition can still be met with a low number of samples, or if some inputs are neglected (Liu et al., 2018a). The latter implies that sample size is therefore independent of problem dimensionality, a particularly useful property when considering costly analyses (Tyan and Lee, 2019). However, if one assumes no prior knowledge regarding the target function to be approximated, it can be difficult to estimate the optimal, or even appropriate initial sample size such that computational resource is not unduly wasted. Furthermore, standard LHS may result in non-uniform distributions (for example, samples allocated along the design space diagonal satisfy both conditions (i) and (ii)). With this in consideration, numerous modifications have been proposed (Beachkofski and Grandhi, 2002; Leary et al., 2003; Ye, 1998; Palmer and Tsui, 2001) to provide more uniform sampling distributions.

Other modern static DoE methodologies include orthogonal array sampling (Queipo et al., 2005), quasi-Monte Carlo sampling (Giunta et al., 2003), and Hammersley sampling (Giunta et al., 2003). Sample distribution can be improved through the incorporation of optimisation techniques that minimise a specific non-uniformity measure. For example, $\sum_{i=1}^{p} \sum_{j=i+1}^{p} d_{ij}^{-2}$ (Leary et al., 2003) where p is the number of sampled points and d_{ij} is the Euclidean distance

between samples i and j.

It should be noted that the majority of static sampling procedures are naturally *model-independent*. That is, the choice of samples is not influenced by the model output characteristics, thus cannot benefit from any functional knowledge obtained during the surrogate modelling process. Importantly, this again suggests a certain level of inherent inefficiency with static approaches. This can be partially addressed by employing *dynamic* sampling techniques.

Dynamic sampling approaches may be classified into two categories; space-filling and adaptive (Liu et al., 2018a). Space-filling methods are usually derived from their static equivalents, but applied sequentially over a series of successive stages, thus reducing the total number of samples required for a particular function (Liu et al., 2018a). For example, to obtain m points with n dimensions, a one-shot sampling criterion needs to solve an $m \times n$ dimensional optimisation problem. This becomes more difficult with increasing dimensionality and sample size. Employing sequential sampling instead decomposes this large problem into a sequence of n-dimensional sub-problems. For example, the approach of Van Dam et al. (2007) considers a sequential *Latin Hypercube Design* (LHD) process as a set of optimisation sub-problems. In this case, the objective is some space-filling criteria an the constraints are a set of one-dimensional distance thresholds. This quasi-optimal sequential LHD strategy has since been widely studied and extended (Xiong et al., 2013; Crombecq, 2011; Liu et al., 2015). Adaptive sampling (or, active learning (Settles, 2009)) selects additional sample points using the current model approximation itself during the learning process, consequently achieving a more accurate approximation with fewer samples. Typically, such a procedure will concentrate additional samples within regions of high prediction error[3] rather than uniformly across the domain. In essence, the sampling procedure adapts to the properties of the target function (Liu et al., 2018a), allowing a globally accurate meta-model to be constructed for a minimum number of samples. Given the inherent potential for increased efficiency, adaptive sampling procedures have been studied with particular interest for modelling computationally expensive functions.

Adaptive sampling routines must compromise between two conflicting objectives; local exploitation and global exploration. Thus most current methods fall broadly into three main categories; those prioritising local accuracy, those prioritising global accuracy and those striking some more definitive balance between the two (Jiang et al., 2020; Liu et al., 2018a; Parr et al., 2012b).

[3] alternatively referred to as "interesting regions" (Liu et al., 2018a), “multimodal regions” (Li et al., 2010), "irregular regions” (Farhang-Mehr and Azarm, 2005), or similar.

2.2.1 Local Exploitation

Local exploitation implies the tendency to represent specific, localised areas of high non-linearity, discontinuity or multi-modal behaviour, normally characterised by a high prediction error. Given the limited nature of design space sampling, this is often achieved to the detriment of global prediction accuracy. It is also in practice difficult to identify these 'interesting' regions *a priori*[4], though their locations may be estimated using a number of techniques. These include using the prediction variance, cross-validation error or local gradients to estimate the true prediction error, for example. Pure exploitation-based sampling may be particularly disadvantageous given the risk of spending a significant amount of computational resource on modelling areas of the search space that are not conducive to locating the optimal design (Koziel and Yang, 2011).

In the context of surrogate based optimisation, additional samples are in practice selected through local optimisation of/using the current surrogate. For instance, the optimum computed using the current surrogate could then be evaluated using the true function and used to update the approximation. In other words, this approach intrinsically aims at the exploitation of certain regions of the design space (the neighbourhood of a first-order stationary point) (Koziel and Yang, 2011). Naturally then, this approach relies upon the surrogate being globally valid. Should this not be guaranteed however, the optimiser risks being trapped in the vicinity of unsatisfactory local solutions.

2.2.1.1 Minimising the predictor

Perhaps the simplest, most direct application of local exploitation involves first assuming that the surrogate model is in fact globally accurate (Forrester and Keane, 2009). This can be supported by various global validation metrics such as cross-validation or MSE-based approaches, ideally evaluated on an independent testing set of samples (Hastie et al., 2009). Under this assumption, the optimum of the surrogate is found in a straightforward manner via a suitably large number of calls to a global optimiser. Finally, the surrogate optimum need only be evaluated by one call to the 'true' function at this location. However, this approach is usually taken as overly optimistic, as sufficient accuracy within the region of the global optimum is most unlikely for all but the simplest problems.

The practice of introducing some sequential sampling approach based on a specified infill criterion has therefore become common practice. After each new infill point the surrogate is regenerated according to the new data set such that the interpolating model will descend towards the optimum.

[4]Though they can occasionally be identified sequentially as a consequence of global sampling.

2.2.1.2 The Trust-Region method

Minimising the predictor will at least locate a local minimum of the surrogate provided it is continuous and differentiable. However, convergence thereof may be lengthy depending upon the characteristics of the function in question.

An alternative approach that can mathematically guarantee convergence under certain conditions is the trust-region approach. Alexandrov et al. (1997) present a series of rigorous proofs of convergence from an arbitrary point for interpolating surrogates that additionally match the gradient of the objective function at the observed locations, using a trust-region method. Simple methods for matching the function gradient include the first and second order scaling approaches by Haftka (1991) and Eldred et al. (2004) respectively. These will be discussed in further detail in Section 2.3.

A generic description of a trust region approach is to begin at an arbitrary point $\mathbf{x}_0$ and search a surrogate $\hat{y}(\mathbf{x})$ in the interval $\mathbf{x}_0 \pm \delta_0$, where the trust-region δ is initialised at some user defined value. At the first iteration $m = 0$, the approximating surrogate can be initialised by fitting across 2 points; the midpoint of the input domain and a point on the boundary of the trust region. The trust region itself may then be iteratively updated as follows. First, the current prediction at iteration m is evaluated:

$$r = \frac{f(\mathbf{x}_{m-1}) - f(\mathbf{x}_m)}{f(\mathbf{x}_{m-1}) - \hat{y}(\mathbf{x}_m)} \tag{2.26}$$

The updated trust-region can then be defined:

$$\delta_m = \begin{cases} c_1 \|\mathbf{x}_m - \mathbf{x}_{m-1}\|, & \text{if } r < r_1 \\ \min\{c_2 \|\mathbf{x}_m - \mathbf{x}_{m-1}\|, \Delta\}, & \text{if } r > r_2 \\ \|\mathbf{x}_m - \mathbf{x}_{m-1}\|, & \text{otherwise} \end{cases} \tag{2.27}$$

where $c_1 < 1$ and $c_2 > 1$ are factors affecting the degree to which the trust-region contracts or expands based on the current predicted accuracy and Δ is an upper bound on the trust radius. The values r_1 and r_2 represent the lowest allowable levels of surrogate performance before the trust-region must be reduced, and conversely, the highest level of performance that must be achieved before the trust region is increased. Typical values are $r_1 = 0.10$ and $r_2 = 0.75$ (Forrester and Keane, 2009). The next sample point $\mathbf{x}_{m+1}$ is then found by minimising $\hat{y}(\mathbf{x})$ in the region $\mathbf{x}_m \pm \delta_m$. This process repeats until some stopping criteria[5] is met. Of note, the preceding discussion outlines only the fundamentals of a generic trust-region approach for surrogate based optimisation. Further detail can be found in (Alexandrov et al., 1997), with a multi-fidelity implementation reported in (Alexandrov et al., 2001).

Although exploitation-based infill criteria are attractive for local optimisation, a prediction based infill criterion may not locate the global optimum of a deceptive objective function

[5] see (Dennis Jr and Schnabel, 1996) for more information on appropriate stopping criteria.

(Forrester et al., 2008). Similarly, although a trust-region approach will find a local optimum from an arbitrary starting point, it may not find the global optimum if $\mathbf{x}_0$ is not in the global basin of attraction. Locating the global optimum in a more robust manner thus requires an element of *exploration*.

2.2.2 Global Exploration

Global exploration is concerned with representing the entire design domain and identifying regions of interest for future sampling. Regarding iterative exploration, this can be thought of as filling in the gaps between existing samples (Forrester and Keane, 2009).

A simple approach to global exploration is to employ a sequentially space-filling plan such as the Sobol sequence[6] (Sobol, 1979, 1967) (though which occasionally exhibits rather poor space filling characteristics for small samples sizes), or indeed any of the dynamic approaches described in Section 2.2. Another method is to select new points based on some min-max criterion (Morris et al., 1993). For example, if surrogate error estimates are available, infill points can be positioned at the locations of maximum estimated error. Such estimates for regressing models, with the exception of those with the modified Gaussian process variance, are of dubious merit here, however (Forrester and Keane, 2009). Gaussian process based surrogate models (such as Kriging, see Section 2.1.4) are an example global exploratory capability integrated into the surrogate construction process.

When considering the estimated error, a condition arises that the surrogate must perfectly represent each sample. That is, the estimated error is zero at each infill location, otherwise the maximum error could occur at a previously sampled point. Of note, outwith deterministic computer experiments, this condition perhaps has some leeway in the world of physical experimentation or when considering other inherently stochastic/noisy processes.

In the context of surrogate-based optimisation, pure global exploration can be a significant drain on computational resource. Time spent accurately modelling sub-optimal regions, or regions ignored by the functional domain, offers diminishing returns at best. However, global exploration is not without it's niche, particularly in applications where design space visualisation is key[7]. Another relevant example is when a globally accurate surrogate is to be implemented within a real-time control system, or even in a more complex overarching calculation (such as aerodynamic response modelling). In both cases however, the argument can still be made that the surrogate needs only be accurate in the region(s) actually used by the higher-level process, whether real-time response prediction or optimisation of some other

[6] A Sobol sequence (also referred to as an LPτ-array (Statnikov and Matusov, 2012)) is a pseudo-random sequence with the property that for all values of N, its subsequence $(x_1, \ldots x_N)$ has a low discrepancy (i.e. are approximately uniformly distributed).

[7] see (Holden, 2004).

objective function.

2.2.3 Balanced Exploitation/Exploration

The conflicting goals of local and global sampling approaches represent the fundamental trade-off in sampling methodology. Any particular approach more weighted towards local exploitation may not represent the general trends of the target function in sufficient detail, leaving the global optimum undiscovered. Alternatively, a method more weighted towards global exploration may fail to represent important, design influencing behaviour, whilst essentially wasting resource on less-important areas.

Therefore, a particularly important goal within surrogate based modelling and optimisation is to strike some appropriate balance between both local and global capabilities. This may be expressed conceptually as:

$$score = w_{local} \times local + w_{global} \times global \tag{2.28}$$

where w_{local} and w_{global} are the local exploitation and global exploration weights respectively, and are assumed to satisfy $w_{local} + w_{global} = 1$. The challenge is to determine the values of these weights such that interesting regions are both detected and adequately sampled (Liu et al., 2016).

Many current approaches assume fixed values for these weights. For example, the SEED approach (Lin et al., 2004) defines a balance factor λ but fixes it at 2. Similarly, the LOLA-Voronoi approach (Crombecq et al., 2011) assigns equal values across both weights. However, other authors balance the two conflicting aspects in a more flexible way. Liu et al. (2018a) classifies such approaches into three general categories, depicted graphically in Figure 2.8. The first (as in (Kim et al., 2009; Turner et al., 2007)) begins by fully exploring the complete domain (i.e. $w_{global} = 1$), then iteratively transfers weight to the local term as interesting regions are discovered. An issue with this approach is that during the final iterations, the global term is close or equal to 0, thus effectively ignored. If considering black-box functions, it is difficult to say whether at this point the function has indeed been sufficiently sampled such that this value is appropriate. The second approach, referred to as a 'switching' (Sasena, 2002) or 'greedy' (Singh et al., 2013) strategy, aims to address this issue. In this case, binary values for both weights are chosen based on some monitored evaluation criteria. For instance, if the uncertainty of the entire domain is below a certain threshold, the strategy switches from global to local. If a particular identified region is determined to be of acceptable accuracy, the strategy switches back to global to identify undetected interesting regions. This process continues iteratively in order to more efficiently improve the quality of the surrogate. The third method (Singh et al., 2013) is a more generalised version of the previous approach. Across successive iterations, the local and global weights are adaptively updated through comparison

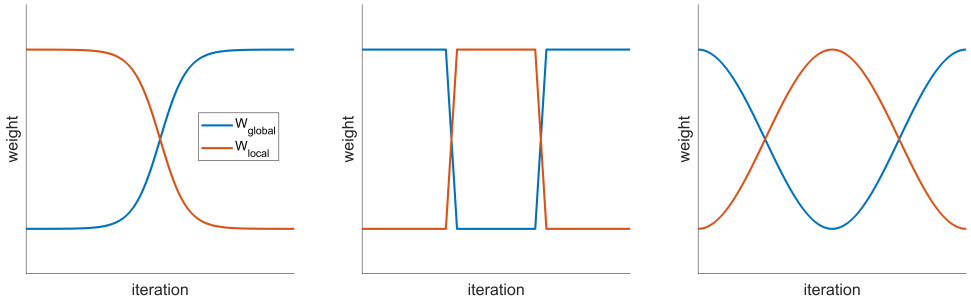


FIGURE 2.8: Three conceptual balancing strategies between local exploitation and global exploration, adapted from (Liu et al., 2018a).

of certain information (e.g. the deviation between the model errors) gathered up until that point.

Another important choice regarding adaptive sampling routines is the number of initial points selected before the procedure begins. If too few initial points are sampled, consequentially poor surrogate predictions can mislead the sampling process to concentrate on inappropriate regions (Kim et al., 2009; Ghoreyshi et al., 2009). On the other hand, too many initial samples may prohibitively waste computational resource (Crombecq et al., 2011). An appropriate choice of initial sample size, however, may not be obvious *a priori*. Such a choice may depend on such factors as the dimensionality and/or complexity of the target function, the computational budget and the characteristics of the selected surrogate modelling technique(s) (Liu et al., 2018a). An extension of this problem relevant to adaptive sampling in particular, is that for some practical instances, a pre-existing initial set of samples may not be space-filling. As an extreme example, consider the case in which all initial samples are clustered within a particular local region. Thus the adaptive sampling routine, by virtue of some appropriate balancing strategy between local and global sampling, must effectively manage and distribute the remaining computational resource accordingly (Liu et al., 2018a).

Finally, an adaptive sampling process must include, or be subject to, an appropriate stopping criterion. A common approach is to simply define a fixed computational budget assigned to sampling (e.g. a maximum number of function evaluations). Another method is to utilise the successive relative improvement (SRI) measured by the difference of some error criteria. For example, one may use the cross-validation error, jackknifing variance (Kleijnen and Van Beers, 2004) or relative absolute error (Kim et al., 2009) during the final iterations. In this case, the sampling process terminates once no further positive gain in SRI is registered.

Modern adaptive sampling techniques can be categorised according to their representation of the 'real' prediction errors (Liu et al., 2018a):

- *Variance-based* adaptive sampling asserts that regions with large prediction variances (i.e. large uncertainty) estimated by a statistical surrogate model are most likely to contain the largest prediction errors, therefore are most suitable for further sampling.
- *Query-By-Committee* (QBC) based adaptive sampling is a semi-variance based strategy that uses the response variances estimated simultaneously from several competing approximations to represent the overall prediction error.
- *Cross-Validation* (CV) based adaptive sampling directly estimates the prediction errors using the cross-validation metric.
- *Gradient-based* adaptive sampling employs local gradient information to implicitly represent the prediction errors.

2.2.3.1 Variance-based Adaptive Sampling

Variance based sampling approaches are most often integrated within Kriging-based or *Gaussian Process* (GP) surrogate modelling techniques. As such, they are considered heavily model-dependant. Variance-based methods generally consider the prediction variance either as obtained from the constructed surrogate or as a practical estimation of the associated (yet assumed unknown) prediction error (Liu et al., 2018a). Intuitively, most variance-based methods aim to sample areas of the domain with higher values of MSE (see Maximising the Mean Square Error, or MMSE) (Jin et al., 2001).

$$\mathbf{x}_{new} = \arg\max_{\mathbf{x} \in D} \hat{\sigma}(\mathbf{x}) \tag{2.29}$$

Particularly within the statistics community, where Kriging and other GP based meta-modelling techniques are most prevalent, many alternative variance-based sampling methods have been proposed. These include the *Integrated Mean Square Error* (IMSE) criterion (Sacks et al., 1989) and *Maximum Entropy* (ME) Shewry and Wynn (1987). Morris et al. (1993) further combined the ME criterion with the known first derivatives to improve the sampling performance. Jones (Jones, 2001) categorises statistical infill criteria into two generalised areas; one-stage and two-stage methods. Two-stage methods (simpler and more common) first fit the surrogate to the available data, then calculate the infill criterion using this model. One-stage approaches instead operate in reverse, generating the surrogate based on the infill criterion.

A simple two-stage approach for balancing exploitation and exploration, respectively with regards to the prediction $\hat{y}(\mathbf{x})$ and it's MSE $s^2(\mathbf{x})$, is to *minimise a statistical lower bound*:

$$LB(\mathbf{x}) = \hat{y}(\mathbf{x}) - As(\mathbf{x}) \tag{2.30}$$

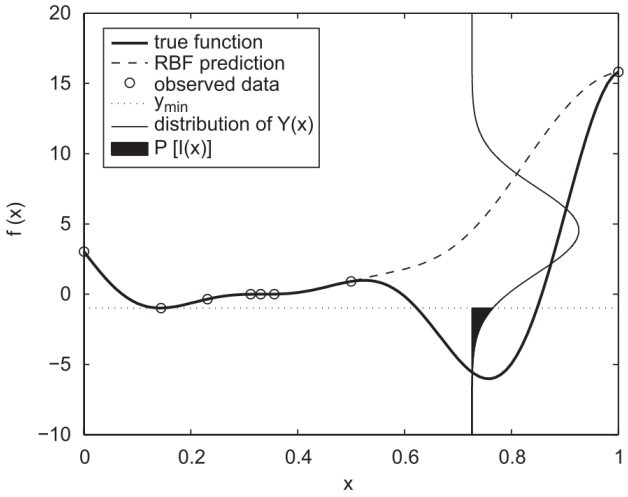


FIGURE 2.9: A graphical interpretation of the probability of improvement (Forrester and Keane, 2009).

where the constant A defines the balance between exploitation and exploration. That is, as $A \to 0, LB(\mathbf{x}) \to \hat{y}(\mathbf{x})$, representing pure exploitation, and as $A \to \infty$, the effect of $\hat{y}(\mathbf{x})$ becomes negligible. In this case, minimising $LB(\mathbf{x})$ is therefore equivalent to maximising $s(\mathbf{x})$, thus representing pure exploration. An issue with this approach is that it can be difficult to estimate an appropriate value for A *a priori* (Forrester and Keane, 2009). A value suitable for one function may cause over exploitation in another. To accommodate this risk, the designer may apply this method using a series of values for A, finally selecting infill points within any resulting clusters of local minima to Eq. (2.30).

An alternative method is to consider $\hat{y}(\mathbf{x})$ as a realisation of a random variable. By doing so, the *probability of improvement*, where an improvement is defined $I = y_{min} - Y(\mathbf{x})$, can be calculated based upon the best observed objective function value so far, y_{min}:

$$P[I(\mathbf{x})] = \frac{1}{s\sqrt{2\pi}} \int_{-\infty}^{0} \mathrm{e}^{(-(I-\hat{y}(\mathbf{x}))^2)/2s^2} \mathrm{d}I \tag{2.31}$$

Figure 2.9 shows a graphical interpretation of the probability of improvement for a Kriging approximation model of a one-dimensional function (Forrester and Keane, 2009). In this example, the vertical Gaussian distribution with variance $s^2(\mathbf{x})$ centred around $\hat{y}(\mathbf{x})$ represents the uncertainty in the prediction $\hat{y}(\mathbf{x})$ and the part of the distribution below the horizontal dotted line indicates the possibility of improving on the best observed value. The probability of improvement is therefore the area enclosed by the Gaussian distribution below the best observed value so far.

Rather than basing decision on where improvement is likely to be, the actual value of *Expected Improvement* (EI) can be used instead, calculated from the mean $\hat{y}(\mathbf{x})$ and variances $s^2(\mathbf{x})$.

The EI is defined according to:

$$E[I(\mathbf{x})] = \begin{cases} (y_{min} - \hat{y}(\mathbf{x}))\Phi(\frac{y_{min}-\hat{y}(\mathbf{x})}{s(\mathbf{x})}) + s\phi(\frac{y_{min}-\hat{y}(\mathbf{x})}{s(\mathbf{x})}) & \text{if } s > 0 \\ 0 & \text{if } s = 0 \end{cases} \tag{2.32}$$

where $\Phi(.)$ and $\phi(.)$ are the normal cumulative distribution function and the probability density function respectively.

Maximising $E[I(\mathbf{x})]$ has become a very popular tool for locating the global optimum (Forrester and Keane, 2009). However, the calculation of $E[I(\mathbf{x})]$ and $P[I(\mathbf{x})]$ usually assumes that the surrogate model parameters (e.g. the correlation parameters for Kriging models) have been estimated accurately based on the observed data[8]. However, in situations where the initial sampling proves particularly unlucky or sparse, and/or the target function includes a deceptively positioned optimum, model parameters are therefore susceptible to suboptimal assignments, potentially compromising the efficacy of the resulting prediction. To illustrate in an extreme sense, Wang et al. (2016) find EI-based infill to give markedly different results when applied with different surrogate modelling techniques approximating the same underlying function. For Kriging surrogates, a common strategy is to estimate the correlation parameters with a maximum likelihood method, then use those estimated parameters directly in the calculation of the Kriging predictor Jones et al. (1998). This issue may also be tackled by employing Bayesian Inference, as described in (Santner et al., 2003).

One-stage approaches attempt to address some of the various problems associated with error-based two-stage approaches. A decoupling of the basic principles introduced via two-stage methods is adopted, where instead of using the surrogate to locate the optimum, the current optimum is instead used to "locate" or *learn* the surrogate. For example, the principle behind the one-stage *goal seeking* approach is to simply search for points at which an objective function value *could* exist, regardless of value expectations or probability for improvement. This method may be used to estimate an appropriate value of the objective function, or search for a specific level of improvement (even without knowing *a priori* whether such an improvement is possible (Jones, 2001)).

Such methods are known as *Maximum Likelihood Estimators* (MLEs), another example of which could be the Kriging predictor (see Section 2.1.4). Rather than estimating the value of $\hat{y}(\mathbf{x})$ at a specific sample point $\mathbf{x}$, it is instead assumed that the surrogate passes through a goal y^g in addition to the sampled data. The input value that best fits this assumption $\mathbf{x}^g$ can be found by maximising the *conditional ln-likelihood*:

$$-\frac{n}{2}\ln(2\pi) - \frac{n}{2}(\hat{\sigma}^2) - \frac{1}{2}|\mathbf{C}| - \frac{(\mathbf{y}-\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{y}-\mathbf{m})^T}{2\hat{\sigma}^2} \tag{2.33}$$

[8]This further assumes that for most practical purposes, the correct choice of model parameters is unclear *a priori*.

where

$$\mathbf{m} = \mathbf{1}\mu + \boldsymbol{\psi}(\hat{y}^g - \mu) \tag{2.34}$$

$$\mathbf{C} = \boldsymbol{\Psi} - \boldsymbol{\psi}\boldsymbol{\psi}^T \tag{2.35}$$

and $\boldsymbol{\psi}$ represents the vector of correlations between $\mathbf{x}^g$ and the observed data $\mathbf{y}$. Eq (2.33) is maximised by varying $\mathbf{x}^g$ and the surrogate model parameters (e.g. if using a Kriging model, the upper/lower bounds of θ). Maximising (2.33) can also be approached in a similar manner to selecting the surrogate model parameters. For example, substituting for the MLE $\hat{\sigma}^2$ gives the concentrated conditional ln-likelihood (see Section 2.1.4):

$$-\frac{n}{2}\ln(\hat{\sigma}^2) - \frac{1}{2}\ln|\mathbf{C}| \tag{2.36}$$

The goal $\mathbf{x}^g$ cannot however be purely arbitrary. Should an overly optimistic value be selected, the method will exhibit excessive exploration, given the equally low likelihood in many areas of the search domain. Conversely, an overly pessimistic goal will result in a primarily local search, though the goal will likely be found quickly (which may well be an acceptable outcome) (Forrester and Keane, 2009).

In cases where an appropriate goal estimate cannot be made, an alternative approach is to instead calculate a lower bound based on the conditional likelihood. Rather than locating the $\mathbf{x}$ that gives the highest likelihood conditional upon $\hat{y}(\mathbf{x})$ passing through a goal, the $\mathbf{x}$ that minimises $\hat{y}(\mathbf{x})$ subject to the conditional likelihood not being too low is found instead (Jones, 2001). Considering points that are hypothesised to belong to the sample data but are not actually observed as such (i.e. imputed), the ratio of the conditional likelihood of $\hat{y}$, L_0, to the conditional likelihood L_{cond} of the surrogate model passing through the imputed point can be compared with the χ^2 distribution. Whether to keep the value of the imputed point can be decided according to whether the following condition is satisfied:

$$2\ln\frac{L_0}{L_{cond}} < \chi^2_{critical}(\text{limit,dof}) \tag{2.37}$$

where the critical value of χ^2 depends on the desired confidence limit (e.g. > 0.95) and the number of model parameters (acting as degrees of freedom). Using this ratio test, a lower confidence bound for the surrogate prediction can be systematically determined, where the minimum of this bound can be used as an infill criterion.

One-stage approaches can still exhibit a number of disadvantages, however. For example, considering a conditional bound approach applied in cooperation with a Kriging surrogate model, the variables $\mathbf{x}$, y^h, σ and the Kriging model parameters must all be varied to minimise the lower bound. The extra $3k + 1$ parameters (in addition to the included sample points) can result in a process far more computationally expensive than a standard two-stage approach

(Forrester and Keane, 2009). If the underlying function is expensive, as in the case of high-fidelity CFD or non-linear FEA, this extra cost may nonetheless be worthwhile (Forrester and Keane, 2009).

It is notable that the variance-based sampling approaches briefly described above rely heavily on the inner properties of the statistical surrogate model(s) employed. In consequence, such approaches are difficult to extend to alternative surrogate modelling techniques.

2.2.3.2 Query-By-Committee based

Employing the *Query-By-Committee* (QBC) methodology as an adaptive sampling technique originated within the machine learning community for application to classification problems (Seung et al., 1992; Freund et al., 1997). A 'committee' in this sense is a collection $\mathcal{C} = \{\hat{f}_i\}_{1 \leq i \leq t}$ of alternative surrogate models, each representing a competing approximation of the underlying function of interest. Candidate points for future sampling are evaluated using each member of the committee in turn, with the single point exhibiting the largest resulting disagreement between members selected as the new candidate for sampling. This level of disagreement is usually defined as the *"variance of the predictions of t committee members"* (Krogh and Vedelsby, 1995). A simple form can be written as:

$$\hat{\sigma}^2_{QBC}(\mathbf{x}) = \frac{1}{t} \sum_{i=1}^{t} \left(\hat{f}_i(\mathbf{x}) - \bar{\hat{f}}(\mathbf{x}) \right)^2 \tag{2.38}$$

where $\bar{\hat{f}}(\mathbf{x}) = \sum_{i=1}^{t} \hat{f}_i(\mathbf{x})$ is the average value of t predictions at point $\mathbf{x}$. An advantage here over variance-based sampling is that QBC is fully model-independent, with little or no restriction on the type of surrogate model that may be employed.

A critical prerequisite of QBC sampling is that the included surrogate modelling techniques must have at least *some* disagreement across the input domain, i.e. *QBC diversity*. From this requirement, QBC-based approaches can be further split into two groups, *homogeneous* and *heterogeneous*. *Homogeneous* refers to the case where committee members are constructed using the same modelling technique. To achieve QBC diversity, multiple models can, for example, be generated each upon a different sample set $\{\mathbf{X}_i\}_{1 \leq i \leq t}$ extracted from the original sample set $\mathbf{X}$. *Heterogeneous* refers to the case where instead, committee members are constructed using different modelling techniques. In this case, QBC diversity can be achieved by building each approximation upon the same sample set. For example, a committee containing models generated using Kriging, RBF, SVR etc. Alternatively, Kriging models built using different basis functions.

The QBC technique has been found to exhibit 'greedy' local exploitation (Liu et al., 2018a). If unconstrained, QBC tends to cluster new samples within regions of large prediction disagreement (Li et al., 2009). To compensate, some authors employ the use of a distance term

d to facilitate global exploration, discarding points that lie too close to previously sampled points (Hendrickx and Dhaene, 2005; Li et al., 2009; Eason and Cremaschi, 2014). A QBC based adaptive sampling process can thus be generally expressed as an iterative selection of new points according to:

$$\mathbf{x} = \arg\max_{x \in D} Score\left(\hat{\sigma}(\mathbf{x}), d(\mathbf{x})\right) \tag{2.39}$$

where the QBC prediction variance $\hat{\sigma}$ controls the level of local exploitation, whilst the distance term d determines the level of global exploration. *"Score"* in this case could be any relevant function defining the superiority of one sample over another. Typically, $Score$ takes the form of a normalised weighting between the inputs (in this case $\hat{\sigma}$ and d), such that potential samples may be ranked accordingly.

2.2.3.3 Cross-Validation based

Model-dependant adaptive sampling approaches utilise specific model properties to determine optimal future sampling, which thus prohibits their use in conjunction with other surrogate modelling approaches. A previously seen example is the prediction variance calculated during a Kriging-based surrogate modelling process used to drive a variance based sampling procedure (see Section 2.2.3.1).

A well-known model-independent approach that requires only the information gained from existing sample points is the *Cross-Validation* (CV) approach. A common example is to employ the *Leave-One-Out* (LOO) CV metric. The point $\mathbf{x}_i$ is selected as a validation point with the remaining $m-1$ points used to fit the surrogate model $\hat{f}^i$ used to obtain the CV error:

$$\hat{e}(\mathbf{x}_i) = \left| f(\mathbf{x}_i) - \hat{f}^i(\mathbf{x}_i) \right|, \quad 1 \leq i \leq m \tag{2.40}$$

The observations at all m points are used to obtain the generalised mean square CV error (GMSE), where $GMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m} \hat{e}(\mathbf{x}_i)^2}$. A small value of $GMSE$ represents a favourable (more accurate) model.

The CV error has been found to somewhat correlate with the local prediction error (Liu et al., 2018a). A small value of $\hat{e}(\mathbf{x}_i)$ suggests an insensitivity to the loss of $\mathbf{x}_i$. In other words, the surrogate has been well fit around $\mathbf{x}_i$. Conversely, a large value of $\hat{e}(\mathbf{x}_i)$ suggests that the region surrounding $\mathbf{x}_i$ does not contain enough sample points such that the loss of $\mathbf{x}_i$ does not significantly affect the prediction quality of the model.

Dependent on the method of CV error estimation, CV-based sampling can be categorised as either *discrete* or *continuous*. *Continuous* CV-based adaptive sampling uses the estimated CV errors to identify interesting regions for guiding the local exploitation of the sampling process (Li et al., 2010; Aute et al., 2013). However, directly maximising the CV-based criteria encourages the selection process to cluster points around specific regions (i.e. too heavily

weighted towards local exploitation) (Jin et al., 2001; Li et al., 2010). Accordingly, a distance-based space filling term $d(\mathbf{x})$ is commonly added to aid global exploration. The resulting auxiliary optimisation problem can be expressed as follows:

$$\begin{aligned} \mathbf{x}_{new} &= \arg\max \hat{f}_e(\mathbf{x}) \hookleftarrow \mathbf{X}, \quad \hat{e}(\mathbf{x}_1) \dots \hat{e}(\mathbf{x}_m) \\ s.t. \quad & ||\mathbf{x} - \mathbf{x}_i|| \geq d, \quad \mathbf{x} \in D \end{aligned} \tag{2.41}$$

The distance term d is often difficult to estimate *a priori* (Xu et al., 2014). Too small, and the clustering phenomenon cannot be overcome. Too large, points are selected more evenly across the entire domain, thus interesting local regions remain undiscovered. Li et al. (2010) suggest the average of the minimal distance of each observed point be used as the threshold d value. Alternatively, Aute et al. (2013) suggest the maximum of the minimum distance between each observed point be taken as the d value instead, giving more of an emphasis on global exploration. *Discrete* CV-based adaptive sampling segments the design space into partitions to facilitate cooperation with discrete CV error calculations. By doing so, the overall computational cost of the sampling procedure is reduced (in comparison to continuous methods) and retains an implicit global exploration ability via dynamic shift of the regions with the largest CV-error (Xu et al., 2014). The approach presented by Xu et al. (2014) divides the domain into Voronoi cells based on the observed points using the Voronoi diagram algorithm (Aurenhammer, 1991). The CV error of each cell is calculated, with the cell exhibiting the largest selected as the "sensitive" cell. A space-filling point is thus selected from within the sensitive cell. This ultimately promotes local exploitation whilst avoiding any local clustering. The general form of a CV-based adaptive sampling routine can be expressed as:

$$\mathbf{x}_{new} = \arg\max_{x \in D} Score\left(\hat{e}(\mathbf{x}), d(\mathbf{x})\right) \tag{2.42}$$

where the values of the CV error control the local exploitation, with global exploration facilitated by the the distance constraints, the areas of partitioned polygons, or the implicit error-pursuing mechanism (Liu et al., 2018a). Of note, the design space D may be either continuous or discrete.

2.2.3.4 Gradient-based

Interesting regions of a function response are most often characterised by highly non-linear, multi-modal or discontinuous behaviour. For these reasons, amongst others, they are difficult to approximate without an appropriate amount of localised sampling, and thus typically result in higher prediction errors. It is easy to imagine therefore that the locations of such regions can be indicated by high gradient values of the target function. Gradient information can thus be used as a driver for adaptive sampling.

A simple and practical approach is to use the gradient information provided by the surrogate model itself. For many conventional surrogate modelling techniques, e.g., RBF modelling with

smooth kernel functions (see Section 2.1.2), gradient information at any point can be simply and inexpensively derived. The general form of a gradient-based adaptive sampling methods can be written as (Liu et al., 2018a):

$$\mathbf{x}_{new} = \arg\max_{x \in D} Score\left(\hat{g}(\mathbf{x}), d(\mathbf{x})\right) \tag{2.43}$$

where $\hat{g}$ represents the estimated first/second-order gradients based on some specific models or some other geometric features (thus acts as the local exploitation term) and the distance criteria d acts as the global exploration term.

Yao et al. (2009) suggested that despite the fact that a surrogate is only an approximation of the target function, it may nonetheless describe the function trends to a reasonable extent, thus can be used to inform the sampling process. Deschrijver et al. (2011) counter-argue instead that gradient-based sampling approaches rely too heavily on the quality of the surrogate approximation and that accordingly, more generic, model-independent approaches are needed. To address this issue, the LOLA-Voronoi algorithm (Crombecq et al., 2011) selects a new point according to:

$$\mathbf{x}_{new} = \arg\max_{x \in D} G(\mathbf{x}) + V(\mathbf{x}) \tag{2.44}$$

where the gradient term $G(\mathbf{x})$, measured by a local linear approximation across neighbouring points, is used to inform local exploitation. The sampling density term $V(\mathbf{x})$, measured using the areas of the partitioned Voronoi-cells, is used to inform global exploration. This approach has been incorporated into the *MATLAB Surrogate Modelling* (SUMO) toolbox (Gorissen et al., 2010) for the adaptive refinement of problem surrogates.

2.2.4 Parallel Infill

The preceding discussion has primarily assumed that infill points are selected in serial (i.e. one point per iteration). This is also referred to as single-selection. Most modern adaptive sampling algorithms adopt single-selection, as their infill criterion is usually based upon selecting the global optimum of an auxiliary optimisation problem (Liu et al., 2018a). If multiple points are selected instead, the sampling algorithm can be easily integrated into a parallel computing environment to improve computational efficiency, particularly meaningful for expensive simulation-based problems. This is known as batch-selection.

Many serial infill criteria can in fact be converted to allow parallel updates. A straightforward example is to employ a searching method that can locate a number of alternative minima, e.g. multi-start hill-climbers or a genetic algorithm with clustering. These minima can thus be simultaneously evaluated to utilise available parallel computing capabilities, with those meeting the specified infill criteria added as infill points (Sóbester et al., 2004).

A more common method is to employ a simple "q-best" strategy. In this case, a collection of candidate sample points are ranked according to some criterion, with the q-best points selected for evaluation. This approach is often inappropriate, however, due to any overlapping information represented by the new points not being considered. This may lead to sample clustering. Such strategies are then usually employed alongside two specific batch selection criteria: *informative* and *diverse*. The former states that points should be sampled within interesting regions, whereas the latter states that points should lie further away from previously observed areas, in order to maximise the information gained about the target function (Liu et al., 2018a). Applied to adaptive sampling, these criteria are, in essence, a local/global balancing extended to batch-selection.

With batch-sampling in general, it is difficult to specify how many infill points will be selected at a given iteration and thus available resource may not all be appropriately used at any one time. The method proposed by Schonlau (1997) searches the infill criterion to find it's global optimum then temporarily includes the surrogate-predicted value at this point into the model training data. The surrogate is then reconstructed and the infill criterion is searched again. This process repeats until the desired number of infill points has been reached. Each point is then evaluated and added to the data set in place of the temporary surrogate predictions. This method makes effective use of parallel computing resources, though samples may not be as well placed as a sequential scheme (Forrester and Keane, 2009).

Importantly, working with surrogate models (including sampling and training) often represents a significant bottleneck in a parallel computing environment (Forrester and Keane, 2009). Even assuming sufficient processing power as to simultaneously evaluate all points selected during the initial sampling stage, each must be used to generate and evaluate the target function surrogate before the parallel calculation of the subsequent infill points can be performed. This then fundamentally limits the amount of resource that can be allocated to such surrogate generation and searching processes.

In summary, single-selection is often the most practical and informative approach, since the searching landscape is updated at every iteration. However, for cases in which the target function is expensive, batch selection offers a meaningful route to exploit parallelisation to reduce the overall computational cost. Of note, the choice of batch size q is largely determined simply by the available computational resource (e.g. the number of processors) rather than the sampling criterion itself (Liu et al., 2018a).

2.2.5 Multi-Response Sampling

Adaptive sampling methods can also be classified according to the number of unique responses that can be simultaneously addressed.

Single-response adaptive sampling methods, comprising the majority of methods currently employed (Liu et al., 2018a), differ primarily in their local exploitation criteria, or more specifically, the way they estimate the local prediction error. For example, variance-based methods (Section 2.2.3.1) assert that regions of high prediction uncertainty (high prediction variance) are more likely to contain areas of high prediction error. An extension of this approach is *Query-By-Committee* (QBC) (Section 2.2.3.2), where the response variance is calculated simultaneously considering multiple, competing approximations. CV based approaches (Section 2.2.3.3) estimate prediction error using the cross-validation metric, usually cheap to implement but suffers with increasing sample size and/or expensive functions. Finally, gradient-based methods (Section 2.2.3.4) employ local gradient information to implicitly represent the prediction errors instead, thus informing the selection of additional points.

In practice however, many disciplines include system models that produce multiple correlated/uncorrelated output responses for a given input. An example is aerodynamic response simulation, which may for a given specification of geometry and flow conditions produce both the coefficients of lift C_L and drag C_D.

Some traditional space-filling sampling approaches (such as LHS, maximin, see Section 2.2) can be easily extended to handle the multi-response definition since they determine the sample distributions without considering the characteristics of the output responses (Liu et al., 2018a) (i.e. are model independent). This modification is more difficult when considering adaptive approaches however, as adaptation to the characteristics of one response does not guarantee applicability to another, particularly if uncorrelated. In fact, for the consequential branch of adaptive methods for multi-response systems in global surrogate modelling, responses are generally assumed uncorrelated to begin with. That is, k output responses of a multi-response system $\mathcal{F} = \{f_1, \ldots, f_k\}$ may be simultaneously considered, but their correlations are ignored. Considering each response individually in this manner requires a specification of weights to assign relative importance to each. This in turn converts the multi-response problem into a series of single-response sampling problems. However, such information on the relative importance of individual responses may in practice be difficult to estimate *a priori* (Liu et al., 2018a).

Various methods have been proposed to address this problem. Reichart et al. (2008) presents two strategies, *alternating selection* and *rank combination*. Alternating selection simply samples each response in succession according to some (possibly separate) sampling criteria. That is, for the i-th iteration, a new point is sampled according to one criterion, and for the $(i+1)$-th iteration, a second alternative criterion is used instead. Rank combination instead introduces a score for each candidate point relative to each response, where the resulting sum represents the complete score for that particular point. The candidate point with the highest score is sampled. Generalised multi-response sampling frameworks often adopt the rank-combination

strategy, most commonly in one of three primary manners:

- Evaluate the final score of each candidate point via the sum (or product) of the scores for all k responses (Rosenbaum, 2013).
- Use the maximum/minimum of the scores for each of the k responses to evaluate the complete score for a particular candidate point (Crombecq, 2011; Lovison and Rigoni, 2010). This method is found particularly useful for systems of many responses of varying complexity/form (Liu et al., 2018a).
- Combine the two previous methods. That is, a fusion between a weighted-sum and a maximum value criterion (Liu et al., 2018c).

These types of general multi-response adaptive sampling criteria are typically considered *compromise* approaches, i.e. offering a reasonable solution considering all responses together, though suiting none of them well individually. The wide applicability of these approaches can perhaps be attributed to their lack of assumptions regarding the responses themselves, thus allowing application in a general manner.

One major drawback however is that due to this lack of assumption, any correlation between responses (which may indeed be reasonable to assume for many real-world systems) is effectively ignored. Thus, newly obtained samples must be evaluated according to each response individually, which may represent a significant drain on computing resources should they be particularly expensive. To take advantage of existing response correlations, a *symmetric* approach to multi-response adaptive sampling can be adopted.

2.2.5.1 Symmetric multi-response adaptive sampling

In real world systems, some level of correlation between the responses of a multi-response system is expected. Response correlations can be utilised to model the k responses as part of a joined, interconnected process, ultimately improving the prediction quality compared with modelling them each individually.

A common approach is to attribute an equal level of importance to each response. This is referred to as *symmetric* modelling. Examples from the literature include ANNs extended to support multi-response systems by specifying how the system inputs are shared across all responses (Caruana, 1994), the general stacked single-target framework and ensemble of regressor chains (Spyromitros-Xioufis et al., 2016) and the multi-response GP framework[9] where response correlations are captured using cross-covariance functions built using a linear

[9] also known as *Multi-Task Gaussian Process* (MTGP).

model of coregionalisation or convolved process (Alvarez and Lawrence, 2009). This latter approach has seen wide use given the integrated generation of prediction mean and variance (under a Bayesian framework).

Despite the development of symmetric modelling approaches, branched research into symmetric sampling for multi-response systems has been less active. A symmetric multi-response adaptive sampling routine should address two fundamental problems:

- *How* to choose the location within the target domain that will yield the most information relative to both responses at a given iteration.
- *Which* responses should be used to evaluate each new sample.

The first issue is currently tackled primarily by extending current sampling methods (particularly variance-based) under the multi-response modelling definition. For example, Li et al. (2006) extend the maximum entropy criterion by using the stacking strategy. In other words, the output responses are treated as extra inputs. Unfortunately, and similarly as for the generalised multi-response methods, such approaches typically evaluate each new sample point using each response, thus efficiency savings are minimal.

An approach that attempts to address the "which" issue involves evaluating the new point(s) against the single response that assigns the highest score. An example here is the entropy-based approach by Zhang et al. (2016), developed under the sparse convolved MTGP framework (Alvarez and Lawrence, 2009) for handling large data-sets. This approach has been shown to perform well in various numerical scenarios. Thanks to the sparse model structure, an exploitation-exploration trade-off expression is introduced to query the most informative point, then to select the proper response to evaluate it.

2.2.5.2 Asymmetric multi-response adaptive sampling

Should there be a clear primary response (i.e. with a definitively higher relative importance than the remaining $k-1$ responses), an *asymmetric* approach may be more suitable. The goal of asymmetric multi-response adaptive sampling is to enhance the modelling performance of the primary response with recourse given to the secondary responses (Liu et al., 2018a). The asymmetric multi-response problem is alternatively referred to as *transfer learning* (Pan and Yang, 2010) within the machine learning community. Within the engineering community, this class of problem is typically confined to research areas encompassing (for example) *Multi-Fidelity Models* (MFMs), or data fusion (Fernández-Godino et al., 2016; Peherstorfer et al., 2016).

Specifically considering the multi-fidelity case, the assumption is thus that for a given target function, there exists several correlated approximations, each with a distinct level of modelling fidelity. The most widely studied examples of such are bi-level cases, where the 'high-fidelity' simulations provides the accurate predictions of the real system response but at a prohibitive computational cost, and the 'low-fidelity' simulation offers coarser response predictions but is available at a significantly reduced cost. The high-level goal of multi-fidelity modelling therefore is to leverage the knowledge gained from the the cheap approximation(s) to enhance the modelling results of the high-fidelity simulator. Similarly to symmetric approaches, methods for multi-fidelity modelling have been the subject of disproportionate research relative to their associated adaptive sampling routines (Liu et al., 2018a).

The fundamental issues facing asymmetric modelling are largely the same as for the symmetric case. That is, *how* to select the sample location and *which* responses to evaluate them with. In the case of multi-fidelity modelling, the choice of response becomes of critical importance given the assumption that the high-fidelity simulation is significantly more expensive to evaluate than the low-fidelity equivalent(s).

Space filling sampling approaches e.g., optimal LHD (Jin et al., 2001) are currently popular for multi-fidelity modelling problems. A common requirement is that the high-fidelity points should be a subset of the low-fidelity points. This allows both the simple estimation of modelling error and the high/low-fidelity model hyper-parameters individually (Kennedy and O'Hagan, 2000). Accordingly, the nested sampling approach (Qian et al., 2009) and the nearest neighbour sampling (Gratiet and Garnier, 2014) have seen wide application. However, as per the earlier discussion (see Section 2.2), these space-filling approaches are generally employed in a one-shot fashion, thus cannot take advantage of information gained during the sampling process.

Several studies have investigated adaptive sampling methods for multi-fidelity modelling problems. Examples include the asymmetric nested lattice approach proposed by Qian et al. (2014), which selects samples based on the inputs that determine greater variability in the response (though does not consider any distinction between low- and high-fidelity responses). The discrepancy based approach presented by Han et al. (2010) selects points at locations of maximum disagreement (or *discrepancy*) between the various levels of fidelity on the assumption that the multi-fidelity and high-fidelity responses will eventually converge given sufficient samples. As a purely error-based strategy, however, this method may lead to excessive local exploitation. As an attempt to avoid this consequence, Zhou et al. (2016) presented a modification which includes an additional distance based penalty term to discourage excessive local exploitation.

Most current asymmetric approaches for multi-fidelity modelling evaluate the selected point

with the high-fidelity response, given a general assumption that the low-fidelity approximation is well fit. This assumption may not necessarily remain valid in practice, especially for target functions of high dimensionality. Furthermore, an efficient multi-fidelity modelling approach, in line with the definition of asymmetric sampling, must consider that sampling the low-fidelity approximation may itself enhance the overall predictive quality of the resulting multi-fidelity model. This mentality can be seen in the Co-Kriging based sampling approach of (Le Gratiet and Cannamela, 2015). This method employs the LOO CV error to actively adjust the original prediction variance such that interesting regions may be more frequently sampled. Rather than select a particular response to evaluate each new point, this method utilises a recursive formulation of the Co-Kriging procedure (Le Gratiet, 2013), which offers the contribution of each fidelity level to the final approximation. By doing this, the relative computational costs of each fidelity level are actively taken into account during the sampling process.

The preceding discussions may be summarised as follows; In most practical implementations involving uncorrelated responses, a general multi-response adaptive sampling approach is suitable. For correlated responses, or those with a clear hierarchy of importance/dependence, the symmetric and asymmetric multi-response adaptive sampling approaches should be considered.

2.2.6 Multi-Objective Sampling

Another complication is the case where the problem solution must be selected according to more than one objective function. Similarly to the case of multi-response surrogate modelling, the sample point most suitable for selection by an adaptive sampling approach may not be the same across all objectives. This issue becomes more pronounced when considering problems where the target function is a constituent component of the system as a whole, rather than the objective function itself. In these cases, the existence of multiple objectives implies the possibility that different areas of the target function domain may be more heavily relied upon to obtain solutions relative to the included objectives.

A common strategy is to evaluate an initial sampling plan (as per the single objective case) but construct separate surrogate models relative to each objective. An example of this is the multi-objective variant of the well known NSGA-II algorithm. For this approach, the search is simply applied to the resulting set of surrogates and used to produce a Pareto set of solutions. Each solution is then used to generate a set of infill points, which are subsequently evaluated and appended to the training data used to tune/refine the surrogates. Although generally effective, this approach is unable to explicitly manage the trade-off between exploitation and exploration when constructing the surrogates, similarly to when employing a prediction-based infill criterion for single objective searching (Forrester and Keane, 2009). However, distance-based (or crowding/niching) measures have been seen to help mitigate this issue.

2.2.6.1 Multi-Objective Expected Improvement

Whilst in principle it is possible to treat the multi-objective sampling problem as a special case of the similarly formed multi-response sampling problem, the following, adapted from Forrester and Keane (2009), details a method with which to consider multiple objectives more explicitly within the context of statistical[10] methods. This demonstration is confined to the case of constructing surrogate(s) to represent specifically the objective functions themselves.

Consider the minimisation of two objective functions $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$, which may be evaluated to yield the outputs y_1 and y_2 respectively. For simplicity, assume that the input vector $\mathbf{x}$ consists of one single design variable. By evaluating a sample set (or sample *plan*) X, the responses $\mathbf{y}_1$ and $\mathbf{y}_2$ are observed. Thus the multi-objective Pareto set of m unique solutions may be identified as:

$$\begin{aligned}\mathbf{y}^*_{1,2} =&\{[y_1^{(1)*}(x^{(1)*}), y_2^{(1)*}(x^{(1)*})], [y_1^{(2)*}(x^{(2)*}), y_2^{(2)*}(x^{(2)*})],\\ &\ldots, [y_1^{(m)*}(x^{(m)*}), y_2^{(m)*}(x^{(m)*})]\}\end{aligned} \tag{2.45}$$

where the superscript $*$ indicates a non-dominated solution. Next, consider the construction of two GP-based models (e.g. Kriging models) to represent these 'samples' in objective space (i.e. modelling each dimension of the Pareto set with its own surrogate). These models are assumed independent (though correlated models could be constructed using, for example, co-Kriging), and according to the definitions of GP-based models (see Section 2.1.4), have means $\hat{y}_1(x)$, $\hat{y}_2(x)$ and variances $s_1^2(x)$, $s_2^2(x)$ respectively. A two-dimensional Gaussian *Probability Density Function* (PDF) can be constructed according to:

$$\begin{aligned}\phi(Y_1, Y_2) =&\frac{1}{s_1(x)\sqrt{2\pi}} \exp\left[-\frac{(Y_1(x) - \hat{y}_1(x))^2}{2s_1^2(x)}\right]\\ &\times \frac{1}{s_2(x)\sqrt{2\pi}} \exp\left[-\frac{(Y_2(x) - \hat{y}_2(x))^2}{2s_2^2(x)}\right]\end{aligned} \tag{2.46}$$

where x is explicitly defined as the location at which an estimate of the objective functions is being sought.

A general criteria by which to select a new sample could be based on the likelihood that the newly evaluated point will be good enough to become a member of the current Pareto set and, if comparing competing points, which will offer the most improvement. First, consider the probability that a new solution at point x will dominate one single member of the current Pareto set, $[y_1^{(1)*}, y_2^{(1)*}]$. For the bi-objective case, this will be the case if the the improvement is registered relative to objective 1, objective 2 or both objectives 1 and 2. Simply the, the probability that the new sample will yield an improvement can be expressed as:

$$P[Y_1(x) < y_1^{(i)} \cap Y_2(x) < y_2^{(i)}] \tag{2.47}$$

[10] variance-based, see Section 2.2.3.1.

which may be calculated by integrating the volume under the associated joint PDF. Secondly, consider the probability that the new point improves the entire Pareto set. This probability may be calculated similarly to the previous case, only modifying the area over which the integration is carried out. The following formulation considers only points which dominate the Pareto set (for alternatives considering Pareto set augmentation, the reader is referred to (Keane, 2006)).

$$\begin{aligned} P[Y_1(x) < \mathbf{y}_1^* \cap Y_2(x) < \mathbf{y}_2^*] = &\int_{-\infty}^{y_1^{*(1)}} \int_{-\infty}^{\infty} Y_1 \phi(Y_1, Y_2) \mathrm{d}Y_2 \mathrm{d}Y_1 \\ &+ \sum_{i=1}^{m-1} \int_{y_1^{*(i)}}^{y_1^{*(i+1)}} \int_{-\infty}^{y_2^{*(i+1)}} Y_1 \phi(Y_1, Y_2) \mathrm{d}Y_2 \mathrm{d}Y_1 \\ &+ \int_{y_1^{*m}}^{\infty} \int_{-\infty}^{y_2^{*m}} Y_1 \phi(Y_1, Y_2) \mathrm{d}Y_2 \mathrm{d}Y_1 \end{aligned} \tag{2.48}$$

This is the multi-objective equivalent of (2.31) from Section 2.2.3.1 (Forrester and Keane, 2009). Importantly, this equation is not dependent on the relative scaling between objectives. However, if used as an infill criterion, exploration will not necessarily be strongly encouraged due to the lack of bias given to the degree of improvement actually being achieved.

To ensure a more explicit exploration capability, the first moment of the joint probability density function integral taken over the area where improvements occur is used as an equivalent improvement metric, calculated about the current Pareto front. Resigning the more rigorous derivation and explanation to (Forrester and Keane, 2009), this may ultimately be expressed as:

$$E[I(\mathbf{x}^*)] = P[I(\mathbf{x}^*)] \sqrt{(\bar{Y}_1(x) - y_1^*(x^*))^2 + (\bar{Y}_2(x) - y_2^*(x^*))^2} \tag{2.49}$$

where

$$\bar{Y}_1(x) = \left\{ \begin{array}{l} \displaystyle\int_{-\infty}^{y_1^{*(1)}} \int_{-\infty}^{\infty} Y_1 \phi(Y_1, Y_2) \mathrm{d}Y_2 \mathrm{d}Y_1 \\ \displaystyle + \sum_{i=1}^{m-1} \int_{y_1^{*(i)}}^{y_1^{*(i+1)}} \int_{-\infty}^{y_2^{*(i+1)}} Y_1 \phi(Y_1, Y_2) \mathrm{d}Y_2 \mathrm{d}Y_1 \\ \displaystyle + \int_{y_1^{*m}}^{\infty} \int_{-\infty}^{y_2^{*m}} Y_1 \phi(Y_1, Y_2) \mathrm{d}Y_2 \mathrm{d}Y_1 \end{array} \right\} / P[I(\mathbf{x}^*)] \tag{2.50}$$

where $\bar{Y}_2(x)$ is similarly defined. According to this formulation, $E[I(\mathbf{x}^*)$ varies with the location of the predicted position of the candidate point $(\hat{y}_1, \hat{y}_2)$ as well as with the estimated errors s_1 and s_2, given that it is these quantities defining the PDF. The closer the new prediction lies to the *utopia* point (i.e. ideological minimum of both objectives simultaneously), the closer the integral becomes to unity (that is, the greater the probability that the new prediction will offer an improvement). This then encourages updates that improve the Pareto front relative to all objectives. If there exists significant gaps between the points comprising the

existing Pareto front, areas within or in close proximity to such gaps will score proportionately higher values of $E[I(\mathbf{x}^*)]$, given the higher Euclidean distances to the nearest points. This then encourages a more even spacing as the Pareto front is updated with additional points. Furthermore, if *all* available samples (i.e. not just those comprising the Pareto front) are used to construct the objective surrogates, error terms will tend larger, thus encourage exploration. It can be seen then that this formulation balances exploitation and exploration, similarly to the single-objective case (Section 2.2.3.1). An unavoidable issue with this approach is that although there is no inherent requirement to appropriately scale the objective functions, failing to do so will diminish the quality of the expected results (specifically, the spacing of points along the Pareto front) given the resulting bias in the calculation of Euclidean distances (Forrester and Keane, 2009).

Of note, this formulation may be extended to encompass problems of higher number of objectives, without loss of generality. However, the resulting increase in dimensionality of the Pareto surfaces does inevitably complicate the expressions used to calculate the improvement metrics, though they remain expressible in closed form (Forrester and Keane, 2009).

2.3 Multi-Fidelity Model Management

In place of expensive simulation models or physical experiments, approximation methods can facilitate the integration of discipline-specific computer analyses and can more easily provide insight into the relationships between design (input) variables and system performance (output) responses given the reduced computational cost (Simpson, 2008). Further advantages include (1) offering improved filtering of numerical noise, the prevalence of which may hinder the convergence of many optimisation algorithms, particularly gradient-based methods (Simpson, 2008; Jiang et al., 2020), (2) improved simulation error detection and identification of non-linear/design-defining regions as the entire input-space may be more easily explored and analysed, and in certain applications (3) the enabling of parallel computing and optimisation techniques given the possibility of reusing the same simulation for many design alternatives (Chen et al., 2006).

The utilisation of mathematical methods for approximating expensive functions in *Multidisciplinary Design Optimisation* (MDO) problems originally emerged within the field of structural engineering. Research was motivated primarily by the large computational expense of FEA, consequentially incompatible with most numerical optimisation techniques of the time (Simpson, 2008). Methods utilising RSMs (first introduced by Box and Wilson (1992) and detailed by Myers and Montgomery (1995)) and ANNs (Hajela and Berke, 1992) were the predominant areas of research leading up to 1990 (Barthelemy and Haftka, 1993; Sobieszczanski-Sobieski and Haftka, 1997). Early contributors included research groups at

the Virginia Institute of Technology, the University of Notre Dame, Rensselaer Polytechnic Institute, Old Dominion University, and the NASA Langley Research Centre (Simpson, 2008).

An important concept from this time was that of sequentially approximating the initial optimisation problem by explicit sub-problems based on the information obtained from response analysis and first order design sensitivity analysis[11] (i.e., values of functions and their derivatives) (Schmit and Farshl, 1974). Examples of this approach include *Sequential Linear Programming* (SLP) (Pedersen, 1980), the *CONvex LINearization* (CONLIN) method (Fleury and Braibant, 1986) and the *Method of Moving Asymptotes* (MMA) (Svanberg, 1987). Subsequent work developed this methodology by basing the approximation instead off of the function value and it's derivatives at the current and previous design points[12] (Haftka et al., 1987). The primary goal being to improve the quality of approximations to the point where the number of iterations (and therefore the total required time) needed to solve the original optimisation problem is reduced. Developed methods include the *Accumulated Approximation* (ACAP) technique (Rasmussen, 1990), which used the objective function values obtained at all previous points, as well as the derivatives at the current point, and the multi-point approximation technique by Toropov (1989), which used the objective function values gained at each iteration at several previous design points.

The 1990s saw a particularly strong concentration of research and development efforts regarding response surface modelling from within the MDO community (in part resulting from NASA-funded research related to the *High Speed Civil Transport* (HSCT)) (Haftka et al., 1998; Simpson et al., 2001). For instance, *Variable Complexity Response Surface Modelling* (VCRSM) (Kaufman et al., 1996), developed primarily at Virginia Tech, concentrated the design space around regions of interest using analyses of varying fidelity, then constructed response surface models of the reduced domain in increasing levels of accuracy. *Concurrent SubSpace Optimization* (CSSO) generated response surface approximations of the design space to form the basis of the subspace coordination procedure for MDO (Wujek et al., 1996).

Accompanying the increasing prevalence of response surface methods and applications was the understanding of it's limitations. Namely, the inevitable *curse of dimensionality* (Koch et al., 1999) and the inability to demonstrate global accuracy for highly non-linear design spaces (Simpson et al., 2001). In attempt to mitigate these issues, various new methods were proposed. These included utilising higher order response surfaces (Venter et al., 1998), mixed polynomial models (Roux et al., 1996), and more efficient sampling schemes (Giunta et al., 2003). Some authors promoted the use of gradient information to facilitate surrogate construction (Liu and Batill, 2000, 2002), whilst others concentrated on developing sequential methods for

[11] Evaluated at the current point within the design space, thus methods of this time may be referred to as *single point approximation* methods.

[12] thus may be referred to as *two-point approximations*.

sampling, building, and optimising surrogates, such as move-limits (Toropov et al., 1996) and trust-region approaches (Alexandrov et al., 1997; Rodriguez et al., 1997). In turn, this prompted the development of more rigorously formulated techniques for model management, such as the Surrogate Management Framework (Booker et al., 1999) developed collaboratively at Boeing, IBM, and Rice University. During this time frame, many commercial entities also started to develop software to facilitate the use of approximation methods in design and optimisation. For example, iSIGHT by Engineous Software, Inc., Visual DOC by Vanderplaats R&D, Inc., Optimus by LMS International, ModelCenter by Phoenix Integration, Design Explorer by The Boeing Company and DAKOTA by Sandia National Laboratories.

Nonetheless, given the inherent limitations with response surface modelling and the growing popularity of alternative surrogate modelling techniques (e.g. radial basis functions (Mullur and Messac, 2006), multivariate adaptive regression splines (Hastie et al., 2009), Kriging (Sacks et al., 1989; Cressie, 2015; Sasena, 2002) and support vector regression (Gunn, 1998)), the latter part of the 1990s saw a focus shift towards new SBO methods incorporating these new approaches. With an increased variety of methods came a similarly increased range of applications, including optimisation under uncertainty, robust design and reliability-based design given the usefulness of inexpensive surrogates for Monte Carlo simulation and uncertainty analysis (Koch et al., 2004). The reader is referred to (Jin et al., 2001) for a more complete evaluation and comparison of methods developed during this period.

However, a common theme within physics-based design and optimisation (and other outer-loop applications, such as inference and uncertainty quantification) is that the computational cost of constructing surrogates of an expensive function through direct sampling can nonetheless remain prohibitive (Zhou et al., 2018). This is especially true if underlying non-linearities, discontinuities and multi-modality necessitate a higher number of samples to provide a useful approximation. With this in mind, many researchers began to investigate the incorporation of multiple sources of information to supplement, and thus alleviate, the expensive function of interest. In fact, within many fields and applications across computational science and engineering, any particular system of interest is likely able to be represented by various alternative numerical models/formulations. Such a collection of methods usually conforms to a natural hierarchical trade-off between evaluation cost and relative accuracy. That is, models more representative of the real, underlying physical process(es) induce a more significant computational cost. Adopting the standard nomenclature, a computationally expensive model that describes the system of interest to the accuracy required by the current application is referred to as the *high-fidelity* model, whereas any less accurate but less computationally demanding model is referred to as *low-fidelity*. This relationship can be visualised as in Figure 2.10.

Multi-fidelity (or variable-fidelity) modelling is a method that has received considerable atten-

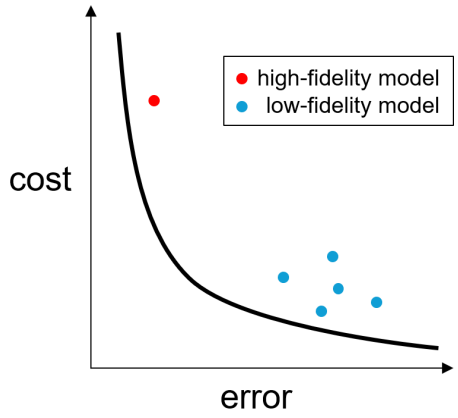


FIGURE 2.10: An example distribution of high/low-fidelity models in terms of computational cost and achievable accuracy.

tion in recent years. The response trends of one or more inexpensive, lower quality numerical models (for example, obtained using a computer simulation based on coarsely-discretised structure of interest) are leveraged to supplement a surrogate construction/refinement process in cooperation with a judiciously limited number of high-fidelity model evaluations, or indeed to entirely replace the original expensive function altogether. The fundamental premise is that the lower-fidelity models can inexpensively supply functional trends with only occasional recourse to the high fidelity model to establish global/local accuracy, apply some correction or adjustment or guarantee outer-loop convergence (e.g. as part of a multi-fidelity optimisation scheme) (Ng and Willcox, 2014; Alexandrov et al., 2001; Koziel and Yang, 2011). In this manner, one may in essence attempt to circumvent the curse of dimensionality associated with black-box surrogate modelling by exploiting domain-specific knowledge (Simpson, 2008; Keane and Nair, 2005).

Two key applications of multi-fidelity modelling can be described via the distinction between *many-query* and *outer-loop* applications (Peherstorfer et al., 2018). *Outer-loop* applications are defined by computational loops encompassing a particular function of interest. At each of multiple iterations, the function of interest is evaluated according to some input specification, with the corresponding response contributing to the calculation of the overall outer-loop result, returned at process termination. A relevant example is that of a typical optimisation process, where the optimiser provides a new set of design variables at every iteration to evaluate the problem objective function(s), constraint functions(s) and gradient information (if required). At the termination of the optimisation process, the optimal design is exported as the outer-loop result. Other examples include uncertainty propagation, inverse problems, data assimilation, control problems, and sensitivity analysis (Peherstorfer et al., 2018). While it is true that most (if not all) outer-loop applications involve a large number of distinct evaluations of the target function(s), the term *'many-query application'* is used in this context to denote those in which

there is no explicit end result despite many repeated evaluations[13]. An example is parameter study, which does not necessarily lead to a specific numerical result.

Describing this distinction is important within the context of multi-fidelity modelling. Specifically, the approximation of a sub-system model using multi-fidelity modelling methods may be classed as *many-query*, whereas the optimisation of an objective function approximated via a multi-fidelity surrogate is naturally concerned with only the *outer-loop* result and it's implications for any specified accuracy and/or convergence criteria. Of note, whilst this example can itself be considered a special case of SBO, one must also consider the case of continuously constructing and refining approximation models of constituent subsystems such that some outer-loop objective function may be minimised with respect to the high-fidelity response(s). This represents a more hybridised many-query/outer-loop application relevant for SBO applications. Regardless of higher-level classification, a multi-fidelity process is typically conducted via two distinct stages:

1. The selection/generation of one or more appropriate low-fidelity models (or surrogates thereof). The assumption here being that the selected model(s) may represent the general trends of the underlying physical process to a reasonable degree for a significantly reduced computational cost in comparison to its high(er)-fidelity counterpart(s).

2. The implementation of an appropriate *model management* strategy to efficient distribute available computational resources amongst the various sources of information to guarantee acceptable numerical accuracy or guide an outer-loop application to convergence.

An effective model management strategy must balance model evaluations between the available levels of fidelity whilst preserving the accuracy of the final outer-loop quantity of interest (e.g. model quality or objective function value) that would be obtained if only the high-fidelity model was used (Peherstorfer et al., 2018). *Multi-Fidelity Model Management* (MFMM) techniques differ primarily in how alternative models are employed during execution of the outer-loop process, and in their approach to combine the varying degrees of fidelity. Based on the definition introduced by Peherstorfer et al. (2018), MFMM techniques may be described as belonging to three succinct classifications; *adaptation*, *fusion* and *filtering*. Included in this classification are multi-fidelity methods used for general model enhancement (i.e. a distinct *many-query* application) out-with any overriding outer-loop context. Figure 2.11 lists several prevalent model management strategies according to this classification.

This section thus considers the remaining area of research concerned with the high-level management of incorporated surrogates. First, a selection of common methods for constructing

[13] In essence then, one may think of outer-loop applications as a subset of many-query applications.

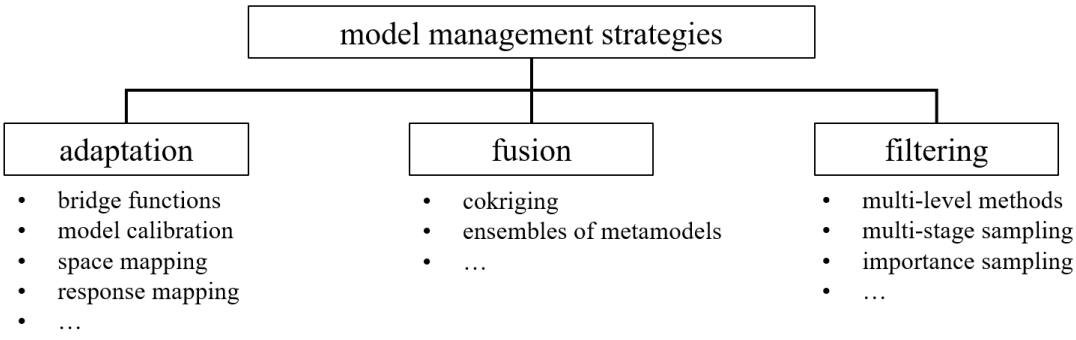


FIGURE 2.11: Classification of multi-fidelity model management strategies, adapted from (Peherstorfer et al., 2018).

low(er)-fidelity approximations is presented and discussed. Next, several distinct multi-fidelity modelling techniques are detailed and discussed, including a selection of relevant examples. Finally, a number of well known management strategies and high-level classifications are introduced and discussed with reference to existing applications.

2.3.1 Low-Fidelity Modelling

Domain specific knowledge of the physical process to be modelled and access to the corresponding high-fidelity simulation method largely determines the available classifications of low-fidelity models that may be constructed. For example, in some cases the high-fidelity has a known structure that may be exploited to create lower fidelity approximations. In other cases, the high-fidelity operates more as a "black box", via which a set of outputs may be obtained for a given input specification without any explicit knowledge of the mathematical process(es) involved. One might expect an inherent reliance of the overall multi-fidelity model prediction quality on the correlation between the constituent low- and high-fidelity approximations (Shi et al., 2020a), thus it becomes important to examine the lower-fidelity modelling options available for a given application.

For simplicity, this section restricts the current characterisation of multi-fidelity modelling methods to include only low-fidelity models that explicitly approximate the same output quantity/quantities as the corresponding high-fidelity alternative. This is in order to distinguish these approaches from methods that embed approximation within the application itself, such as the quasi-Newton (Davidon, 1959) and Krylov subspace methods (Broyden, 1965). Another relevant example is that of multi-resolution/multi-grid methods. While largely related to the current discussion, these methods will be considered in more detail in Part II of this thesis. Low-fidelity models may be constructed based on three high-level classifications (Peherstorfer et al., 2018); *Simplification*, *Projection* and *Data-fitting*.

Simplification refers to models constructed based on simplifying assumptions relative to their high-fidelity counterparts. This implies exploitation of domain expertise and an in-depth knowledge of the high-fidelity implementation. For example, simplifying the analysis model (a coarse instead of refined mesh discretisation (Alexandrov et al., 2001; Choi et al., 2005; Brooks et al., 2011), neglected non-linearities or an appropriately modified early stopping criteria (Jonsson et al., 2015; Koziel and Leifsson, 2013a)), simplifying the domain or level of abstraction (two-dimensional versus three-dimensional models (Forrester et al., 2006; Lazzara et al., 2009), additional boundary conditions (Vitali et al., 1998, 2002) or simplified geometry (Forrester, 2010)) or simplifying the mathematical or physical description (Euler non-cohesive equations instead of the viscous Navier–Stokes equations (Zahir and Gao, 2013)).

This approach naturally allows multiple low-fidelity approximations of various computational costs and fidelities to be constructed via the same base computations. Thus it is easy to compile model hierarchies. Many such hierarchies exist in the field of CFD, for example. For the purpose of modelling and analysing turbulent flow regimes, a typical hierarchy could be (in order of decreasing fidelity): *Direct Numerical Simulations* (DNS), *Large Eddy Simulations* (LES), and *Reynolds Averaged Navier-Stokes* (RANS). Relative to DNS, which resolves the complete spatial and temporal domains to the scale of the modelled turbulence, LES first eliminates small-scale behaviour to simplify the simulation, then RANS simplifies even further by applying the Reynolds decomposition to average the computed values over time (Peherstorfer et al., 2018). More relevant for aerodynamic design problems, modelling fidelity level could be decreased even further by including Euler equations and potential theory to the hierarchy (Han et al., 2013). In Choi et al. (2009), a supersonic aerodynamic design problem is solved by utilising a model hierarchy consisting of the Euler equations, a vortex lattice model, and a classical empirical model.

Projection based models instead take advantage of the problem structure, as opposed to expert domain knowledge (Peherstorfer et al., 2018). In this case, the input/output relationship (or governing equations) of the high-fidelity model is *projected* onto a lower-dimensional subspace, yielding a *Reduced Order Model* (ROM) exploiting the general trends and features of the higher fidelity target. This is often achieved by virtue of detailed knowledge regarding the high-fidelity model structure. *Proper Orthogonal Decomposition* (POD) (Berkooz et al., 1993; Chaturantabut and Sorensen, 2010) constructs a basis for the low-dimensional subspace from selected responses of the high-fidelity model. A method based on *Centroidal Voronoi Tessellation* (CVT) is presented by Du et al. (1999), where a reduced basis is derived using a Voronoi clustering of individual "snapshots" of the target function. See the work by Burkardt et al. (2006) for more details regarding CVT-based ROM. Additional methods include RBFs (Rozza et al., 2007), dynamic mode decomposition (Schmid, 2010) and Krylov Subspace methods (Feldmann and Freund, 1994). See the work by Benner et al. (2015) for an in-depth review of

projection based reduced order models.

Data-fit models are derived directly from a finite set of input/output correspondence. An advantage here is that there is no need for any additional formulations, as samples may be collected directly using the existing black-box high-fidelity simulation. In many scenarios, data-fit models may be constructed as linear combinations of basis functions or surrogates (Peherstorfer et al., 2018). Traditional interpolation or single-fidelity surrogate modelling techniques (such as Polynomial *Response Surface Model* (RSM) (Myers and Montgomery, 1995), *Radial Basis Functions* (RBFs) (Buhmann, 2003), Kriging (Kennedy and O'Hagan, 2000), *Support Vector Regression* (SVR) (Shi et al., 2020a) and *Artificial Neural Networks* (ANNs) (Beale et al., 2017)) may thus be employed as basis functions. The coefficients of the linear combination may be simply evaluated via interpolation or regression, the latter with reference to the recorded set of high-fidelity samples. Perhaps the most common approach is to use simple polynomial basis functions e.g. Lagrange polynomials. Employing polynomials as part of a piecewise interpolation allows the use of lower-order polynomials, thus avoiding over-fitting problems related to global, higher order bases e.g., Runge's phenomenon. As a closely related approach, *Radial Basis Function* (RBF) (Section 2.1.2) are unsurprisingly also a popular choice for constructing data-fit low-fidelity models. When employing a Gaussian density function as a basis, a numerically stable computation of coefficients of the linear combination of the data-fit model can be achieved (Forrester and Keane, 2009). However, similarly to other forms of surrogate modelling, RBFs depend on a suitable choice of hyperparameters (e.g. the bandwidth of the Gaussian density function) in order to provide an accurate representation of the provided data. Choosing appropriate values for such hyperparameters is often difficult *a priori* and the optimisation thereof may be computationally expensive (Forrester and Keane, 2009).

2.3.2 Adaptation

Adaptation refers to the process of directly correcting or enhancing an existing low-fidelity model (or set of models) using selective high-fidelity samples. Common forms of adaptive correction include bridge functions (additive, multiplicative or comprehensive) (Han et al., 2013; Fischer et al., 2018; Gano et al., 2005) , model calibration (Kennedy and O'Hagan, 2001a; March and Willcox, 2012b), space mapping (Bandler et al., 1994; Koziel et al., 2008), shape preserving response prediction (Koziel and Leifsson, 2014) and adaptive response correction (Koziel et al., 2009; Leifsson et al., 2016). The following briefly reviews the more prevalent examples of adaptation for multi-fidelity surrogate modelling.

2.3.2.1 Bridge functions

The most widely used form of multi-fidelity model adaptation is a correction-based approach, alternatively referred to as "bridge functions", "scaling functions" or "calibration". The first

widely accepted example of this methodology was introduced by Haftka in the 1990s (Haftka, 1991; Chang et al., 1993) as part of early research into incorporating approximations into structural optimisation problems. Motivated by an aim to extend the range of applicability of a local derivative-based approximation of a high-fidelity response at a current design point, Haftka proposed a *multiplicative* scaling factor based on the interaction between the high- and low-fidelity models. A multiplicative bridge function can be written as:

$$\phi(\mathbf{x}) = \frac{\mathbf{f}_H(\mathbf{x})}{\mathbf{f}_L(\mathbf{x})} \tag{2.51}$$

where $\mathbf{f}_H$ and $\mathbf{f}_L$ represent the high- and low-fidelity functions respectively. The bridge function $\phi(\mathbf{x})$ is considered as an unknown, and must itself be approximated (i.e. $\hat{\phi}(\mathbf{x})$) using (for example) some appropriate surrogate modelling technique, or a first-order Taylor series expansion (Gano et al., 2005; Eldred et al., 2004). The high-fidelity function may thus be locally approximated according to:

$$\hat{\mathbf{f}}_H(\mathbf{x}) = \hat{\phi}(\mathbf{x})\mathbf{f}_L(\mathbf{x}) \tag{2.52}$$

where the low-fidelity function $\mathbf{f}_L(\mathbf{x})$ may similarly be replaced with a surrogate approximation $\hat{\mathbf{f}}_L(\mathbf{x})$. However, it can be seen that this approach may encounter difficulty if the value of a low-fidelity function sample is close or equal to zero. In such cases, both the high- and low-fidelity values should be shifted to avoid error (Fischer et al., 2018).

To avoid this problem, Lewis and Nash (2000) proposed instead the *additive* bridge function for global approximation, rather than local. The additive bridge function can be written as:

$$\gamma(\mathbf{x}) = \mathbf{f}_H(\mathbf{x}) - \mathbf{f}_L(\mathbf{x}) \tag{2.53}$$

Similarly as for the multiplicative approach, the additive bridge function $\gamma(\mathbf{x})$ (essentially representing the error between the two responses) is unknown and must be approximated as $\hat{\gamma}(\mathbf{x})$. Again, the low fidelity function may also be approximated as $\hat{\mathbf{f}}(\mathbf{x})$. The resulting approximation of the high-fidelity response is thus written:

$$\hat{\mathbf{f}}_H(\mathbf{x}) = \mathbf{f}_L(\mathbf{x}) + \hat{\gamma}(\mathbf{x}) \tag{2.54}$$

The additive bridge function has been proven to be more accurate and robust than the multiplicative approach (Han et al., 2013), and has since become perhaps the most popular of all adaptation approaches for mutifidelity surrogate modelling (Choi et al., 2005, 2009; Peherstorfer et al., 2018; Simpson, 2008). However, despite generally offering a better approximation, Gano et al. (2005) showed that this was not true for all cases, and that the multiplicative approach remained viable. Accordingly, the authors proposed a *hybrid* approach, aiming to incorporate the strengths of both additive and multiplicative approaches. In this manner, a

more complicated correlation between high- and low-fidelity responses could be represented. A hybrid (or *comprehensive*) bridge function may be incorporated as follows:

$$\hat{\mathbf{f}}_H(\mathbf{x}) = \omega[\hat{\phi}(\mathbf{x})\mathbf{f}_L(\mathbf{x})] + (1-\omega)[\mathbf{f}_L(\mathbf{x}) + \hat{\gamma}(\mathbf{x})] \tag{2.55}$$

where ω is a weight coefficient representing the relative importance attributed to the additive and multiplicative components. Controlling the relative weighing between components with a single parameter allows a continuous adaptation to always best suit a given problem, determined by the optimiser rather than the user (Fischer et al., 2018).

The determination of ω is naturally an important topic. Eldred et al. (2004) proposed an adaptive method of determining and updating the value of ω based on the previously evaluated sample:

$$\omega = \frac{\mathbf{f}_H(\mathbf{x}_{old}) - [\hat{\mathbf{f}}_L(\mathbf{x}_{new}) + \hat{\gamma}(\mathbf{x}_{new})]}{\hat{\phi}(\mathbf{x}_{new})\hat{\mathbf{f}}_L(\mathbf{x}_{new}) - [\hat{\mathbf{f}}_L(\mathbf{x}_{new}) + \hat{\gamma}(\mathbf{x}_{new})]} \tag{2.56}$$

where $\mathbf{f}_H(\mathbf{x}_{old})$ denotes the optimum of the previous optimisation step. During an optimisation process, the value of ω is thus updated at each iteration, allowing the bridge function to adopt the most suitable configuration for the region surrounding the current optimum in design space. The hybrid bridge function is formulated in this manner in order to be particularly suited to applications within optimisation. If one were concerned with instead creating a global approximation of a large parameter space using a hybrid approach, a more general formulation is required (Han et al., 2013).

To address this issue, Han et al. (2013) proposed the *Generalised Hybrid Bridge Function* (GHBF), aiming to more fully exploit the underlying trends of the low-fidelity data without requiring an iterative outer-loop. In this method, the relationship between the low- and high-fidelity functions is expressed more generally as:

$$\hat{\mathbf{f}}_H(\mathbf{x}) = \phi(\mathbf{x}) * \mathbf{f}_L(\mathbf{x}) + \gamma(\mathbf{x}) \tag{2.57}$$

where γ is an additive bridge function and ϕ is a multiplicative scaling function. Both γ and ϕ can be basis functions or surrogate models generated from available samples. In (Han et al., 2013), the authors represent the multiplicative scaling ϕ with a low-order polynomial RSM. The additive correction γ, thus representing the difference between the high-fidelity model and the *scaled* low-fidelity model, is given by a Kriging model trained on the available high-fidelity samples. The authors additionally offer the following basic conditions that remain largely relevant for the general construction of multi-fidelity surrogates (Han et al., 2013):

- *Interpolation condition*: The high-fidelity sampled data should be exactly reproduced, as random errors do not exist for a deterministic computer experiment.

- *Convergence condition*: With an increasing number of high-fidelity samples, the influence of the low-fidelity data on the resulting multi-fidelity model should be gradually decreased. I.e. the surrogate modelling process should converge to the exact functional relationship.

In (Fischer et al., 2018), the hybrid correction approach is updated further in order to determine the relative scaling between multiplicative and additive correction based on more than just one sample, in addition to data other than that used to generate the corrective surrogates. This latter point is essential given that most surrogate modelling techniques aim to meet the above *interpolation condition*, that is, that the modelling error is zero at the locations of all high-fidelity samples. The authors propose a Bayesian posterior updating approach for the weighting coefficients that incorporates the available data within the weighting region of the current subproblem optimisation, defined according to a trust-region based framework. More specifically, the weighting region encompasses and surrounds the trust region according to some specified bounding conditions. Adopting the formulation of Kennedy and O'Hagan (2001a), the hybrid bridge function is given as:

$$\tilde{\phi}(\mathbf{x}) = \omega_{mult}\tilde{\phi}(\mathbf{x}) + \omega_{add}\tilde{\phi}(\mathbf{x}) \tag{2.58}$$

By exploiting the fact that a trust-region model management scheme will converge upon subproblem optima that lie at the respective trust region bounds, the authors calculate the bridge function scaling coefficients using the data that surrounds the trust region, thus ultimately selecting the model that is most accurate within the regions the optimum is expected to reside. The weighting coefficients are calculated as:

$$\omega_{mult}^{k} = \frac{\omega_{mult}^{k-1}\psi_{mult}(\mathbf{x})}{\omega_{add}^{k-1}\psi_{add}(\mathbf{x}) + \omega_{mult}^{k-1}\psi_{mult}(\mathbf{x})} \tag{2.59}$$

and

$$\omega_{add}^{k} = \frac{\omega_{add}^{k-1}\psi_{add}(\mathbf{x})}{\omega_{add}^{k-1}\psi_{add}(\mathbf{x}) + \omega_{mult}^{k-1}\psi_{mult}(\mathbf{x})} \tag{2.60}$$

If one defines the initial values $\omega_{mult}^{0} = \omega_{add}^{0}$, this approach allows for an adaptive update, essential to a *Bayesian-Inspired Multi-Fidelity Optimisation* (BIMFO) methodology. The functions $\psi_{mult}(\mathbf{x})$ and $\psi_{add}(\mathbf{x})$ represent the model likelihoods of $\hat{\phi}_{mult}(\mathbf{x})$ and $\hat{\phi}_{add}(\mathbf{x})$, defined respectively as:

$$\psi_i(\mathbf{x}) = \left(\frac{1}{2\pi\hat{\sigma}_{i,MLE}^{2}}\right)^{n/2} \exp\left(-\frac{n}{2}\right) \tag{2.61}$$

and

$$\hat{\sigma}_{i,MLE}^{2} = \frac{\sum_{j=1}^{n}(\mathbf{f}_H(\mathbf{x}_j) - \tilde{\phi}_i(\mathbf{x}_j))^2}{n} = \text{SSE} \tag{2.62}$$

where n is the number of samples available from previous iterations that lie within the current weight region, i denotes either the multiplicative or additive variations, and j represents the j-th sample within the current weight region.

A Bayesian updating technique such as this allows the use of multi-point surrogates. This has proven to enable faster convergence for complex optimisation problems (Fischer et al., 2018). Importantly, this technique is also able to adaptively redefine the bounds of the trust region, should the corrected multi-fidelity approximation be discovered as less than acceptably accurate.

2.3.2.2 Model Calibration

Another adaptation method most commonly referenced within the engineering community is *model calibration* (Kennedy and O'Hagan, 2001a; March and Willcox, 2012b). *Calibration* refers to the process of adjusting/estimating unknown model input parameters in order to match the model response (as closely as possible) with observed data (Kennedy and O'Hagan, 2001a). By setting these context-specific inputs to their estimated values, the behaviour of the target process may then be predicted at unobserved locations, though residual uncertainty should remain a consideration.

Whilst calibration has been successfully applied to improve model prediction by adjusting physical parameters to best agree with experimental data (for example see the works by Kosonen and Shemeikka (1997); McFarland et al. (2008); Christensen and Willcox (2012); Yoo and Choi (2013)), Fernández-Godino et al. (2016) argue that the method itself, in a pure sense, cannot be considered as a multi-fidelity approach given the inherent lack of multi-fidelity data sources. However, Kennedy and O'Hagan (2001a) presented a Bayesian calibration approach that includes a discrepancy function, thus calibration parameters may be simply treated as other hyper-parameters, and marking this approach more as a multi-fidelity adaptation method. This approach potentially can result in non-physical values for the calibration parameters, though only if by doing so converges on a stronger approximation of the included data. See the works by Qian and Wu (2008); Biehler et al. (2015) for examples of successful application of this methodology.

2.3.2.3 Space Mapping

An alternative approach is to establish a transformation, or *mapping*, from the parameter space of one model to that of another (Simpson, 2008; Koziel et al., 2008). Utilising such a transformation, a low-fidelity approximation with 'mapped' parameters can accurately reflect the response behaviour of the high-fidelity target function. A helpful analogy is that of drawing the low-fidelity model on a rubberised sheet, which is then distorted to topologically align with the high-fidelity response (Keane and Nair, 2005). Both linear and non-linear mappings

have been considered in the literature (see, e.g., Bandler et al. (1994), Bakr et al. (2002) and Koziel et al. (2006a) for details).

A design-space mapping can occasionally be obvious and/or problem-specific (Robinson et al., 2008a). For instance, if the high- and low-fidelity models are represented by the same set of equations, but defined over grids of different resolutions. In this case, the input vectors are the geometric parameters defining each grid, thus the low-fidelity input vector can be a subset of the high-fidelity input vector. Alternatively, the high-fidelity input vector can be an interpolation of the low-fidelity input vector. In most cases however, no clear mathematical relationship exists between the two input vectors, thus an empirical mapping is needed (Robinson et al., 2008a). An example is the multi-fidelity supersonic business jet problem presented by Choi et al. (2005). Given that the SBO methods of the time could not be applied to problems in which the low- and high-fidelity models required different input variables, the authors opted to use the two models sequentially. An initial optimisation was performed using only the low-fidelity model with an applied Kriging-based correction, with the result used as an initial guess for a subsequent optimisation employing only the high-fidelity model.

The more rigorously formulated *Space Mapping* (SM) methodology was originally presented by Bandler et al. (1994, 2003) for the optimal design of microwave engineering applications, for which it was functionally more appropriate to make corrections to the input of the system model, rather than to its output. The original approach implemented a simple linear correction of the low-fidelity input space with the goal being to generate an appropriate transformation of the vector of high-fidelity model parameters $\mathbf{x}_{HF}$ to the vector of low-fidelity model parameters $\mathbf{x}_{LF}$:

$$\mathbf{x}_{LF} = f(\mathbf{x}_{HF}) \tag{2.63}$$

Formulating in this manner, the vectors $\mathbf{x}_{HF}$ and $\mathbf{x}_{LF}$ can be of different dimensions. Determining the transformation $\mathbf{F}$ is usually an iterative process. The goal is to satisfy the condition:

$$\|f_{HF}(\mathbf{x}_{HF}) - f_{LF}(\mathbf{x}_{LF})\| \leq \epsilon \tag{2.64}$$

within some local region, where ϵ represents an appropriate tolerance setting. This method proved particularly suitable for engineering applications where model discrepancy could be primarily attributed to second-order effects, such as the presence of parasitic components in electrical engineering problems (Koziel and Yang, 2011). In such applications (i.e. those based on physical processes), the responses from models of varying levels of fidelity are often similar in general form, though slightly distorted/shifted with respect to certain input parameters.

Space Mapping and it's associated optimisation-based applications (see Figure 2.12) have since been developed into a distinct family of alternative methods (Fernandez et al., 2017;

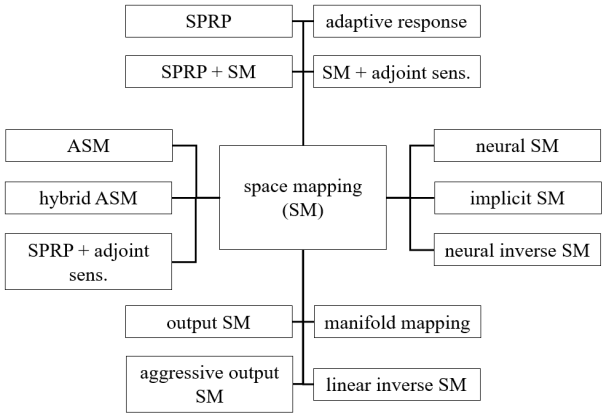


FIGURE 2.12: Fundamental design optimisation methods stemming from the *Space Mapping* (SM) concept, including *Aggressive Space Mapping* (ASM) and *Shape Preserving Response Prediction* (SPRP) based approaches, adapted from (Rayas-Sánchez, 2015).

Bakr et al., 2002, 2000; Koziel and Leifsson, 2016; Koziel et al., 2006a), including implicit space mapping (Bandler et al., 2004), neural-based space mapping (Zhang et al., 2008), ASMs (Bandler, 1995), corrected space mapping (Robinson et al., 2008b) and *Shape Preserving Response Prediction* (SPRP) (Koziel and Leifsson, 2014), to name only a few. The reader is referred to the reviews (Bandler et al., 2003; Rayas-Sánchez, 2015) for a more comprehensive compendium of this class of methods. To distinguish between the large variety of related methods, Koziel and Yang (2011) offer the following classifications to encompass the majority of well-known approaches:

- *Input Space Mapping* (Bandler et al., 2003): The response correction is based on an affine transformation on the low-fidelity input parameter space, i.e. $\hat{f}_{HF}(\mathbf{x}, \mathbf{p}) = f_{LF}(\mathbf{B}\mathbf{x}+\mathbf{c})$ where $\mathbf{p}(\mathbf{B}, \mathbf{c})$. This classification is most similar to the original formulation of Space Mapping.

- *Output Space Mapping* (Koziel et al., 2006a): The response correction is based on an affine transformation on the low-fidelity model response, i.e. $\hat{f}_{HF}(\mathbf{x}, \mathbf{p}) = \mathbf{A}f_{LF}(\mathbf{x}) + \mathbf{d}$ where $\mathbf{p}(\mathbf{A}, \mathbf{d})$. This classification is functionally similar to a comprehensive adaptation (Section 2.3.2).

- *Implicit Space Mapping* (Koziel and Leifsson, 2010): Additional parameters $\mathbf{x}_p \in R^{n_p}$ of the low-fidelity model $f_{LF}(\mathbf{x}, \mathbf{x}_p)$ are tuned to align the low- and high-fidelity responses, i.e. $\hat{f}_{HF}(\mathbf{x}, \mathbf{x}_p) = f_{LF}(\mathbf{x}, \mathbf{x}_p)$. Note: the parameters $\mathbf{x}_p$ are distinguished as distinctly separate from optimisation variables.

- *Custom corrections* (Bandler et al., 2003): The structure of the given design problem is exploited. For example, the model responses could be obtained through the sweeping of

some parameter t (e.g. when t represents time or frequency (Koziel and Yang, 2011)):

$$\begin{aligned}\mathbf{f}_{HF}(\mathbf{x},t) &= [f_{HF}(\mathbf{x},t_1), f_{HF}(\mathbf{x},t_2), \dots f_{HF}(\mathbf{x},t_m)]^T, \\ \mathbf{f}_{LF}(\mathbf{x},t) &= [f_{LF}(\mathbf{x},t_1), f_{LF}(\mathbf{x},t_2), \dots f_{LF}(\mathbf{x},t_m)]^T\end{aligned} \tag{2.65}$$

2.3.2.4 Shape Preserving Response Prediction

Shape Preserving Response Prediction (SPRP) is a multi-fidelity adaptation method aimed at further exploiting low-fidelity physics-based models (Koziel and Leifsson, 2014). Part of the extended family of Space Mapping based approaches, SPRP was also developed within the field of microwave engineering, but has since found application in other fields, such as antenna design (Koziel et al., 2012) and aerodynamic design (Koziel and Leifsson, 2012, 2013a).

Differing from previously discussed approaches, SPRP is a non-parametric method, relying instead on the *change* in low-fidelity response relative to a specified change in input to construct an appropriate surrogate. The predicted change in response for an unknown input translation can then be applied to a reference high-fidelity model to subsequently estimate the corresponding response. By exploiting the general trends of the low-fidelity model, SPRP can be generalised and applied to a wider variety of problems compared to other physics-based surrogate modelling methods.

Naturally, it is of crucial importance that the low-fidelity model is physically-based, thus able to roughly represent the base physical phenomena constituting the modelled process. This ensures that the effects of varying the input parameters on the resulting response is similar for both the low- and high-fidelity simulations.

The change in low-fidelity response is represented by a finite set of translation vectors corresponding to a set of *characteristic points*, placed across the model response. These translation vectors can thus be used to estimate the high-fidelity model response when applied to the same set of characteristic points evaluated using the high-fidelity model. The original SPRP formulation (in the context of a microwave engineering problem) (Koziel, 2009) can be expressed as follows. Let the set of high- and low-fidelity samples be respectively written as:

$$\begin{aligned}\mathbf{f}_{HF}(\mathbf{x}) &= [f_{HF}(\mathbf{x},\omega_1), \dots, f_{HF}(\mathbf{x},\omega_m)]^T \\ \mathbf{f}_{LF}(\mathbf{x}) &= [f_{LF}(\mathbf{x},\omega_1), \dots, f_{LF}(\mathbf{x},\omega_m)]^T\end{aligned} \tag{2.66}$$

where $\omega_j, j = 1, \dots, m$ is the frequency sweep (assuming the microwave models may be parameterised by frequency). The sets of characteristic points of $\mathbf{f}_{HF}(\mathbf{x}^{(i)})$, $\mathbf{f}_{LF}(\mathbf{x}^{(i)})$ and

$\mathbf{f}_{HF}(\mathbf{x})$ are respectively defined as:

$$\begin{aligned}\mathbf{p}_j^{HF} &= [\omega_j^{HF}, r_j^{HF}]^T \\ \mathbf{p}_j^{LF_0} &= [\omega_j^{LF_0}, r_j^{LF_0}]^T \\ \mathbf{p}_j^{LF} &= [\omega_j^{LF}, r_j^{LF}]^T\end{aligned} \tag{2.67}$$

where $j = 1, \ldots, K$. Here, ω and r represent the frequency and magnitude of the corresponding sample point. The translation vectors of the low-fidelity model are defined as:

$$\mathbf{t}_j = [\omega_j^t, r_j^t]^T, \quad j = 1 \ldots, K \tag{2.68}$$

where

$$\begin{aligned}\omega_j^t &= \omega_j^{LF} - \omega_j^{LF_0} \\ r_j^t &= r_j^{LF} - r_j^{LF_0}\end{aligned} \tag{2.69}$$

The SPRP surrogate may thus be written as:

$$\hat{\mathbf{f}}_{HF}(\mathbf{x}) = [\hat{f}_{HF}^{(i)}(\mathbf{x}, \omega_1), \ldots, \hat{f}_{HF}^{(i)}(\mathbf{x}, \omega_m)]^T \tag{2.70}$$

where

$$\hat{f}_{HF}^{(i)}(\mathbf{x}, \omega_j) = \bar{f}_{HF}(\mathbf{x}^{(i)}, F(\omega_j, \{-\omega_k^t\}_{k=1}^K)) + R(\omega_j, \{r_k^t\}_{k=1}^K) \tag{2.71}$$

for $j = 1, \ldots, m$. $\bar{f}_{HF}(\mathbf{x}, \omega)$ represents an interpolation of $\{f_{HF}(\mathbf{x}, \omega), \ldots, f_{HF}(\mathbf{x}, \omega)\}$ onto the frequency interval $[\omega_1, \omega_m]$. The scaling function F interpolates the data pairs:

$$\{\omega_1, \omega_1\}, \{\omega_1^{HF}, \omega_1^{HF} - \omega_1^t\}, \ldots, \{\omega_K^{HF}, \omega_K^{HF} - \omega_K^t\}, \{\omega_m, \omega_m\}$$

onto the frequency interval $[\omega_1, \omega_m]$. The function R performs a similar interpolation for data pairs:

$$\{\omega_1, r_1\}, \{\omega_1^{HF}, r_1^{HF} - r_1^t\}, \ldots, \{\omega_K^{HF}, r_K^{HF} - r_K^t\}, \{\omega_m, r_m\}$$

where

$$\begin{aligned}r_1 &= f_{LF}(\mathbf{x}, \omega_1) - f_{LF}(\mathbf{x}^r, \omega_1) \\ r_m &= f_{LF}(\mathbf{x}, \omega_m) - f_{LF}(\mathbf{x}^r, \omega_m)\end{aligned} \tag{2.72}$$

The function F essentially translates the frequency components of the characteristic points of $\mathbf{f}_{HF}(\mathbf{x}^{(i)})$ to the frequencies at which they should be located according to the translation vectors $\mathbf{t}_j$. On the other hand, the function R adds the necessary magnitude component. The interpolation onto $[\omega_1, \omega_m]$ is necessary because the original frequency sweep is a discrete set.

2.3.3 Fusion

Multi-fidelity *fusion* involves the direct combination of various distinct levels of fidelity into a single multi-fidelity surrogate, without any explicit correction. Examples of this methodology include the work by Allaire and Willcox (2014), where the authors approximate each level of fidelity via a separate Kriging surrogate, then use the technique proposed by Winkler (1981) to combine the random variables associated with each Kriging approximation, thus producing a single fused model. More recently, Zhou et al. (2019) construct a Kriging surrogate based on the fusion of both low- and high-fidelity samples first selected according to objective space crowding, then used to evaluate the fitness of population individuals. The surrogate is refined based on both an individual and a generation-based updating strategy.

2.3.3.1 Ensembles of Metamodels

Perhaps conceptually the simplest method of multi-fidelity data fusion is to take a weighted average of multiple responses. For this approach, each distinct source of data is used to construct a separate surrogate model. Each model is then leveraged into cooperation as part of an *ensemble* (Goel et al., 2007). Assuming that such an ensemble can be generated without significant computational cost (at least in relation to the cost of acquiring data from each source), the ensemble approach effectively infers the overriding trends of the target function. Furthermore, being fundamentally based on the disagreement between competing models, this method can be used as an uncertainty quantifier and/or a method to inform an infill criteria.

Averaging the responses from multiple surrogates is motivated by an inability to find a unique solution to the non-linear inverse problem of identifying a suitable mathematical relationship from a limited set of data (Queipo et al., 2005). Goel et al. (2007) describe the applications of an *ensemble of meta-models* as conforming to one of two major areas.

- Identifying regions of large uncertainty. This idealisation largely overlaps, at least conceptually, with the preceding discussion on multi-response sampling given a heterogeneous/homogeneous *committee* of competing approximations (see Section 2.2.3.2).
- Employing the ensemble via a weighted averaging or the selection of the single best surrogate model based on error statistics. This ultimately may give a more robust approximation than any one individual surrogate.

Considering the first application, regions of large uncertainty/disagreement can be identified as follows. Given an ensemble of n_{SM} surrogate models, the standard deviation of a prediction at point $\mathbf{x}$ is given as:

$$\sigma\left(\hat{\mathbf{y}}(\mathbf{x})\right) = \sqrt{\frac{\sum_{i=1}^{n_{SM}} \left(\hat{y}_i - \bar{y}_i(\mathbf{x})\right)^2}{n_{SM} - 1}} \tag{2.73}$$

where

$$\bar{y}_i(\mathbf{x}) = \frac{\sum_{i=1}^{n_{SM}} \hat{y}_i(\mathbf{x})}{n_{SM}} \tag{2.74}$$

In regions where surrogate predictions differ more dramatically, the standard deviation will be higher, thus these areas are most suitable for future sampling. Of note, while a high value of standard deviation may indicate a region of high uncertainty, a low standard deviation does not imply accuracy. That is, the ensemble may unanimously agree upon an incorrect prediction. The weighted average of the ensemble may be used to describe a single surrogate according to:

$$\hat{\mathbf{y}}(\mathbf{x}) = \sum_{i}^{n_{SM}} w_i(\mathbf{x})\ \hat{y}_i(\mathbf{x}) \tag{2.75}$$

where $\hat{\mathbf{y}}(\mathbf{x})$ is the predicted response of the weighted average surrogate model, $\hat{y}_i(\mathbf{x})$ is the response of the i-th surrogate, and $w_i(\mathbf{x})$ is the weight associated with the i-th surrogate at design point $\mathbf{x}$. Importantly, the sum of all weights must equal unity:

$$\sum_{i=1}^{n_{SM}} w_i = 1 \tag{2.76}$$

such that a case of unanimous agreement across the ensemble is correspondingly reflected in $\hat{\mathbf{y}}(\mathbf{x})$. A surrogate that is deemed of higher accuracy is given a larger weight, and conversely, lower accuracy models are ensured to have less influence on the fused predictions.

Assessing the individual qualities of the ensemble can be accomplished using surrogate validation metrics (see Section 2.1.6). In this context, the measures of interest are more broadly categorised as those of (1) global vs local accuracy and (2) predicted vs observed data (Goel et al., 2007). These local measures of fit quality are used to determine the weights values as functions of design space, i.e. $w_i = w_i(\mathbf{x})$. For instance, the weights can be determined using local surrogate-based metrics (e.g. prediction variance, mean squared error etc) or data-based metrics (e.g. interpolated cross-validation errors). Alternatively, global surrogate-based metrics include the RMSE (if employing polynomial response surface approximations, for example), or the process variance (if employing Kriging), and global data-based metrics include the CV error. Of note, weights selection based on global metrics naturally requires that the weights remains fixed relative to input space, i.e. $w_i(\mathbf{x}) = C_i, \forall \mathbf{x}$. Despite this variety, an appropriate selection of weight functions remains a formidable task for most applications. Examples of this approach in the literature include the work of Zerpa et al. (2005), where a local weighted average model is constructed from an ensemble consisting of a polynomial RSM, a Kriging model and an RBF surrogate. The resulting ensemble is used for the optimisation of an alkali–surfactant-polymer flooding process, where the point-wise estimate of the variance is predicted by the ensemble. Another example is the work by Zhou et al. (2016), where three alternative surrogate modelling techniques are chosen based on their individual strengths. Specifically, Kriging is chosen for it's suitability for non-linear problems, RBF for the only slight influence of sample size on

the average accuracy and robustness and SVR for it's stronger performance for high-order problems. In this case, weight factors are determined by minimising the *Generalised Mean Square Leave-One-Out Error* (GMSELOO) of the ensemble.

Zhou et al. (2016) additionally note that the choice of how many competing models to include within the ensemble is an important consideration. Two may not be enough to provide sufficient levels of disagreement, especially if either surrogate is not wholly suitable for the problem at hand. On the other hand, four or more competing surrogates may considerably increase the computational cost, reducing the overall efficiency. Furthermore, Viana and Haftka (2008) point out that (i) the potential gains from using weighted surrogates diminish substantially in high dimensions, and (ii) the poor quality of the information given by CV errors in low dimensions makes the gain very difficult in practice. However, Viana and Haftka (2008) conclude that the use of multiple surrogate (with or without fusion) is particularly appealing for design optimisation problems due to the fact that the best surrogate may not lead to the best result. Furthermore, fitting many surrogates and repeating optimisations may still be significantly less expensive compared to collecting data via simulation.

2.3.3.2 Co-Kriging

One of the more prevalent approaches for multi-fidelity data fusion is the Co-Kriging method (Myers, 1984). Whilst essentially an extension of the standard Kriging methodology to incorporate several responses, the fundamental idea behind Co-Kriging remains applicable as an enhancement to any parametric RBF (Forrester and Keane, 2009).

Considering two sets of data, representing high- and low-fidelity responses respectively, the Co-Kriging formulation begins by combining the two into one shared set of sample locations:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{LF} \\ \mathbf{X}_{HF} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_{LF}^{(1)} \\ \vdots \\ \mathbf{x}_{LF}^{(n_{LF})} \\ \\ \mathbf{x}_{HF}^{(1)} \\ \vdots \\ \mathbf{x}_{HF}^{(n_{HF})} \end{bmatrix} \tag{2.77}$$

As for the standard Kriging formulation, the response value evaluated at $\mathbf{X}$ is considered the

realisation of a stochastic process. For Co-Kriging, we therefore have the random field:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_{LF}(\mathbf{X}_{LF}) \\ \mathbf{Y}_{HF}(\mathbf{X}_{HF}) \end{bmatrix} = \begin{bmatrix} Y_{LF}(\mathbf{x}_{LF}^{(1)}) \\ \vdots \\ Y_{LF}(\mathbf{x}_{LF}^{(n_{LF})}) \\ \\ Y_{HF}(\mathbf{x}_{HF}^{(1)}) \\ \vdots \\ Y_{LF}(\mathbf{x}_{HF}^{(n_{HF})}) \end{bmatrix} \tag{2.78}$$

This formulation assumes the use of an *auto-regressive* model (Kennedy and O'Hagan, 2000), which states that the covariance conforms to:

$$\text{cov}\{Y_{HF}(\mathbf{x}^{(i)}, Y_{LF}(\mathbf{x}))|Y_{LF}(\mathbf{x}^{(i)})\} = 0, \forall \mathbf{x} \neq \mathbf{x}^{(i)} \tag{2.79}$$

This is also known as a *Markov* property, which assumes the high-fidelity simulation is correct, thus any inaccuracies must lie within the low-fidelity model. Thus, if the high-fidelity response at $\mathbf{x}^{(i)}$ is known, no more can be learned by evaluating the same point with the low-fidelity model.

Local features of the low- and high-fidelity responses are expressed using the Gaussian functions Z_{LF} and Z_{HF}. Using the auto-regressive model, the high-fidelity response can be approximated in a manner reminiscent of the comprehensive correction approach of adaptation methods (Section 2.3.2), that is:

$$Z_{HF}(\mathbf{x}) = \rho Z_{LF}(\mathbf{x}) + Z_d(\mathbf{x}) \tag{2.80}$$

where ρ is a constant scaling factor and Z_d is another Gaussian process, this time representing the discrepancy between ρZ_{LF} and Z_{HF}. In standard Kriging, the covariance matrix is:

$$\text{cov}\{\mathbf{Y}(\mathbf{X}), \mathbf{Y}(\mathbf{X})\} = \sigma^2 \mathbf{\Psi}(\mathbf{X}, \mathbf{X}) \tag{2.81}$$

For Co-Kriging however, the covariance matrix is now written as:

$$\mathbf{C} = \begin{bmatrix} \sigma_{LF}^2 \mathbf{\Psi}_{LF}(\mathbf{X}_{LF}, \mathbf{X}_{LF}) & \rho\sigma_{LF}^2 \mathbf{\Psi}_{LF}(\mathbf{X}_{LF}, \mathbf{X}_{HF}) \\ \rho\sigma_{LF}^2 \mathbf{\Psi}_{LF}(\mathbf{X}_{HF}, \mathbf{X}_{LF}) & \rho^2\sigma_{LF}^2 \mathbf{\Psi}_{LF}(\mathbf{X}_{HF}, \mathbf{X}_{HF}) + \sigma_d^2 \mathbf{\Psi}_d(\mathbf{X}_{HF}, \mathbf{X}_{HF}) \end{bmatrix} \tag{2.82}$$

The notation $\mathbf{\Psi}_{LF}(\mathbf{X}_{HF}, \mathbf{X}_{LF})$ denotes a matrix of correlations of the form ψ_{LF} between the data $\mathbf{X}_{HF}$ and $\mathbf{X}_{LF}$. Correlations are expressed in the same form as for Kriging:

$$\psi^{(i)} = \text{cor}[Y(\mathbf{x}^{(i)}), Y(\mathbf{x})] = \exp\left(-\sum_{j=1}^{k} \theta_j \|x_j^{(i)} - x_j\|^{p_j} \right) \tag{2.83}$$

where $Y(.)$ are random variables which the observed responses are assumed to be. Recall from the Kriging formulation that the variance of the basis function can be controlled for each of the k dimensions of the design space by θ_j and the exponent can be varied (again in each dimension) by p_j. However, now that we have two correlations, ψ_{LF} and ψ_d, we have a corresponding increase in the number of parameters to be estimated, i.e. $\boldsymbol{\theta}_{LF}, \boldsymbol{\theta}_d, \mathbf{p}_{LF}, \mathbf{p}_d$ and the scaling parameter ρ.

The low-fidelity data is assumed independent of the high-fidelity data, thus the MLEs for $\mu_{LF}, \sigma^2_{LF}, \boldsymbol{\theta}_{LF}$ and $\mathbf{p}_{LF}$ can be found by maximising the concentrated ln-likelihood:

$$-\frac{n_{LF}}{2}\ln(\hat{\sigma}^2_{LF}) - \frac{1}{2}\ln\|\det(\boldsymbol{\Psi}_{LF}(\mathbf{X}_{LF}, \mathbf{X}_{LF}))\| \tag{2.84}$$

where

$$\hat{\sigma}^2_{LF} = (\mathbf{y}_{LF} - \mathbf{1}\hat{\mu}_{LF})^T \boldsymbol{\Psi}_{LF}(\mathbf{X}_{LF}, \mathbf{X}_{LF})^{-1}(\mathbf{y}_{LF} - \mathbf{1}\hat{\mu}_{LF})/n_{LF} \tag{2.85}$$

To estimate $\mu_d, \sigma^2_d, \boldsymbol{\theta}_d, \mathbf{p}_d$ and ρ, define:

$$\mathbf{d} = \mathbf{y}_{HF} - \rho\, \mathbf{y}_{LF}(\mathbf{X}_{HF}) \tag{2.86}$$

where $\mathbf{y}_{LF}(\mathbf{X}_{HF})$ are the values of the low-fidelity response $\mathbf{y}_{LF}$ at the sample locations common to the high-fidelity sample set $\mathbf{X}_{HF}$ (as implied by the Markov property). Note: if the low-fidelity response cannot be evaluated at the high-fidelity sample locations, a Kriging estimate of the low-fidelity response $\hat{\mathbf{y}}_{LF}(\mathbf{X}_{HF})$ (using the already determined model parameters $\boldsymbol{\theta}_{LF}$ and $\mathbf{p}_{LF}$), may be used at an additional training cost. The concentrated ln-likelihood of the high-fidelity data can now be written as:

$$-\frac{n_{HF}}{2}\ln(\hat{\sigma}^2_d) - \frac{1}{2}\ln\|\det(\boldsymbol{\Psi}_d(\mathbf{X}_{HF}, \mathbf{X}_{HF}))\| \tag{2.87}$$

where

$$\hat{\sigma}^2_d = (\mathbf{d} - \mathbf{1}\hat{\mu}_d)^T \boldsymbol{\Psi}_d(\mathbf{X}_{HF}, \mathbf{X}_{HF})^{-1}(\mathbf{d} - \mathbf{1}\hat{\mu}_d)/n_{HF} \tag{2.88}$$

Again similarly to Kriging, both ln-likelihood functions must be numerically maximised, which may result in the computational cost of the Co-Kriging approach to impractical for higher-dimensional problems (Guo et al., 2022). For large values of dimensions k and/or samples n, constant values of $\theta_{LF,j}$ and $\theta_{d,j}$ may be used to reduce the time required to evaluate the MLEs, though naturally this will affect the quality of the resulting approximation. Finally, the Co-Kriging predictions are given by:

$$\hat{y}_{HF}(\mathbf{x}) = \hat{\mu} + \mathbf{c}^T\mathbf{C}^{-1}(\mathbf{y} - \mathbf{1}\hat{\mu}) \tag{2.89}$$

where

$$\hat{\mu} = \mathbf{1}^T\mathbf{C}(\mathbf{X}, \mathbf{X})^{-1}\mathbf{y}/\mathbf{1}^T\mathbf{C}(\mathbf{X}_{HF}, \mathbf{X}_{HF})^{-1}\mathbf{1} \tag{2.90}$$

and $\mathbf{c}$ is a column vector of the covariance between $\mathbf{X}$ and $\mathbf{x}$ (Forrester et al., 2007). Eq. (2.89) can be seen as an interpolator of the high-fidelity data (as in ordinary Krging), but will regress the low-fidelity data unless it coincides with $\mathbf{y}_{HF}$.

Co-Kriging is suitable for more complex corrections as long as the discrepancy function Z_d is of a simpler form (i.e. easier to model) than that of Z_{HF}. Additionally, with the ability to incorporate error estimates into the modelling process, Co-Kriging is generally seen as more effective than simple combinations of independent discrepancy-based surrogates (Forrester and Keane, 2009). The Co-Kriging methodology can be extended to incorporate more than two levels of fidelity, with corresponding cost increase due to the additional parameters. The reader is referred to (Forrester et al., 2007) for more details.

2.3.4 Filtering

Filtering refers to a hierarchical management of models, with no explicit generation of a multi-fidelity surrogate. The 'multi-fidelity' aspect then is therefore the manner that model surrogates are constructed (if at all) and/or applied in order to obtain appropriate cost savings (Fernández-Godino et al., 2016). Hierarchical management may take the form of evaluating the high-fidelity model only when the low-fidelity approximation falls outwith a specified acceptable accuracy/applicability range (Choi et al., 2005; Singh and Grandhi, 2009), or alternatively, when the candidate point meets some specified criterion based on the low-fidelity evaluation (Narayan et al., 2014).

Examples include the multistage *Markov-Chain Monte-Carlo* (MCMC) algorithm (Andrés Christen and Fox, 2005), in which the high-fidelity model may only evaluate a candidate sample following the acceptance of that sample by a likelihood criterion determined by the low-fidelity model. Another example is multi-fidelity importance sampling (Peherstorfer et al., 2015), where the low-fidelity model is used to construct an importance-based biasing distribution to guide the sampling of the high-fidelity model. In (Rethore et al., 2013), the majority of the optimisation process is performed using simpler/faster cost functions and a coarse resolution of the design domain. The domain resolution and model complexity is subsequently increased where/when needed.

2.3.4.1 Multi-level methods

A traditional approach to hierarchical model management is the use of multilevel methods (Peherstorfer et al., 2018). Included in this categorisation are multi-grid methods (Brandt, 1977; Briggs et al., 2000), multilevel pre-conditioners (Bramble et al., 1990; Dahmen and Kunoth, 1992) and multilevel function representations (Bank et al., 1988; Daubechies, 1998; Bungartz

and Griebel, 2004). Such methods typically switch between different levels of fidelity[14] by varying a certain parameter. For instance, if the parameter were mesh density, the hierarchy would consist of a series of coarse-grid approximations, successively increasing in resolution level (Peherstorfer et al., 2018).

By definition, a hierarchy may often necessitate the *ranking* of those low-fidelity models constituting the series. Such a ranking may be determined via metrics based on approximation quality and/or cost, for instance. A common approach is to define a relative ranking of the model hierarchy using *rates*, then to use these rates to distribute work accordingly. However, using rates only may be too coarse a description and may entirely miss relevant behavioural aspects of the model(s) in question (Peherstorfer et al., 2018). In a more general sense, a major difficulty is that information regarding accuracy and cost metrics may not always be readily available. Finally, in many applications, the low-fidelity modelling approaches may be limited, or even fixed, for a give scenario. In these cases, they cannot therefore be easily generated by simply varying a parameter (e.g., discretisation).

A critical observation with regards to the hierarchical optimisation of models of varying fidelity is that an improvement detected via the low-fidelity model does not guarantee an improvement relative to the high-fidelity equivalent (Robinson et al., 2008a).

2.4 Multi-Fidelity Surrogate-Based-Optimisation

The fundamental goal of optimisation is to determine the *input* that leads to an optimal *output* with respect to a specific *objective* (performance) function (Peherstorfer et al., 2018). Numerical methods to solve optimisation problems involve an iterative evaluation of the various constituent functions defining the problem (including the objective function(s) and any/all system/sub-system models). Consider the following non-linear minimisation problem representing a simulation-driven design process:

$$\mathbf{x}^* = \arg\min_x f(\mathbf{x}) \tag{2.91}$$

where $f(\mathbf{x})$ denotes the objective function to be minimised, evaluated at the design point $\mathbf{x} \in R^n$. The function f is typically of the form $f(\mathbf{x}) = U(\mathbf{R}_f(\mathbf{x}))$, where $\mathbf{R}_f \in R^m$ denotes the response vector of the system of interest (in particular, one may have $m > n$ or even $m \gg n$ (Bandler et al., 2004)), whereas U is a given scalar merit function, defined through a norm that measures the distance between $\mathbf{R}_f(\mathbf{x})$ and a target vector $\mathbf{y}$. An optimal design vector is denoted by $\mathbf{x}^*$. In many cases, $\mathbf{R}_f$ is obtained through computationally expensive computer simulations. For problems in which this becomes prohibitive, approximation methods can be incorporated. Where the approximation is centred on the problem constituent models (as

[14] More commonly thought of as a hierarchy of low-fidelity approximations of the single high-fidelity function.

opposed to the mathematical formulation of the problem, for example), the general trends of lower-fidelity, computationally less demanding surrogate models can be leveraged to accelerate the outer-loop process. A successful SBO method reaches a solution that satisfies the optimality conditions associated with the original, high(er)-fidelity problem within a significantly reduced computational budget (Simpson, 2008).

The simplest form of this approach is to construct a fixed surrogate of the time consuming function(s) in advance of the optimisation process using some suitable combination of sampling scheme (Section 2.2) and modelling technique (Section 2.1) (Jiang et al., 2020). The optimisation is then performed using only the pre-constructed surrogate, thus the accuracy of the final solution depends on the accuracy of the surrogate(s). Since the optimal solution is most often considered unknown *a priori*, each surrogate must guarantee a high prediction accuracy across the entire design space, thus ensuring the corresponding solution will lie at least in the vicinity of the true optimal solution. Consequently, this approach requires the evaluation of a large number of sample points using the expensive target function, which may not always be computationally practical, or even possible. Indeed, many earlier attempts to apply ROM-based surrogates in optimisation were almost exclusively dependant on such *offline* sampling processes, naturally resulting in extremely high computational costs (Benner et al., 2015; Yondo et al., 2018).

A more efficient methodology, constituting the general application of SBO algorithms, is an *offline-online* (Yondo et al., 2018) approach to iteratively update the surrogate approximations during the optimisation process. Each surrogate is supplemented with additional samples of the expensive function(s), selected at each iteration according to information derived from the current optimal solution. Sample selection is most often performed to balance the relationship between global exploration (sampling in areas with few existing sample points) and local exploitation (sampling near the current optimal solution), where the point with most potential to improve the current surrogate(s) (according to some specified metric) is selected. This approach in general aims to maximise the information gained from each subsequent sample, thus typically requires fewer total function evaluations to converge upon the optimal solution of the original problem.

A standard, sequential SBO framework can be formulated as follows: (Koziel and Yang, 2011):

$$\mathbf{x}^{(i+1)} = \arg\min_{x} s^{(i)}(\mathbf{x}) \tag{2.92}$$

where the sequence of design points $\mathbf{x}^{(i)}$ iteratively converges to a solution (or acceptable approximation thereof) of the design problem (2.91). Each design $\mathbf{x}^{(i+1)}$ is the optimal solution evaluated using the surrogate model $s^{(i)}$, itself assumed to be a computationally cheap and sufficiently reliable representation of the expensive model f (particularly in the neighbourhood of the current design $\mathbf{x}^{(i)}$). The target function (the expensive function to be approximated)

response is typically evaluated once per iteration (i.e. at every new design $\mathbf{x}^{(i+1)}$), with the resulting data used to validate and/or update the current approximation until some acceptable accuracy criteria has been met. Under the assumption that the surrogate model is computationally inexpensive, the cost associated with (2.92) can be viewed as negligible.

If the surrogate model is continuous and smooth (Echeverría and Hemker, 2008), satisfies zero- and first-order consistency conditions with the target function i.e., $s^{(i)}(\mathbf{x}^{(i)}) = f(\mathbf{x}^{(i)})$ and $\nabla s^{(i)}(\mathbf{x}^{(i)}) = \nabla f(\mathbf{x}^{(i)})$ (Alexandrov et al., 2001), (the verification of the latter requires model sensitivity data), and the surrogate-based algorithm is enhanced by, for example, a trust region method (Conn et al., 2000), then the sequence of intermediate solutions is provably convergent to a local optimum of the target function (Alexandrov et al., 1997). The generalised approach toSBO can therefore be summarised as follows:

1. Initial sampling of the associated design space(s).
2. Generation of initial surrogates using some appropriate modelling technique(s).
3. Optimisation using the constructed surrogates.
4. New points at which to evaluate data are identified according to some specified infill criteria.
5. Surrogates are updated with new information.
6. Evaluation of accuracy/convergence criteria. If not met, return to step 4.

Within this framework, numerous options are available to the designer. For instance, the initial DoE sampling procedure used to construct the initial surrogate(s) (Section 2.2), the selection of a suitable surrogate modelling technique (Section 2.1), the selection of the optimisation methodology to employ (Section 1.2) and an infill criterion with which to adaptively update surrogates with new information (Section 2.2). These decisions may be dealt with individually, or collectively as part of a Model Management Framework (Section 2.3). Numerous works reviewing the fields of surrogate modelling, optimisation frameworks and their associated processes can be found in (Forrester and Keane, 2009; Fernández-Godino et al., 2016; Koziel and Yang, 2011; Simpson, 2008; Jiang et al., 2020; Jones, 2001; Liu et al., 2018a; March, 2012; Park et al., 2017; Peherstorfer et al., 2018; Robinson et al., 2008b; Viana et al., 2014; Yondo et al., 2018).

Many outer-loop and (by definition) many-query applications such as optimisation, inference, and uncertainty quantification require multiple evaluations of problem constituent functions at many alternative input points. This often leads to computational demands that exceed available resources, particularly if only the high-fidelity model is used. Despite a wealth of surrogate

modelling and optimisation techniques to improve efficiency such that this hindrance is minimised, a remaining issue (particularly for physics-based design and optimisation problems) is that the computational cost of constructing surrogates of an expensive function only through direct sampling can nonetheless remain extremely prohibitive (Zhou et al., 2018). This is compounded if underlying non-linearities, discontinuities and multi-modality necessitate a higher number of samples to provide a useful approximation. Given the computational benefits associated with the application of multi-fidelity methods to surrogate modelling, the incorporation of multiple sources of information to support a surrogate based optimisation process can further alleviate the numerical burden of one or more expensive constituent functions.

Recall the fundamental premise of multi-fidelity surrogate modelling: to inexpensively supply functional trends using one or more lower-fidelity models with only occasional recourse to the high-fidelity model to establish global/local accuracy, apply some correction or adjustment or guarantee outer-loop convergence to the high-fidelity optimum (Ng and Willcox, 2014; Alexandrov et al., 2001; Koziel and Yang, 2011). In this manner, one may attempt to somewhat circumvent the curse of dimensionality associated with black-box surrogate modelling by exploiting domain-specific knowledge (Simpson, 2008; Keane and Nair, 2005). Multi-fidelity modelling methods have been widely applied in surrogate-based optimisation (Peherstorfer et al., 2018; Fernández-Godino et al., 2016; Koziel and Leifsson, 2016; Huang et al., 2013). Approaches include utilising low-fidelity models to accelerate the search for the global optimum (Booker et al., 1999; Keane, 2003; Forrester and Keane, 2009), leveraging a low-fidelity model with adaptive corrections within a *Trust Region Model Management* (TRMM) scheme (Alexandrov et al., 1997, 2001; Booker et al., 1999; March and Willcox, 2012a), or directly optimising a multi-fidelity approximation of the objective function.

A critical component of the characterisation of multi-fidelity methods for outer-loop problems used within this section is the use of explicit low-fidelity models that approximate the same output quantity as the high-fidelity model. This distinguishes from methods that embed approximations within the solver itself (e.g. the mathematical description of the problem), such as the quasi-Newton and Krylov subspace methods (Peherstorfer et al., 2018).

2.4.1 Classifications and Terminology

In the current literature, the wide applicability of surrogate modelling and multi-fidelity methods across numerous fields has naturally given rise to a certain degree of mixed terminology and conflicting classifications. For the purposes of this thesis, and to suit a wider mathematical context, it is necessary to specify certain terminology and how it shall be employed in the following discussion.

The term *Surrogate Based Optimisation* (SBO) is often used an an umbrella term to describe frameworks incorporating both single- and multi-fidelity methods, often with little or no distinction given to whether this refers to the surrogate approximations or the overriding framework itself. Relevant to this work, the fields of multi-fidelity surrogate modelling and SBO therefore exhibit a degree of cross-pollination in associated terminology. In an effort to simplify the following discussion, the terms *single-* and *multi-fidelity* are employed to describe both the underlying surrogate modelling approach and the outer-loop/many-query SBO process.

Consider specifically the field of surrogate modelling. Methods can be broadly distinguished according to the number of data-sources[15] available to the surrogate construction process. A surrogate constructed from one single source of information is therefore referred to as *single-fidelity*. This single source may be of any fidelity level of relevance to the designer, whether some inexpensive approximation based upon simplified assumptions, reduced-order modelling or simple data-fitting (Section 2.3.1), or some computationally intensive simulation intended to represent the underlying functional process to as fine a degree as is practical/possible. Accordingly, a surrogate instead incorporating information from more than one distinct source, often of alternative degrees of numerical complexity, is referred to as *multi-fidelity*. Multi-source information may be incorporated into a single surrogate by means of either *adaptation* (Section 2.3.2) or *fusion* (Section 2.3.3) methods for multi-fidelity modelling, or employed within a *filtering* (Section 2.3.4), hierarchical architecture.

The terms *single-* and *multi-fidelity* may have slightly different meaning when describing a SBO process. For this work, *single-fidelity SBO* is defined as an SBO process where the constructed surrogates represent only one level of fidelity available to the optimiser. Similarly, where the constructed surrogates instead represent multiple levels of fidelity available to the optimiser, the overriding process is defined as *multi-fidelity SBO*. The muddling of terminology becomes most apparent when considering that a surrogate based optimisation process, either single or multi-fidelity, can be performed using surrogate models that are themselves either single or multi-fidelity. Therefore, the majority of approaches relevant to the following discussion can be classified as one of four generalised descriptions:

1. Single-Fidelity SBO with single-fidelity surrogates: replacing an expensive function with an offline/online surrogate. If online, the current optimum is used to update the surrogate at each iteration.

2. Single-Fidelity SBO with multi-fidelity surrogates: replacing an expensive function with offline/online multi-fidelity surrogates. If online, the current optimum is used to inform

[15] If one data-source is truly unique compared to another, it is assumed that there must be some inherent difference in modelling fidelity (or equivalent thereof), even if slight.

an adaptive surrogate correction process at each iteration.

3. Multi-Fidelity SBO with single-fidelity surrogates: optimising a single-fidelity surrogate of an expensive function, where the optimiser controls/switches the level of fidelity (e.g. hierarchical, trust-region etc).

4. Multi-Fidelity SBO with multi-fidelity surrogates: optimising a multi-fidelity surrogate of an expensive function, where the optimiser controls the level of fidelity of further adaptive sampling used to update the surrogate.

For the purposes of discussion, we again adopt the classifications offered by Peherstorfer et al. (2018), who categorise multi-fidelity optimisation approaches into either *global* or *local* methods. Recall from Section 1.2 the general distinction between global and local optimisation methods; global methods aim to search the complete design domain $\mathcal{Z}$ for the global optimum $\mathbf{z}^*$, whereas local methods typically locate only the nearest local optimum within the neighbourhood of the initial guess $\mathbf{z} \in \mathcal{Z}$, most often via fast, gradient-based searching methods. The use of gradient-based algorithms similarly defines local multi-fidelity optimisation approaches[16], where their incorporation often leads to a more efficient search process and generally is more scalable to problems with higher-dimensional input. Global multi-fidelity methods do not typically rely on gradient-based information, allowing them a particular advantage in scenarios where the target function of interest is given as a black-box, or where gradient approximations are subject to a large amount of inherent noise (Peherstorfer et al., 2018). Of note, the multi-fidelity solution of constrained optimisation problems can be achieved using penalty functions to convert the constrained problem into an unconstrained one, although some methods admit more sophisticated approaches with which to approximate the constraint functions (Peherstorfer et al., 2018).

Another important consideration is that optimisation using surrogates offers no guarantees that the solution is even close to the true optimum. In practice, SBO proceeds sequentially with the surrogate-predicted optimum point evaluated using the high-fidelity model, the surrogates then refit using the new data, and the optimisation performed again until convergence is achieved. This supports why sequential sampling frameworks have been the target of considerable research. For example, *Efficient Global Optimisation* (EGO) (Jones et al., 1998) (discussed further in Section 2.4.2.1) and its derivatives (Viana et al., 2014).

2.4.2 Global Methods

As with global optimisation techniques and global sampling strategies, a critical requirement of global multi-fidelity optimisation is to strike an appropriate balance between *global exploration*

[16]Though some local methods do not require gradient information.

and *local exploitation*. In the multi-fidelity context, this can refer respectively to adapting the low-fidelity response with information gathered from high-fidelity evaluations (exploration), and actively locating the optimum of the included objective function(s) using the low/multi-fidelity approximation (exploitation). A careful distinction should also be made here specifically considering global multi-fidelity *modelling* (rather than optimisation) where the goal is to accurately represent the entire high-fidelity response using an inexpensive surrogate model incorporating response information from multiple sources.

A large number of global multi-fidelity optimisation methods incorporate *adaptation*-based approaches as their multi-fidelity modelling strategy of choice (Peherstorfer et al., 2018). Applying this management strategy within an optimisation context usually takes the general form of minimising the problem objective according to an adaptively refined low-fidelity model, where high-fidelity information is additionally used to guarantee the resulting optimal solution is also optimal with respect to the original, expensive, target function.

However, global methods incorporating multi-fidelity *fusion* and *filtering*-based management strategies have also been investigated (Queipo et al., 2005). The work by Allaire and Willcox (2014) combines response information gathered from a hierarchy of multiple Kriging approximations. In this case, the random variables defining each level of the hierarchy are fused together using the technique introduced by Winkler (1981). An *ensemble of meta-models* (Section 2.3.3.1) approach is employed by Goel et al. (2007), where a weighted average of the included low-fidelity models is used to perform the outer-loop optimisation. Weights are derived using the respective errors of the low-fidelity models individually. The pattern-search *Surrogate Management Framework* (SMF) presented by Booker et al. (1999) uses a hierarchy of low-fidelity models to inform the search directions, whilst ensuring convergence to the high-fidelity optimum. This method is discussed further in Section 2.4.3.3. Similarly, the work by Taddy et al. (2009) introduces a Bayesian approach to pattern-search with GP models. In this case, the search is again guided by the low-fidelity approximations.

Another interesting approach is to partition the input domain into several regions, where each is modelled using separate, local, Gaussian-process approximations (Gramacy and Lee, 2008, 2009). This 'treed' implementation is particularly suited cases where sampling the high-fidelity target function entails large-scale, parallel computations.

2.4.2.1 Efficient Global Optimisation (EGO)

A method that has gained particular popularity in recent years is *Efficient Global Optimisation* (EGO) (Jones et al., 1998; Kleijnen and Van Beers, 2004; Schonlau, 1997). This method employs an exploitation/exploration balancing strategy to allow not only the global searching of the input domain, but also the ability to locally refine the objective function surrogate in promising

regions using the *Expected Improvement* (EI) criteria. Such areas are processed using separate, localised optimisation sub-problems and as such offer a number of efficiency benefits by virtue of parallelisation.

Similarly to other approaches, EGO sequentially searches for the global high-fidelity optimum whilst iteratively updating a Kriging surrogate of the objective function with additional (assumed expensive) information until convergence is achieved. EGO is often combined with the EI criteria to inform the sequential sampling procedure and infer convergence. That is, the stopping criteria can be defined as when the maximum value of EI is less than some absolute tolerance, or alternatively, the ratio of the maximum EI to the current best solution is less than a specific relative tolerance.

Assume the input z minimises the low-fidelity response f_{LF} at the current iteration. A new point $z' \in \mathcal{Z}$ is sampled by evaluating z' using the high-fidelity target function $f_{HF}(z')$. The output response gained from this evaluation is subsequently used to update the low-fidelity model. EGO primarily distinguishes itself from other approaches via its selection methodology of the new sample point z'. Prior to evaluation, the high-fidelity response $f_{HF}(x')$ is assumed uncertain. This uncertainty is modelled as a Gaussian random variable Y with mean and standard deviation supplied by the low-fidelity Kriging approximation. The improvement at point z' is given by $I(z') = \max\{f_{HF}(z') - Y, 0\}$, thus the expected improvement can be written as:

$$\mathrm{E}[I(z')] = \mathrm{E}[\max\{f_{HF}(z') - Y, 0\}] \tag{2.93}$$

The expected improvement at point z' can be inexpensively evaluated by exploiting the low-fidelity Kriging model. Specifically the in-built computation of the MSE, and therefore the standard deviation, of the approximation (Jones et al., 1998; Forrester and Keane, 2009). The complete optimisation loop can be summarised as follows:

- Construct an initial Kriging approximation
- Select the new sample point z' that maximises the expected improvement
- Evaluate the high-fidelity model at the newly selected point z'
- Update the Kriging model with the new information
- Repeat until the stopping criteria is met

EGO can be made globally convergent and does not require high-fidelity model derivatives (Jones, 2001). However, the authors of (Jones, 2001; March, 2012) point out that EGO remains sensitive to the samples used to construct the initial approximation. This suggests that a fairly exhaustive search within the vicinity of the initial points may be necessary prior to the global

search. It has been suggested (Peherstorfer et al., 2018) that this behaviour may be due to the fact that the optimisation is performed by assuming the Kriging approximation is 'correct' at the current iteration, thus earlier iterations (particularly the initial iterations) may have a disproportionate effect on the overall search.

Numerous improvements to the base EGO algorithm have been proposed to address these effects (Jones, 2001; Forrester and Keane, 2009). For example, the work (Rajnarayan et al., 2008) builds upon EGO into the context of multi-objective, gradient-free optimisation. Later work (Rajnarayan, 2009) incorporates the consideration of improvement and risk when selecting new points. In (Chaudhuri and Haftka, 2014), the EGO algorithm is equipped with an adaptive target method that uses information from previous iterations to learn the extent of the improvement to expect in subsequent iterations.

2.4.3 Local Methods

Adaptation remains a popular model management strategy for *local* multi-fidelity optimisation approaches. Knill et al. (1998) state that their low-fidelity adaptive correction procedure applied within a gradient-based optimisation scheme was effective in reducing the overall computational cost. However, befitting a gradient-based method, their approach was negatively affected by the the presence of numerical noise in the aerodynamic and structural responses. An attempt to circumvent this issue was offered by Giunta et al. (1995), where a thorough design space exploration was first performed using a low-fidelity model. Using the gathered information, "nonsense" regions were identified and promptly excluded from the problem design domain. Following these initial steps, a polynomial RSM of the high-fidelity simulation was constructed across the remaining "ribbon-like" sub-domain of the original design space, requiring a significantly reduced number of samples. This approach was developed further by incorporating a polynomial RSM of the correction function, using a relatively small number of high-fidelity function evaluations (compared to the number of samples used for low-fidelity evaluations). Using this data, the low fidelity model was modified by applying the correction to it's polynomial approximation (Mason et al., 1998) or to it's associated training data. The authors of (Venkataraman et al., 1998) demonstrated the effectiveness of correcting inexpensive, low-fidelity analyses using more expensive, accurate models for the design of shell structures for buckling.

A general approach is to perform the optimisation only using the surrogate, but evaluate the high fidelity function at the location of the obtained optimum. The information gained from this evaluation is then used to adapt the surrogate approximation, and the process is repeated until convergence. Convergence itself can be mathematically guaranteed in specific circumstances (Barthelemy and Haftka, 1993; Queipo et al., 2005; Forrester et al., 2008). A variation of this approach is to apply the surrogate adaptation only if the computed point lies

within a specified trust-region. This fundamental idea has since expanded to include an entire class of methods, which is described in more detail in the following section.

2.4.3.1 Trust Region Methods

Perhaps the most widely accepted class of multi-fidelity optimisation methods, Trust-Region methods solve optimisation problems using the low-fidelity model (or some form thereof), but restrict the size of the search step to lie within a region of acceptability, known as the "trust region".

Trust Region methods were originally developed in the field of *Non-Linear Programming* (NLP) (Conn et al., 2000). A classical example is to derive a quadratic approximation of the high-fidelity model (most often, the objective function), where the model gradient and Hessian are defined at the centre of the trust-region. The trust region may then be expanded or contracted depending on the approximation quality of the quadratic (Conn et al., 2000). Early formulations can be found in the papers by Levenberg (1944) and Marquardt (1963), with the papers by Powell (1968, 1970, 1975) establishing and proving convergence of the trust-region concept.

The first applications of this ideology to surrogate-based optimisation were given by Alexandrov et al. (1997, 2001). In this work, the authors introduce the *Approximation Model Management Optimisation* (AMMO) framework, in which the trust-region methodology for the numerical optimisation of NLP problems was combined with an adaptive low-fidelity multiplicative linear correction step (see Section 2.3.2.1). Importantly, this work contains the formulation of the first-order consistency requirement, which states that (as in the case of conventional mathematics) if a low- and high-fidelity approximation have equal values of gradient at the centre of the trust region, the algorithm is provably globally convergent to a first-order stationary point of the original (high-fidelity) optimisation problem (Alexandrov et al., 1997). Eldred et al. (2004) showed that if second order consistency is additionally satisfied, convergence rates can be improved. Second-order correction methods were implemented in the DAKOTA software (Adams et al., 2022) and compared to other surrogate modelling techniques in (Eldred et al., 2004).

A general trust-region algorithm may be summarised in Algorithm 0: The parameters η_1 and η_2, define the size (maximum and minimum, respectively) of the trust region, and γ_1 and γ_2 control when to modify (contract and expand, respectively) the trust region. The value Δ^* is the upper bound of the step size. For each iteration i, a low-fidelity model is constructed such that the zero- and first-order requirements are met. That is, both low- and high-fidelity models produce the same output and have the same gradient at the current point z_i (Alexandrov et al., 1997). It is shown in (Alexandrov et al., 1997) that an additive or multiplicative correction is sufficient to correct an arbitrary low-fidelity model such that these requirements are met.

Algorithm 1 A Multi-Fidelity Trust Region Algorithm.

1: **procedure** TRUSTREGION($\mathbf{f}_{HF}, \eta_1, \eta_2, \gamma_1, \gamma_2, \Delta^*$)
2: Choose a starting point z_1 and initial step size (trust-region radius) $\Delta_1 > 0$
3: **for** i=1,2,3,... until convergence **do**
4: Construct a low-fidelity model $f_{LF}^{(i)}$

$$f_{LF}^{(i)}(z_i) = f_{HF}(z_i), \quad \text{and} \quad \nabla f_{LF}^{(i)}(z_i) = \nabla f_{HF}(z_i)$$

5: Find $s_i \in \mathcal{Z}$ that solves

$$\begin{aligned} &\min f_{LF}^{(i)}(z_i + s_i) \\ &\text{s.t. } \|s_i\| \leq \Delta_i \end{aligned}$$

6: Update point

$$z_{i+1} = \begin{cases} z_i + s_i & \text{if } f_{HF}(z_i + s_i) < f_{HF}(z_i) \\ z_i & \text{otherwise} \end{cases}$$

7: Compare actual and estimated decrease in f_{HF} :

$$\gamma = \frac{f_{HF}(z_i) - f_{HF}(z_i - s_i)}{f_{HF}(z_i) - f_{LF}^{(i)}(z_i - s_i)}$$

8: Update step size

$$\Delta_{i+1} = \begin{cases} \eta_1 \|s_i\|, & \gamma < \gamma_1 \\ \min\{\eta_2 \Delta_i, \Delta^*\}, & \gamma > \gamma_2 \\ \|s_i\|, & \text{otherwise} \end{cases}$$

9: **end for**
10: **return** z_1, z_2, z_3
11: **end procedure**

The proposed step s_i can only be accepted if point $z_i + s_i$ decreases the high fidelity function value with respect to the current point z_i. The size of the trust region (i.e. the size of the step that may be applied) is thus evaluated based on the ratio of the actual decrease in f_{HF} relative to the decrease predicted by the scaled low-fidelity approximation. Step 5 may be adjusted to accommodate additional constraints. The authors offer a modification in (Alexandrov et al., 2001) for cases where the constraints are expensive to evaluate, thus are also approximated using surrogates.

An application of the AMMO methodology was presented in (Alexandrov et al., 2001) for the optimisation of a two-element aerofoil utilising a RANS code and an Euler code as high- and low-fidelity simulation tools, respectively. The authors reported a 55:1 run-time ratio (excluding sensitivity analysis performed by an adjoint approach), with a five-fold saving in

computational effort. Rodríguez et al. (1998) demonstrated that the multi-fidelity surrogates constructed within the CSSO MDO strategy can be effectively managed by a trust-region model management strategy. The authors provided a proof of convergence for the outer-loop SBO algorithm that has since been applied to MDO test problems. The authors then extend this work into a comparative study of alternative sampling strategies for surrogate modelling (Rodríguez et al., 2001). The authors of (Yue and Meerbergen, 2013) define the trust region using the error bounds of interpolative reduced-order models. These bounds are also used to trigger a model refinement process such that convergence to the high-fidelity optimum can be guaranteed. Fischer et al. (2018) correct the response of a low-fidelity model using a correction model within a trust-region model management optimisation framework. The correction model itself is only constructed within the trust-region, and specifically includes both additive and multiplicative corrections. Robinson et al. (2008a) present a trust-region approach for combining low- and high-fidelity models defined over different domains (i.e. with different numbers of design variables). The alternative design spaces are mapped using the Space Mapping approach (Bandler et al., 2003). This method is proved convergent following an additional correction step applied to the mapping.

An important consideration of the trust-region approach is that if gradient information is either unavailable or too costly to compute, a management framework relying on first-order consistency of the low-fidelity model (such as that of (Alexandrov et al., 1997)) cannot be applied. Accordingly, a gradient-free variation of the trust-region framework was presented by March and Willcox (2012b) for solving constrained optimisation problems. This approach built upon the gradient-free trust region methods of (Conn et al., 2009; Conn, 2009) and the tailored radial basis function modelling of (Wild and Shoemaker, 2013). Interpolating radial basis functions are employed as low-fidelity models, with a multi-fidelity approximation of the problem objective iteratively updated throughout the optimisation process. The resulting error interpolation allows the conditions to ensure convergence (to the high-fidelity optimum) be satisfied.

2.4.3.2 Space Mapping

Incorporating the Space-Mapping multi-fidelity modelling approach (Section 2.3.2.3) into an SBO paradigm involves optimising within the low-fidelity space, then inverting the computed mapping to determine a trial point in high-fidelity space. The high-fidelity evaluation of this trial point (and other in the near vicinity) is used to construct the mapping for the subsequent iteration, and so on.

As another example of a sequential approach, Space Mapping optimisation operates by identi-

fying the sequence of surrogates:

$$s^{(0)}(\mathbf{x}) = U(R_{LF}(\mathbf{x})) \tag{2.94}$$

$$s^{(i)}(\mathbf{x}) = U(R_{MF}(\mathbf{x}; \mathbf{p}_{SM}^{(i)})) \tag{2.95}$$

where $i > 0$ and the parameters $\mathbf{p}_{SM}$ are obtained via a parameter extraction procedure. As with other correction based methods, the quality of the resulting multi-fidelity surrogate is heavily dependent on the prediction quality of the low-fidelity model response (Koziel et al., 2006a). Bandler et al. (2003) have observed (as have others) that the number of low-fidelity model samples required to offer a suitable corrected multi-fidelity surrogate is of the order of the number of optimisation variables.

Space-Mapping has been integrated and developed into various alternative SBO frameworks and applied across a range of disciplines (Bandler et al., 2003; Viana et al., 2014). Leary et al. (2001) demonstrated the use of space-mapping for the structural optimisation of a simple beam. Redhe and Nilsson (2004, 2006) developed a multi-point space-mapping variation where high-fidelity responses are used to iteratively update the mapping, subsequently applied using response-surface methodology. The authors demonstrate this approach in application to a structural optimisation problem representing the crash-worthiness of a vehicle. Koziel and Leifsson (2010) offer a modification to the space-mapping approach incorporating a trust-region sampling scheme intended to improve convergence performance. In a similar manner to the AMMO algorithm, the optimisation of the space-mapping surrogate is confined (using the Euclidean norm) to the proximity of $\mathbf{x}^{(i)}$ as follows:

$$\mathbf{x}^{(i+1)} = \arg\min_{\mathbf{x}} s^{(i)}(\mathbf{x}) \tag{2.96}$$

$$s.t. \|\mathbf{x} - \mathbf{x}^{(i)}\|_2 \leq \boldsymbol{\delta}^{(i)} \tag{2.97}$$

where $\boldsymbol{\delta}^{(i)}$ is the radius of the trust-region at the i-th iteration, updated every iteration according to a specified criteria (Conn et al., 2000).

A notable modification of the space-mapping methodology incorporated into an SBO framework is *Aggressive Space Mapping* (ASM) (Bandler, 1995). ASM has been shown to descend towards the problem optimum much more quickly than standard output space-mapping, but requires the assumption that the mapping between the spaces is bijective[17], and that the input vector is of the same dimension in each space. Despite examples of successful application (primarily in microwave circuit design (Bakr, 1999), but also other disciplines (Bandler et al., 1994)), this method does not ensure first-order accuracy, thus the proofs of convergence (even to only local minima of the high-fidelity space) associated with trust-region methods do not apply here (Robinson et al., 2008b). However, the authors of (Madsen et al., 2004) presented

[17] A function $f : A \to B$ is said to be bijective if both the *injective* and *surjective* properties are satisfied. That is, for the function $f : A \to B$, where $a \in A$ and $b \in B$ the condition $f(a) = b$ is met.

a provably convergent variant in which the model surrogate is a convex combination of the space-mapped low-fidelity function and a Taylor-series approximation of the high-fidelity function.

Given that convergence may be ensured with any surrogate that is first-order accurate, one approach is to combine the space-mapping methodology with an adaptive correction step (Section 2.3.2) to force at least first-order accuracy. However, if the input parameters are first selected to match the output function at some number of control points and a correction of this kind is then applied, it is likely that the correction will unnecessarily distort the match performed in the space-mapping step (Robinson et al., 2008b). Robinson et al. (2008b) address this issue by embedding the the correction process within the space mapping algorithm, thus the mapping and correction may be performed simultaneously. In other words, using the remaining degrees of freedom following an adaptive correction, the input mapping is varied such that the best possible match to the control (high-fidelity) points is made. A *Corrected Space-Mapping* optimisation may be formulated as:

$$\mathbf{p}^k \in \arg\min_{\mathbf{p}} \sum_{i=1}^{q} \|\beta(\mathbf{x}^i) - \tilde{\beta}^k(P(\mathbf{x}^i, \mathbf{p}))\|_2 \tag{2.98}$$

where $\tilde{\beta}^k$ represents the corrected surrogate of the high-fidelity function for the k-th sub-problem and $\mathbf{x}^i$ denotes the i-th sample (or *control*) point. The selection of sample points and the functional form of the mapping P are designer-defined[18]. The correction must be updated for each new value of $\mathbf{p}$, given the dependence on the Jacobian and Hessian matrices of the mapping P (Robinson et al., 2008b).

One final mention here is *Manifold Mapping* (MM), itself a special case of *Output Space-Mapping* (OSM) supported by convergence theory and not requiring any parameter extraction step (Echeverría and Hemker, 2005, 2008). MM can be integrated into a SBO process by combining the surrogate model $s^{(i)}(\mathbf{x}) = U(\mathbf{R}_s^{(i)}(\mathbf{x}))$ with the following response correction step:

$$\mathbf{R}_s^{(i)}(\mathbf{x}) = \mathbf{R}_f^{(i)}(\mathbf{x}^{(i)}) + \mathbf{S}^{(i)}(\mathbf{R}_c(\mathbf{x}) - \mathbf{R}_c(\mathbf{x}^{(i)})) \tag{2.99}$$

where $i \geq 0$ and $\mathbf{S}^{(i)}$ is defined as a $m \times m$ matrix:

$$\mathbf{S}^{(i)} = \boldsymbol{\Delta}\mathbf{F}\boldsymbol{\Delta}\mathbf{C}^{\dagger} \tag{2.100}$$

where

$$\boldsymbol{\Delta}\mathbf{F} = [\mathbf{R}_f(\mathbf{x}^{(i)} - \mathbf{R}_f(\mathbf{x}^{(i-1)}), \ldots, \mathbf{R}_f(\mathbf{x}^{(i)} - \mathbf{R}_f(\mathbf{x}^{(\max\{i-n,0\})})] \tag{2.101}$$

$$\boldsymbol{\Delta}\mathbf{C} = [\mathbf{R}_c(\mathbf{x}^{(i)} - \mathbf{R}_c(\mathbf{x}^{(i-1)}), \ldots, \mathbf{R}_c(\mathbf{x}^{(i)} - \mathbf{R}_c(\mathbf{x}^{(\max\{i-n,0\})})] \tag{2.102}$$

[18]The authors of (Robinson et al., 2008b) employ a linear mapping $\hat{\mathbf{x}} = P(\mathbf{x}) = \mathbf{M}\mathbf{x} + \mathbf{b}$, where $\mathbf{M}$ is a matrix of $\hat{n} \times n$ elements and $\mathbf{b}$ is a vector of length $\hat{n}$ for a total of $\hat{n} \times (n+1)$ space-mapping parameters. The sample points are selected as the previous q steps in the trust-region managed algorithm $(\mathbf{x}^{k-q+1}, \ldots, \mathbf{x}^k)$, for which high-fidelity function values are already available.

The matrix $\mathbf{S}^{(0)}$ is generally assumed as an identity matrix $\mathbf{I}_m$. The superscript † represents the *pseudoinverse*[19], defined for $\mathbf{\Delta C}$ as:

$$\mathbf{\Delta C}^{\dagger} = \mathbf{V}_{\mathbf{\Delta C}} \sum\nolimits_{\mathbf{\Delta C}}^{\dagger} \mathbf{U}_{\mathbf{\Delta C}}^{T} \tag{2.103}$$

where $\mathbf{U}_{\mathbf{\Delta C}}$, $\sum_{\mathbf{\Delta C}}$ and $\mathbf{V}_{\mathbf{\Delta C}}$ are the factors in the single value decomposition of $\mathbf{\Delta C}$. The matrix $\sum_{\mathbf{\Delta C}}^{\dagger}$ is the result of inverting the non-zero entries in $\sum_{\mathbf{\Delta C}}$, leaving the zeros invariant (Echeverría and Hemker, 2005). The response correction $\mathbf{R}_s^{(i)}(\mathbf{x})$ is an approximation of:

$$\mathbf{R}_s^*(\mathbf{x}) = \mathbf{R}_f(\mathbf{x}^*) + \mathbf{S}^*(\mathbf{R}_c(\mathbf{x}) - \mathbf{R}_c(\mathbf{x}^*)) \tag{2.104}$$

with $\mathbf{S}^*$ defined as the $m \times m$ matrix:

$$\mathbf{S}^* = J_f(\mathbf{x}^*) J_c^{\dagger}(\mathbf{x}^*) \tag{2.105}$$

where $J_f(\mathbf{x}^*)$ and $J_c(\mathbf{x}^*)$ are the Jacobians of the fine and coarse responses respectively, each evaluated at $\mathbf{X}^*$. Given that neither $\mathbf{x}^*$ nor $\mathbf{S}^*$ can be known beforehand, an iterative approximation is used.

A MM model alignment problem can be expressed as a least-squares optimisation:

$$U(\mathbf{R}_f(\mathbf{x})) = \|\mathbf{R}_f(\mathbf{x} - \mathbf{y}\|_2^2 \tag{2.106}$$

where $\mathbf{y} \in R^m$ representing the given input specification. The low-fidelity model is assumed to require a negligible computational cost relative to the high-fidelity counterpart, allowing the MM surrogate to be explored globally. Considering global exploration, the MM algorithm includes some mitigation strategies (supported by convergence theory (Echeverría and Hemker, 2008)) for avoiding local minima. These strategies rely on four fundamental factors:

- Model responses are smooth (continuous).
- The surrogate optimisation problem is well-posed.
- The discrepancy of the optimal model response $\mathbf{R}_f(\mathbf{x}^*)$ with respect to the input specification is sufficiently small.
- The low-fidelity model is an acceptable approximation of the high-fidelity model.

The requirements corresponding to the first three points are generally satisfied in general application. Since low-fidelity models are often based on expert knowledge, the correlation between responses is often amenable to allowing convergence.

[19] A generalisation of the inverse matrix, see (Ben-Israel and Greville, 2003)

The MM algorithm demonstrates reliable convergence characteristics if the merit function U is sufficiently smooth. Since the response correction does not utilise the merit function U, MM can yield acceptable results even if U is not explicitly differentiable. Including gradient information has been seen to improve convergence (Echeverría and Hemker, 2008). Alternatively, incorporating a Levenberg-Marquardt strategy can be considered a convergence safeguard analogous to a trust-region approach (Hemker and Echeverría, 2007; Conn et al., 2000).

Various other modifications to the original Space-Mapping SBO methodology have been proposed in the literature, such as enforcing zero-order and approximate/exact first order consistency conditions (Koziel and Leifsson, 2010) and adaptively constrained parameter extraction (Koziel et al., 2010). However, although the traditional OSM SBO can achieve both the zero- and first-order consistency conditions, some of these alternative frameworks in practice fail to do so, and in some cases exhibit distinct convergence problems (Koziel et al., 2008; Robinson et al., 2008b). The selection of the most appropriate Space-Mapping (or thereof derived) technique can therefore often remain unclear to the designer (Koziel et al., 2008).

2.4.3.3 Surrogate Management Framework

The *Surrogate Management Framework* (SMF) (Booker et al., 1999) was the first mathematically rigorous formulations to manage the use of approximation models in optimisation. Developed collaboratively by researchers at Boeing, IBM, and Rice University, the SMF was motivated by increasing research into sequential techniques for sampling, building, and optimising approximation models. Methods such as move limits (Toropov et al., 1996) and trust region approaches (Alexandrov et al., 1997; Rodriguez et al., 1997) were applied to NLP problems, where the approximation was included as part of the mathematical formulation of the problem.

SMF is based on the pattern-search optimisation algorithm (Kolda et al., 2003). Pattern search optimisation is a derivative-free family of methods that explores the problem search space via a structured set of points (the pattern, or stencil) that is modified across subsequent iterations. Pattern search has been proven globally convergent to first-order stationary points. As an example, consider the pattern-search SMF presented in (Booker et al., 1999). Each iteration consists of two distinct steps: *search* and *poll*. A pattern of size Δ is centred at point $\mathbf{x}^{(i)}$ every iteration i. The search stage selects a small number of points from the search space (usually via a surrogate model, as in (Marsden et al., 2004)), and evaluates each using the cost function $f(\mathbf{x})$. If any of these trial points improve the cost function, the search is deemed successful and the current pattern is shifted to align it's centre accordingly. The search step then repeats. However, if no improvement is registered, the poll step is initiated. The poll step evaluates each point in the current pattern using $f(\mathbf{x})$. As before, if any improvement is detected, the poll is deemed successful and the pattern is translated. This time, if no improvement is registered, the iteration is declared unsuccessful and the stopping criteria (typically defined using the

pattern size Δ) is checked. If the process is to continue, the value of Δ is reduced and a new searching step is initiated.

The set of vectors formed by each pattern point and the pattern centre must be a *generating* (or *positive spanning*) *set* to guarantee convergence to a stationary point(Kolda et al., 2003; Marsden et al., 2004). A *generating set* for R^n consists of a set of vectors whose non-negative linear combinations span R^n. If centred at $\mathbf{x}(i)$ and $\nabla f(\mathbf{x}^{(i)}) \neq 0$, at least one vector in the generating set defines a feasible descent direction (Kolda et al., 2003). Thus, assuming $f(\mathbf{x})$ is continuous and Δ is of a suitable small value, one may expect the current pattern to offer an improved cost value $f(\mathbf{x}^{(i)})$. This property is crucial in proving convergence.

Pattern-search based algorithms are usually associated with a large number of function evaluations, especially compared to gradient-based alternatives. Naturally however, the use of surrogate responses within the searching stage have been shown to significantly reduce this dependence.

2.4.4 Stopping Criteria

A final consideration in the design of a SBO framework (multi-fidelity or otherwise) is the choice of an appropriate stopping criterion (Chaudhuri and Haftka, 2013). Fundamentally, the outer-loop optimisation process should be terminated at the instance where any additional high-fidelity function evaluations do not offer an improvement in surrogate quality, nor in the objective function response. This latter aspect can prove complicated when considering that in some cases, the objective function value computed using a less-accurate surrogate model may be superior to that of the real high-fidelity optimum.

Classical stopping criteria include simple limits on various aspects of the process expenditure. For instance, a specified maximum allowable computational budget, maximum number of function evaluations or maximum number of iterations. These limits cannot guarantee that the problem optimum will ever in fact be obtained, whether by slow convergence or computationally expensive iterations (particularly for large-scale problems), for example. For the EGO framework (Section 2.4.2.1), Jones et al. (1998) proposed a stopping criteria based on the EI metric. When the maximum value of EI is below an absolute tolerance, or, when the ratio of the maximum value of EI to the current best solution is below a certain relative tolerance, the sequential sampling process is terminated. Similarly, Chaudhuri and Haftka (2013) introduced a criterion that allows the outer-loop to continue only if a specified improvement in the objective function value is expected. In other examples, the stopping criterion has been shifted away from deterministic calculation, with (Zhang et al., 2013) presenting a support decision scheme where the designer has direct control throughout the process, and (Boender and Rinnooy Kan, 1987; Betro, 1991) introducing stochastic process based criteria.

2.5 Conclusion

This chapter has offered a brief introduction to several areas of research relevant to multi-fidelity surrogate modelling.

In Section 2.1, some prominent surrogate modelling techniques, such as *Radial Basis Functions* (RBFs), *Artificial Neural Networks* (ANNs) and Kriging, were introduced. Each method was discussed in terms of both it's mathematical formulation and implementation. The scope of discussion is limited to the general approximation of expensive non-linear functions defined by a finite set of input/output correspondence. The reason for this is twofold. First, it is of interest to the current study to examine techniques for multi-fidelity modelling in a general sense, without any explicit reliance on one surrogate modelling approach over another. Secondly, this allows design of a modular multi-fidelity surrogate management framework, where both surrogate modelling methods and multi-fidelity techniques can be easily swapped in/out independently for research purposes. Acknowledgement is made of methods that conversely utilise multi-fidelity information within the surrogate construction process itself, though further discussion of these approaches is reserved for later chapters. Also included is a brief discussion of various methods with which to numerically validate such approaches.

A common component of many surrogate modelling techniques is a user-specification of several hyper-parameters or settings that can be used to tune the resulting approximation to better represent a particular set of data, or a particular type of target function. This poses several hurdles for an autonomously adaptive surrogate management framework to overcome. First, the appropriate choice of hyperparameters may require *a priori* knowledge of the function under consideration, which may not always be available. Second, a correct choice of hyperparameters to model one subsystem of a complex system may not be the same as for another. Even with modern hyperparameter optimisation techniques, this nonetheless requires repeated (and potentially costly) extra computational effort.

Another issue posed against autonomously adaptive frameworks is the variable dependence of most surrogate modelling techniques on properties of the target function, such as dimensionality, non-linearity, multi-modality or presence of discontinuities.

Specifically considering continuous optimal control problems, it is highly desirable that system models are continuously differentiable given that a continuous/smooth representation of the dynamic system state is often required to reach feasible solutions. Common methods of generating continuously differentiable surrogates suited to multi-fidelity applications are Kriging (Sacks et al., 1989) and *Response Surface Models* (RSMs) (Myers and Montgomery, 1995). However, these methods often require the user-specification of hyper-parameters such as (in the case of Kriging) regression and correlation functions (including an initial

guess of the correlation function parameters), which may require *a priori* knowledge of the function under consideration. Whilst for physics-based subsystem models a certain degree of prior knowledge may be assumed, the most appropriate choice of hyper-parameters may not necessarily remain constant across separate subsystems. This work aims to implement a more generalised methodology (detailed in the following chapter), suitable for multiple models of unknown functional form within multiple subsystems.

Section 2.2 overviews several static and dynamic sampling methodologies with which a surrogate modelling process may be informed. A particular focus is made on the introduction and discussion of the balance between local exploitation and global exploration, though special considerations (e.g. serial vs. parallel infill, multi-response sampling and multi-objective sampling) relevant to multi-fidelity surrogate modelling are included. This section forms the theoretical basis for the sampling strategies implemented in the following chapter (specifically, Section 3.2.).

Section 2.3 introduces the fundamental premises behind common methods of simultaneously managing alternate levels of modelling fidelity. After an initial section detailing the definition and generation of "low-fidelity" models specifically, multi-fidelity methods are introduced using three primary classifications: adaptation, fusion and filtering based approaches. This discussion follows into the definition of several distinct multi-fidelity correction 'stages' used in the following chapters.

Finally, Section 2.4 introduces a number of approaches with which the above methodologies may be seamlessly integrated into high-level frameworks for numerical optimisation. This includes consideration of global vs. local accuracy as a primary driver (with examples in each case), and a brief description of several commonly used stopping criteria implemented within such frameworks. The next chapter thus collates these discussions and introduces the multi-fidelity management framework developed and implemented as part of this work.

2.6 Bibliography

B. M. Adams, W. J. Bohnhoff, K. R. Dalbey, M. S. Ebeida, M. S. Eldred, J. D. Jakeman, L. P. Swiler, J. A. Stephens, D. M. Vigil, and T. M. Wildey. DAKOTA, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.16 User's Manual. *Sandia Technical Report SAND2010-2183*, (Updated November 7, 2013 Dakota,), 2022.

T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.

N. Alexandrov, M. D. O. Branch, N. Larc, J. E. Dennis, R. Michael, L. Icase, and V. Torczon. A Trust Region Framework for Managing the Use of Approximation Models in Optimization. *Engineering*, (97), 1997.

N. M. Alexandrov, R. M. Lewis, C. R. Gumbert, L. L. Green, and P. A. Newman. Approximation and model management in aerodynamic optimization with variable-fidelity models. *Journal of Aircraft*, 38(6):1093–1101, 2001. ISSN 00218669. doi: 10.2514/2.2877.

D. Allaire and K. Willcox. A mathematical and computational framework for multifidelity design and analysis with computer models. *International Journal for Uncertainty Quantification*, 4(1):1–20, 2014. ISSN 21525099. doi: 10.1615/Int.J.UncertaintyQuantification.2013004121.

M. Alvarez and N. D. Lawrence. Sparse convolved Gaussian processes for multi-output regression. *Advances in Neural Information Processing Systems 21 - Proceedings of the 2008 Conference*, pages 57–64, 2009.

J. Andrés Christen and C. Fox. Markov chain Monte Carlo using an approximation. *Journal of Computational and Graphical Statistics*, 14(4):795–810, 2005. ISSN 10618600. doi: 10.1198/106186005X76983.

F. Aurenhammer. Voronoi Diagrams - A Survey of a Fundamental Data Structure. *ACM Computing Surveys (CSUR)*, 23(3):345–405, 1991.

V. Aute, K. Saleh, O. Abdelaziz, S. Azarm, and R. Radermacher. Cross-validation based single response adaptive design of experiments for Kriging metamodeling of deterministic computer simulations. *Structural and Multidisciplinary Optimization*, 48(3):581–605, 2013. ISSN 16151488. doi: 10.1007/s00158-013-0918-5.

M. Bakr, J. W. Bandler, K. Madsen, and J. Sondergaard. An Introduction to the Space Mapping Technique. 103(3):239–248, 2002. ISSN 00138703. doi: 10.1023/A.

M. H. Bakr. A hybrid aggressive space-mapping algorithm for em optimization. *IEEE Transactions on Microwave Theory and Techniques*, 47(12):2440–2449, 1999. ISSN 00189480. doi: 10.1109/22.808991.

M. H. Bakr, J. W. Bandler, K. Madsen, J. E. Rayas-Sanchez, and J. Sondergaard. Space mapping optimization of microwave circuits exploiting surrogate models. *IEEE MTT-S International Microwave Symposium Digest*, 3(June):1785–1788, 2000. ISSN 0149645X. doi: 10.1109/mwsym.2000.862325.

J. W. Bandler. Electromagnetic Optimization Exploiting Aggressive Space Mapping. 1995.

J. W. Bandler, R. M. Biernacki, S. H. Chen, P. A. Grobelny, and R. H. Hemmers. Space mapping technique for electromagnetic optimization. *IEEE Transactions on Microwave Theory and Techniques*, 1994.

J. W. Bandler, Q. S. Cheng, S. A. Dakroury, A. S. Mohamed, M. H. Bakr, K. Madsen, and J. Søndergaard. Space Mapping: The State of The Art. *SBMO/IEEE MTT-S International Microwave and Optoelectronics Conference Proceedings*, 52(1):951–956, 2003.

J. W. Bandler, Q. S. Cheng, N. K. Nikolova, and M. A. Ismail. Implicit space mapping optimization exploiting preassigned parameters. *IEEE Transactions on Microwave Theory and Techniques*, 52(1 II):378–385, 2004. ISSN 00189480. doi: 10.1109/TMTT.2003.820892.

R. E. Bank, T. F. Dupont, and H. Yserentant. The hierarchical basis multigrid method. *Numerische Mathematik*, 52(4):427–458, 1988. ISSN 0029599X. doi: 10.1007/BF01462238.

J. F. Barthelemy and R. T. Haftka. Approximation concepts for optimum structural design - a review. *Structural Optimization*, 5(3):129–144, 1993. ISSN 09344373. doi: 10.1007/BF01743349.

R. R. Barton. Metamodels for simulation input-output relations. 9(x):49–54, 1992.

B. Beachkofski and R. Grandhi. Improved distributed hypercube sampling. In *43rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, page 1274, 2002.

M. H. Beale, M. T. Hagan, and H. B. Demuth. *Neural Network Toolbox ™ User's Guide R2017b*. 2017. ISBN 0971732108.

A. Ben-Israel and T. N. Greville. *Generalized inverses: theory and applications*. Springer, 2003.

P. Benner, S. Gugercin, and K. Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Review*, 57(4):483–531, 2015. ISSN 00361445. doi: 10.1137/130932715.

G. Berkooz, P. Holmes, and J. L. Lumley. The proper orthogonal, decomposition in the analysis of turbulent flows. (1971), 1993. ISSN 00664189. doi: 10.1146/annurev.fluid.25.1.539.

B. Betro. A Stochastic Technique for Global Optimization. *Computers and Mathematics with Applications*, 21(6):127–133, 1991.

A. Bhosekar and M. Ierapetritou. Advances in surrogate based modeling, feasibility analysis, and optimization: A review. *Computers and Chemical Engineering*, 108:250–267, 2018. ISSN 00981354. doi: 10.1016/j.compchemeng.2017.09.017. URL https://doi.org/10.1016/j.compchemeng.2017.09.017.

J. Biehler, M. W. Gee, and W. A. Wall. Towards efficient uncertainty quantification in complex and large-scale biomechanical problems based on a Bayesian multi-fidelity scheme. *Biomechanics and Modeling in Mechanobiology*, 14(3):489–513, 2015. ISSN 16177940. doi: 10.1007/s10237-014-0618-0. URL http://dx.doi.org/10.1007/s10237-014-0618-0.

C. G. Boender and A. H. Rinnooy Kan. Bayesian stopping rules for multistart global optimization methods. *Mathematical Programming*, 37(1):59–80, 1987. ISSN 14364646. doi: 10.1007/BF02591684.

A. J. Booker, J. E. Dennis, P. D. Frank, D. B. Serafini, V. Torczon, and M. W. Trosset. A rigorous framework for optimization of expensive functions by surrogates. *Structural Optimization*, 17(1):1–13, 1999. ISSN 09344373. doi: 10.1007/BF01197708.

M. A. Bouhlel, N. Bartoli, A. Otsmane, and J. Morlier. Improving kriging surrogates of high-dimensional design models by Partial Least Squares dimension reduction. *Structural and Multidisciplinary Optimization*, 53(5):935–952, 2016. ISSN 1615-147X.

G. E. P. Box and N. R. Draper. *Empirical model-building and response surfaces*. John Wiley & Sons, 1987.

G. E. P. Box and K. B. Wilson. On the experimental attainment of optimum conditions. In *Breakthroughs in statistics: methodology and distribution*, pages 270–310. Springer, 1992.

J. H. Bramble, J. E. Pasciak, and J. Xu. Parallel Multilevel Preconditioners. *Mathematics of Computation*, 55(191):1, 1990. ISSN 00255718. doi: 10.2307/2008789.

A. Brandt. Multi-Level Adaptive Solutions to Boundary-Value Problems. *Mathematics of Computation*, 31(138):333, 1977. ISSN 00255718. doi: 10.2307/2006422.

W. L. Briggs, V. E. Henson, and S. F. Mccormick. A Multigrid Tutorial, 2nd Edition FOSLS/LL* View project Adaptive Algebraic Multigrid Methods View project SEE PROFILE. (January), 2000. URL https://www.researchgate.net/publication/220690328.

C. J. Brooks, A. I. Forrester, A. J. Keane, and S. Shahpar. Multi-fidelity design optimisation of a transonic compressor rotor. *9th European Conference on Turbomachinery: Fluid Dynamics and Thermodynamics, ETC 2011 - Conference Proceedings*, 2(2000):1267–1276, 2011. ISSN 24104833.

D. S. Broomhead and D. Lowe. Radial basis functions, multi-variable functional interpolation and adaptive networks. (4148), 1988.

C. G. Broyden. A Class of Methods for Solving Nonlinear Simultaneous Equations. *Mathematics of Computation*, 19(92):577, 1965. ISSN 00255718. doi: 10.2307/2003941.

M. D. Buhmann. *Radial basis functions: theory and implementations*, volume 12. Cambridge university press, 2003. ISBN 1139435248.

H. J. Bungartz and M. Griebel. Sparse grids. *Acta Numerica*, 13:147–269, 2004. ISSN 09624929. doi: 10.1017/S0962492904000182.

J. Burkardt, M. Gunzburger, and H. C. Lee. POD and CVT-based reduced-order modeling of Navier-Stokes flows. *Computer Methods in Applied Mechanics and Engineering*, 196(1-3): 337–355, 2006. ISSN 00457825. doi: 10.1016/j.cma.2006.04.004.

R. Caruana. Learning Many Related Tasks at the Same Time With Backpropagation. *NIPS 1994: Proceedings of the 7th International Conference on Neural Information Processing Systems*, pages 657–664, 1994.

K. J. Chang, R. T. Haftka, G. L. Giles, and P.-J. Kao. Sensitivity-based scaling for approximating structural response. *Journal of Aircraft*, 30(2):283–288, 1993.

S. Chaturantabut and D. C. Sorensen. Nonlinear model reduction via discrete empirical interpolation. 32(5):2737–2764, 2010.

A. Chaudhuri and R. T. Haftka. A stopping criterion for surrogate based optimization using ego. 72, 2013.

A. Chaudhuri and R. T. Haftka. Efficient global optimization with adaptive target setting. *AIAA Journal*, 52(7):1573–1577, 2014. ISSN 00011452. doi: 10.2514/1.J052930.

R. Chen, J. Xu, S. Zhang, C.-H. Chen, and L. H. Lee. An effective learning procedure for multi-fidelity simulation optimization with ordinal transformation. In *2015 IEEE International Conference on Automation Science and Engineering (CASE)*, pages 702–707. IEEE, 2015. ISBN 1467381837.

V. C. Chen, K. L. Tsui, R. R. Barton, and M. Meckesheimer. A review on design, modeling and applications of computer experiments. *IIE Transactions (Institute of Industrial Engineers)*, 38 (4):273–291, 2006. ISSN 0740817X. doi: 10.1080/07408170500232495.

B. Cheng and D. M. Titterington. Neural Networks : A Review from a Statistical Perspective. *Statistical Sinica*, 9(1):2–54, 1994.

Q. S. Cheng, S. Koziel, and J. W. Bandler. Simplified space-mapping approach to enhancement of microwave device models. *International Journal of RF and Microwave Computer-Aided Engineering*, 16(5):518–535, 2006. ISSN 10964290. doi: 10.1002/mmce.20172.

V. Cherkassky and F. M. Mulier. *Learning from data: concepts, theory, and methods*. John Wiley & Sons, 2007.

S. Choi, J. J. Alonso, and I. M. Kroo. Multi-fidelity design optimization studies for supersonic jets using surrogate management frame method. *Collection of Technical Papers - AIAA Applied Aerodynamics Conference*, 2:960–977, 2005. ISSN 10485953. doi: 10.2514/6.2005-5077.

S. Choi, J. J. Alonso, and H. M. Kroo. Two-level multifidelity design optimization studies for supersonic jets. *Journal of Aircraft*, 46(3):776–790, 2009. ISSN 00218669. doi: 10.2514/1.34362.

D. E. Christensen and K. Willcox. *Multifidelity Methods for Multidisciplinary Design Under Uncertainty*. PhD thesis, 2012.

H. S. Chung and J. J. Alonso. Using gradients to construct cokriging approximation models for high-dimensional design optimization problems. *40th AIAA Aerospace Sciences Meeting and Exhibit*, 2002.

A. R. Conn. *Introduction to Derivative-Free Optimization*, volume 01. 2009.

A. R. Conn, N. I. M. Gould, and P. L. Toint. *Trust region methods*. SIAM, 2000.

A. R. Conn, K. Scheinberg, and I. S. N. Vicente. Downloaded 11 / 24 / 23 to 130 . 159 . 237 . 115 . Redistribution subject to SIAM license or copyright ; see https://epubs.siam.org/terms-privacy Copyright © by SIAM . Unauthorized reproduction of this article is prohibited . Downloaded 11 / 24 / 23 to 13. 20(1):387–415, 2009.

I. Couckuyt, A. Forrester, D. Gorissen, F. De Turck, and T. Dhaene. Blind Kriging: Implementation and performance analysis. *Advances in Engineering Software*, 49:1–13, 2012. ISSN 0965-9978.

P. Craven and G. Wahba. Smoothing noisy data with spline functions - Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik*, 31(4):377–403, 1978. ISSN 0029599X. doi: 10.1007/BF01404567.

N. Cressie. *Statistics for spatial data*. John Wiley & Sons, 2015.

K. Crombecq. Surrogate modeling of computer experiments with sequential experimental design. 2011.

K. Crombecq, D. Gorissen, D. Deschrijver, and T. O. M. Dhaene. A Novel Hybrid Sequential Design Strategy for Global Surrogate Modelling of Computer Experiments. 33(4):1948–1974, 2011.

A. Da Ronch, M. Panzeri, M. A. A. Bari, R. D'Ippolito, and M. Franciolini. Adaptive design of experiments for efficient and accurate estimation of aerodynamic loads. *Aircraft Engineering and Aerospace Technology*, 89(4):558–569, 2017. ISSN 00022667. doi: 10.1108/AEAT-10-2016-0173.

W. Dahmen and A. Kunoth. Multilevel preconditioning. *Numerische Mathematik*, 63(1):315–344, 1992. ISSN 0029599X. doi: 10.1007/BF01385864.

I. Daubechies. Ten Lectures on Wavelets - Ingrid Daubechies (1992).pdf, 1998.

W. C. Davidon. Variable Metric Method for Minimization. Technical report, 1959.

J. E. Dennis Jr and R. B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*. SIAM, 1996.

D. Deschrijver, K. Crombecq, H. M. Nguyen, and T. Dhaene. Adaptive sampling algorithm for macromodeling of parameterized S-parameter responses. *IEEE Transactions on Microwave Theory and Techniques*, 59(1):39–45, 2011. ISSN 00189480. doi: 10.1109/TMTT.2010.2090407.

Q. Du, V. Faber, and M. Gunzburger. Centroidal Voronoi tessellations: Applications and algorithms. *SIAM Review*, 41(4):637–676, 1999. ISSN 00361445. doi: 10.1137/S0036144599352836.

J. Eason and S. Cremaschi. Adaptive sequential sampling for surrogate model generation with artificial neural networks. *Computers and Chemical Engineering*, 68:220–232, 2014. ISSN 00981354. doi: 10.1016/j.compchemeng.2014.05.021. URL http://dx.doi.org/10.1016/j.compchemeng.2014.05.021.

D. Echeverría and P. W. Hemker. Space Mapping and Defect Correction. *Computational Methods in Applied Mathematics*, 5(2):107–136, 2005. ISSN 16099389. doi: 10.2478/cmam-2005-0006.

D. Echeverría and P. W. Hemker. Manifold mapping: A two-level optimization technique. *Computing and Visualization in Science*, 11(4-6):193–206, 2008. ISSN 14329360. doi: 10.1007/s00791-008-0096-y.

M. S. Eldred, A. A. Giunta, and S. S. Collis. Second-order corrections for surrogate-based optimization with model hierarchies. *Collection of Technical Papers - 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, 3(October):1754–1768, 2004. doi: 10.2514/6.2004-4457.

A. Farhang-Mehr and S. Azarm. Bayesian meta-modeling of engineering design simulations: A sequential approach with adaptation to irregularities in the response behaviour. *International Journal for Numerical Methods in Engineering*, 62(15):2104–2126, 2005. ISSN 00295981. doi: 10.1002/nme.1261.

P. Feldmann and R. W. Freund. Efficient linear circuit analysis by Pade approximation via the Lanczos process, 1994.

G. Fernandez, C. Park, N. H. Kim, and R. Haftka. Review of multi-fidelity models. (March), 2017.

M. G. Fernández-Godino, C. Park, N.-H. Kim, and R. T. Haftka. Review of multi-fidelity models. 2016. URL http://arxiv.org/abs/1609.07196.

C. C. Fischer, R. V. Grandhi, and P. S. Beran. Bayesian-enhanced low-fidelity correction approach to multifidelity aerospace design. *AIAA Journal*, 56(8):3295–3306, 2018. ISSN 00011452. doi: 10.2514/1.J056529.

R. Fisher. Arrangment of Field Experiments. *Journal of the Ministry of Agriculture*, (33):503–513, 1926.

C. Fleury and V. Braibant. Structural optimization: A new dual method using mixed variables. *International Journal for Numerical Methods in Engineering*, 23(3):409–428, 1986. ISSN 10970207. doi: 10.1002/nme.1620230307.

F. D. Foresee and M. T. Hagan. Gauss-Newton Approximation to Bayesian Learning. *School of Electrical and Computer Engineering*, pages 1930–1935, 1997.

A. Forrester, A. Sobester, and A. Keane. *Engineering design via surrogate modelling: a practical guide*. John Wiley & Sons, 2008. ISBN 0470770791.

A. I. Forrester. Black-box calibration for complex-system simulation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1924):3567–3579, 2010. ISSN 1364503X. doi: 10.1098/rsta.2010.0051.

A. I. Forrester and A. J. Keane. Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences*, 45(1-3):50–79, 2009. ISSN 03760421. doi: 10.1016/j.paerosci.2008.11.001.

A. I. Forrester, N. W. Bressloff, and A. J. Keane. Optimization using surrogate models and partially converged computational fluid dynamics simulations. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 462(2071):2177–2204, 2006. ISSN 14712946. doi: 10.1098/rspa.2006.1679.

A. I. Forrester, A. Sóbester, and A. J. Keane. Multi-fidelity optimization via surrogate modelling. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 463 (2088):3251–3269, 2007. ISSN 14712946. doi: 10.1098/rspa.2007.1900.

Y. Freund, H. S. Seung, E. Shamir, and N. Tishby. Selective Sampling Using the Query by Committee Algorithm. *Machine Learning*, 28(2-3):133–168, 1997. ISSN 08856125. doi: 10.1023/A:1007330508534.

S. E. Gano, J. E. Renaud, and B. Sanders. Hybrid variable fidelity optimization by using a kriging-based scaling function. *AIAA Journal*, 43(11):2422–2430, 2005. ISSN 00011452. doi: 10.2514/1.12466.

M. Ghoreyshi, K. J. Badcock, and M. A. Woodgate. Accelerating the numerical generation of aerodynamic models for flight simulation. *Journal of Aircraft*, 46(3):972–980, 2009. ISSN 15333868. doi: 10.2514/1.39626.

A. A. Giunta, R. Narducci, S. Burgee, B. Grossman, W. H. Mason, L. T. Watson, and R. T. Haftka. Variable-complexity response surface aerodynamic design of an HSCT wing. *13th Applied Aerodynamics Conference*, pages 994–1002, 1995. doi: 10.2514/6.1995-1886.

A. A. Giunta, S. F. Wojtkiewicz, and M. S. Eldred. Overview of modern design of experiments methods for computational simulations. *41st Aerospace Sciences Meeting and Exhibit*, (January), 2003. doi: 10.2514/6.2003-649.

T. Goel, R. T. Haftka, W. Shyy, and N. V. Queipo. Ensemble of surrogates. *Structural and Multidisciplinary Optimization*, 33(3):199–216, 2007. ISSN 1615147X. doi: 10.1007/s00158-006-0051-9.

J. Golinski. An adaptive optimization system applied to machine synthesis. *Mechanism and Machine Theory*, 8(4):419–436, 1973. ISSN 0094114X. doi: 10.1016/0094-114X(73)90018-9.

D. Gorissen, I. Couckuyt, P. Demeester, T. Dhaene, and K. Crombecq. A surrogate modeling and adaptive sampling toolbox for computer based design. *Journal of Machine Learning Research*, 11:2051–2055, 2010. ISSN 15324435.

M. Grabe, S. Fasoulas, and K. Hannemann. Minimal Induced Drag for Non-planar Lifting Surfaces with Moderate and Small Aspect Ratio Zusammenfassung Fulltext Preview Abstract - SpringerLink. *Notes on Numerical Fluid Mechanics*, (May):84–85, 2012. doi: 10.1007/978-3-642-14243-7.

R. B. Gramacy and H. K. Lee. Bayesian treed Gaussian process models with an application to computer modeling. *Journal of the American Statistical Association*, 103(483):1119–1130, 2008. ISSN 01621459. doi: 10.1198/016214508000000689.

R. B. Gramacy and H. K. Lee. Adaptive design and analysis of supercomputer experiments. *Technometrics*, 51(2):130–145, 2009. ISSN 00401706. doi: 10.1198/TECH.2009.0015.

L. L. Gratiet and J. Garnier. Recursive co-kriging model for design of computer experiments with multiple levels of fidelity. *International Journal for Uncertainty Quantification*, 4(5): 365–386, 2014. ISSN 21525099. doi: 10.1615/int.j.uncertaintyquantification.2014006914.

S. Gunn. *Support Vector Machines for classification and regression*. PhD thesis, 1998.

M. Guo, A. Manzoni, M. Amendt, P. Conti, and J. S. Hesthaven. Multi-fidelity regression using artificial neural networks: Efficient approximation of parameter-dependent output quantities. *Computer Methods in Applied Mechanics and Engineering*, 389:114378, 2022. ISSN 00457825. doi: 10.1016/j.cma.2021.114378. URL https://doi.org/10.1016/j.cma.2021.114378.

R. T. Haftka. Combining global and local approximations. *AIAA Journal*, 29(9), 1991.

R. T. Haftka, J. A. Nachlas, L. T. Watson, T. Rizzo, and R. Desai. Two-point constraint approximation in structural optimization, 1987. ISSN 00457825.

R. T. Haftka, E. P. Scott, and J. R. Cruz. Optimization and experiments: a survey. 1998.

P. Hajela and L. Berke. Neural networks in structural analysis and design: An overview. *Computing Systems in Engineering*, 3(1-4):525–538, 1992. ISSN 09560521. doi: 10.1016/0956-0521(92)90138-9.

Z. H. Han, S. Görtz, and R. Hain. A Variable-Fidelity Modeling Method for Aero-Loads Prediction. *New results in numerical and experimental fluid mechanics*, (7):17–25, 2010.

Z. H. Han, S. Görtz, and R. Zimmermann. Improving variable-fidelity surrogate modeling via gradient-enhanced kriging and a generalized hybrid bridge function. *Aerospace Science and Technology*, 25(1):177–189, 2013. ISSN 12709638. doi: 10.1016/j.ast.2012.01.006. URL http://dx.doi.org/10.1016/j.ast.2012.01.006.

T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

D. O. Hebb. The Organization of Behavior; A Neuropsychological Theory. *The American Journal of Psychology*, 63(4):633, 1949. ISSN 00029556. doi: 10.2307/1418888.

P. W. Hemker and D. Echeverría. A trust-region strategy for manifold-mapping optimization. *Journal of Computational Physics*, 224(1):464–475, 2007. ISSN 10902716. doi: 10.1016/j.jcp.2007.04.003.

W. Hendrickx and T. Dhaene. Sequential design and rational metamodelling. *Proceedings - Winter Simulation Conference*, 2005(June):290–298, 2005. ISSN 08917736. doi: 10.1109/WSC.2005.1574263.

C. M. Holden. *Visualization methodologies in aircraft design*. PhD thesis, 2004.

J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982. ISSN 0027-8424.

L. Huang, Z. Gao, and D. Zhang. Research on multi-fidelity aerodynamic optimization methods. *Chinese Journal of Aeronautics*, 26(2):279–286, 2013. ISSN 10009361. doi: 10.1016/j.cja.2013.02.004. URL http://dx.doi.org/10.1016/j.cja.2013.02.004.

P. Jiang, Q. Zhou, and X. Shao. *Surrogate-Model-Based Design and Optimization*. 2020. ISBN 9789811507304. doi: 10.1007/978-981-15-0731-1_7.

R. Jin, W. Chen, and T. W. Simpson. Comparative studies of metamodelling techniques under multiple modelling criteria. *Structural and Multidisciplinary Optimization*, 23(1):1–13, 2001. ISSN 1615147X. doi: 10.1007/s00158-001-0160-4.

D. R. Jones. A Taxonomy of Global Optimization Methods Based on Response Surfaces. *Journal of Global Optimization*, 21(4):345–383, 2001. ISSN 09255001. doi: 10.1023/A:1012771025575.

D. R. Jones, M. Schonlau, and W. J. Welch. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization*, 13:455–492, 1998.

I. M. Jonsson, L. Leifsson, S. Koziel, Y. A. Tesfahunegn, and A. Bekasiewicz. Shape optimization of trawl-doors using variable-fidelity models and space mapping. *Procedia Computer Science*, 51(1):905–913, 2015. ISSN 18770509. doi: 10.1016/j.procs.2015.05.223.

V. R. Joseph, Y. Hung, and A. Sudjianto. Blind kriging: A new method for developing metamodels. *Journal of mechanical design*, 130(3), 2008. ISSN 1050-0472.

M. Kaufman, V. Balabanov, S. L. Burgee, A. A. Giunta, B. Grossman, R. T. Haftka, W. H. Mason, and L. T. Watson. Variable-complexity response surface approximations for wing structural weight in HSCT design. *Computational Mechanics*, 18(2):112–126, 1996. ISSN 01787675. doi: 10.1007/BF00350530.

A. Keane and P. Nair. *Computational approaches for aerospace design: the pursuit of excellence*. John Wiley & Sons, 2005. ISBN 0470855479.

A. J. Keane. Wing optimization using design of experiment, response surface, and data fusion methods. *Journal of Aircraft*, 40(4):741–750, 2003. ISSN 0021-8669.

A. J. Keane. Statistical improvement criteria for use in multiobjective design optimization. *AIAA Journal*, 44(4):879–891, 2006. ISSN 00011452. doi: 10.2514/1.16875.

M. C. Kennedy and A. O'Hagan. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87(1):1–13, 2000. ISSN 00063444. doi: 10.1093/biomet/87.1.1.

M. C. Kennedy and A. O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001. ISSN 1369-7412. doi: 10.1111/1467-9868.00294.

M. C. Kennedy, A. O'Hagan, and N. Higgins. Bayesian Analysis of Computer Code Outputs. *Quantitative Methods for Current Environmental Issues*, pages 227–243, 2002. doi: 10.1007/978-1-4471-0657-9_11.

B. Kim, Y. Lee, and D. H. Choi. Construction of the radial basis function based on a sequential sampling approach using cross-validation. *Journal of Mechanical Science and Technology*, 23 (12):3357–3365, 2009. ISSN 1738494X. doi: 10.1007/s12206-009-1014-z.

J. P. Kleijnen and W. C. Van Beers. Application-driven sequential designs for simulation experiments: Kriging metamodelling. *Journal of the Operational Research Society*, 55(8): 876–883, 2004. ISSN 01605682. doi: 10.1057/palgrave.jors.2601747.

D. L. Knill, A. A. Giunta, C. A. Baker, B. Grossman, W. H. Mason, R. T. Haftka, and L. T. Watson. HSCT configuration design using response surface approximations of supersonic euler aerodynamics. *36th AIAA Aerospace Sciences Meeting and Exhibit*, (January 1998), 1998. doi: 10.2514/6.1998-905.

P. N. Koch, T. W. Simpson, J. K. Allen, and F. Mistree. Statistical approximations for multidisciplinary design optimization: The problem of size. *Journal of Aircraft*, 36(1):275–286, 1999. ISSN 15333868. doi: 10.2514/2.2435.

P. N. Koch, R. J. Yang, and L. Gu. Design for six sigma through robust optimization. *Structural and Multidisciplinary Optimization*, 26(3-4):235–248, 2004. ISSN 1615147X. doi: 10.1007/s00158-003-0337-0.

T. Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59–69, 1982. ISSN 03401200. doi: 10.1007/BF00337288.

T. G. Kolda, R. M. Lewis, and V. Torczon. Optimization by direct search: New perspectives on some classical and modern methods. *SIAM Review*, 45(3):385–482, 2003. ISSN 00361445. doi: 10.1137/S003614450242889.

A. N. Kolmogorov. On the representation of continuous functions of several variables as superpositions of continuous functions of one variable and addition. *Doklady Akademii. Nauk USSR*, 114:679–681, 1957. doi: 10.18411/lj-12-2018-148.

R. Kosonen and J. Shemeikka. The use of a simple simulation tool for energy analysis. *Proceeding of Building Simulation, 5h Conference of International Building Performance Simulation Association, Prague, Czech Republic, 8-10 September*, pages 103–110, 1997.

S. Koziel. Efficient optimization of microwave circuits using shape-preserving response prediction. *IEEE MTT-S International Microwave Symposium Digest*, pages 1569–1572, 2009. ISSN 0149645X. doi: 10.1109/MWSYM.2009.5166010.

S. Koziel and L. Leifsson. Shape-preserving response prediction for engineering design optimization. *IEEE Transactions on Microwave Theory and Techniques*, 58(11), 2010. ISSN 18770509. doi: 10.1016/j.procs.2013.05.253.

S. Koziel and L. Leifsson. Knowledge-based airfoil shape optimization using space mapping. In *30th AIAA applied aerodynamics conference*, page 3016, 2012.

S. Koziel and L. Leifsson. Multi-level CFD-based airfoil shape optimization with automated low-fidelity model selection. *Procedia Computer Science*, 18:889–898, 2013. ISSN 18770509. doi: 10.1016/j.procs.2013.05.254. URL http://dx.doi.org/10.1016/j.procs.2013.05.254.

S. Koziel and L. Leifsson. Shape-Preserving Response Prediction for Surrogate Modeling and Engineering Design Optimization. *Springer Proceedings in Mathematics and Statistics*, 97: 241–253, 2014.

S. Koziel and L. Leifsson. *Simulation-driven design by knowledge-based response correction techniques*. 2016. ISBN 9783319301150. doi: 10.1007/978-3-319-30115-0.

S. Koziel and X.-S. Yang. *Computational Optimization, Methods and Algorithms*. 2011. ISBN 9783642175565.

S. Koziel, J. W. Bandler, and K. Madsen. A space-mapping framework for engineering optimization - Theory and implementation. *IEEE Transactions on Microwave Theory and Techniques*, 54(10):3721–3730, 2006. ISSN 00189480. doi: 10.1109/TMTT.2006.882894.

S. Koziel, Q. Cheng, and J. W. Bandler. Space mapping. *IEEE Microwave Magazine*, (December): 105–122, 2008.

S. Koziel, J. W. Bandler, and K. Madsen. Space mapping with adaptive response correction for microwave design optimization. *Society*, 57(2):478–486, 2009.

S. Koziel, Q. S. Cheng, and J. W. Bandler. Implicit space mapping with adaptive selection of preassigned parameters. *IET Microwaves, Antennas and Propagation*, 4(3):361–373, 2010. ISSN 17518725. doi: 10.1049/iet-map.2009.0080.

S. Koziel, S. Ogurtsov, and S. Szczepanski. Rapid antenna design optimization using shape-preserving response prediction. *Bulletin of the Polish Academy of Sciences: Technical Sciences*, 60(1):143–149, 2012. ISSN 02397528. doi: 10.2478/v10175-012-0020-y.

A. Krogh and J. Vedelsby. Neural Network Ensembles, Cross Validation, and Active Learning Anders. pages 6–7, 1995.

D. S. Lazzara, R. Haimes, and K. Willcox. Multifidelity geometry and analysis in aircraft conceptual design. *19th AIAA Computational Fluid Dynamics Conference*, (June):1–22, 2009.

L. Le Gratiet. *Multi-fidelity Gaussian process regression for computer experiments*. PhD thesis, 2013.

L. Le Gratiet and C. Cannamela. Cokriging-Based Sequential Design Strategies Using Fast Cross-Validation Techniques for Multi-Fidelity Computer Codes. *Technometrics*, 57(3): 418–427, 2015. ISSN 15372723. doi: 10.1080/00401706.2014.928233.

S. Leary, A. Bhaskar, and A. Keane. A constraint mapping approach to structural optimization of an expensive model uisng surrogates, 2001.

S. J. Leary, A. Bhaskar, and A. J. Keane. A knowledge-based approach to response surface modelling in multifidelity optimization. *Journal of Global Optimization*, 26(3):297–319, 2003. ISSN 09255001. doi: 10.1023/A:1023283917997.

L. Leifsson, S. Koziel, and Y. Tesfahunegn. Fast Multi-Objective Aerodynamic Optimization Using Space-Mapping-Corrected Multi-Fidelity Models and Kriging Interpolation. 153, 2016. doi: 10.1007/978-3-319-27517-8. URL http://link.springer.com/10.1007/978-3-319-27517-8.

K. Levenberg. A Method for the Solution of Certain Non-Linear Problems in Least Squares. *Quarterly of Applied Mathematics*, 2(2):164–168, 1944. ISSN 0022-0574.

J.-C. Lévesque, A. Durand, C. Gagné, and R. Sabourin. Bayesian optimization for conditional hyperparameter spaces. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 286–293. IEEE, 2017.

R. Lewis and S. Nash. A multigrid approach to the optimization of systems governed by differential equations. In *8th symposium on multidisciplinary analysis and optimization*, page 4890, 2000.

B. Li, L. Peng, and B. Ramadass. Accurate and efficient processor performance prediction via regression tree based modeling. *Journal of Systems Architecture*, 55(10-12):457–467, 2009. ISSN 13837621. doi: 10.1016/j.sysarc.2009.09.004. URL http://dx.doi.org/10.1016/j.sysarc.2009.09.004.

G. Li, S. Azarm, A. Farhang-Mehr, and A. R. Diaz. Approximation of multiresponse deterministic engineering simulations: A dependent metamodeling approach. *Structural and Multidisciplinary Optimization*, 31(4):260–269, 2006. ISSN 1615147X. doi: 10.1007/s00158-005-0574-5.

G. Li, V. Aute, and S. Azarm. An accumulative error based adaptive design of experiments for offline metamodeling. *Structural and Multidisciplinary Optimization*, 40(1-6):137–155, 2010. ISSN 1615147X. doi: 10.1007/s00158-009-0395-z.

Y. Lin, F. Mistree, J. K. Allen, K. L. Tsui, and V. C. Chen. A sequential exploratory experimental design method: Development of appropriate empirical models in design. *Proceedings of the ASME Design Engineering Technical Conference*, 1:1021–1035, 2004. doi: 10.1115/detc2004-57527.

D. J. Linse and R. F. Stengel. Identification of aerodynamic coefficients using computational neural networks. *Journal of Guidance, Control, and Dynamics*, 16(6):1018–1025, 1993. ISSN 07315090. doi: 10.2514/3.21122.

H. Liu, S. Xu, and X. Wang. Sequential sampling designs based on space reduction. *Engineering Optimization*, 47(7):867–884, 2015. ISSN 10290273. doi: 10.1080/0305215X.2014.928816. URL https://doi.org/10.1080/0305215X.2014.928816.

H. Liu, S. Xu, Y. Ma, X. Chen, and X. Wang. An Adaptive Bayesian Sequential Sampling Approach for Global Metamodeling. *Journal of Mechanical Design*, 138(1), 2016. ISSN 10500472. doi: 10.1115/1.4031905.

H. Liu, Y. S. Ong, and J. Cai. A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design, 2018a. ISSN 16151488.

H. Liu, S. Xu, X. Wang, S. Yang, and J. Meng. A multi-response adaptive sampling approach for global metamodeling. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 232(1):3–16, 2018b. ISSN 20412983. doi: 10.1177/0954406216672250.

J. Liu, Z. Han, and W. Song. Comparison of Infill Sampling Criteria in Kriging-Based Aerodynamic Optimization. *28th ICAS*, pages 1–10, 2014.

W. Liu and S. M. Batill. Gradient-enhanced neural network response surface approximations. *8th Symposium on Multidisciplinary Analysis and Optimization*, (February), 2000. doi: 10.2514/6.2000-4923.

W. Liu and S. M. Batill. Gradient-enhanced response surface approximations using kriging models. *9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, pages 1–11, 2002. doi: 10.2514/6.2002-5456.

A. Lovison and E. Rigoni. Adaptive sampling with a Lipschitz criterion for accurate metamodeling. *Communications in Applied and Industrial Mathematics*, 1(2):110–126, 2010. doi: 10.1685/2010CAIM545.

D. J. C. MacKay. A Practical Bayesian Framework for Backpropagation Networks. *Neural Computation*, 472(1):448–472, 1992. URL http://authors.library.caltech.edu/13793/1/MACnc92b.pdf.

T. J. MacKman, C. B. Allen, M. Ghoreyshi, and K. J. Badcock. Comparison of adaptive sampling methods for generation of surrogate aerodynamic model. *AIAA Journal*, 51(4):797–808, 2013. ISSN 00011452. doi: 10.2514/1.J051607.

K. Madsen, J. Bandler, S. Dakroury, M. Bakr, J. Sondergaard, Q. Cheng, and A. Mohamed. Space Mapping: The State of the Art. *IEEE Transactions on Microwave Theory and Techniques*, 52 (1):337–361, 2004. ISSN 0018-9480. doi: 10.1109/tmtt.2003.820904.

A. March and K. Willcox. Provably convergent multifidelity optimization algorithm not requiring high-fidelity derivatives. *AIAA Journal*, 50(5):1079–1089, 2012a. ISSN 00011452. doi: 10.2514/1.J051125.

A. March and K. Willcox. Constrained multifidelity optimization using model calibration. *Structural and Multidisciplinary Optimization*, 46(1):93–109, 2012b. ISSN 1615147X. doi: 10.1007/s00158-011-0749-1.

A. March, K. Willcox, and Q. Wang. Gradient-based multifidelity optimisation for aircraft design using Bayesian model calibration. *Aeronautical Journal*, 115(1174):729–738, 2011. ISSN 00019240. doi: 10.1017/S0001924000006473.

A. I. March. *Multifidelity Methods for Multidisciplinary System Design*. PhD thesis, 2012. URL https://dspace.mit.edu/bitstream/handle/1721.1/76145/820204781-MIT.pdf?sequence=2.

D. Marquardt. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.

A. L. Marsden, M. Wang, J. E. Dennis, Jr., and P. Moin. Optimal Aeroacoustic Shape Design Using the Surrogate Management Framework. *Optimization and Engineering*, 5(2):235–262, 2004. doi: 10.1023/b:opte.0000033376.89159.65.

J. Martínez-Frutos and D. Herrero-Pérez. Kriging-based infill sampling criterion for constraint handling in multi-objective optimization. *Journal of Global Optimization*, 64(1):97–115, 2016. ISSN 15732916. doi: 10.1007/s10898-015-0370-8.

B. H. Mason, R. T. Haftka, E. R. Johnson, and G. L. Farley. Variable complexity design of composite fuselage frames by response surface techniques. *Thin-Walled Structures*, 32(4): 235–261, 1998. ISSN 02638231. doi: 10.1016/S0263-8231(98)00016-0.

W. S. McCulloch and W. H. Pitts. Logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 1943.

J. McFarland, S. Mahadevan, V. Romero, and L. Swileir. Calibration and uncertainty analysis for computer simulations with multivariate output. *AIAA Journal*, 46(5):1253–1265, 2008. ISSN 00011452. doi: 10.2514/1.35288.

X. Meng and G. E. Karniadakis. A composite neural network that learns from multi-fidelity data: Application to function approximation and inverse pde problems. *Journal of Computational Physics*, 401:109020, 2020.

T. Mengistu and W. Ghaly. Aerodynamic optimization of turbomachinery blades using evolutionary methods and ANN-based surrogate models. *Optimization and Engineering*, 9(3): 239–255, 2008. ISSN 13894420. doi: 10.1007/s11081-007-9031-1.

E. Minisci and M. Vasile. Robust Design of a Re-entry Unmanned Space Vehicle by Multi-Fidelity Evolution Control. 51:4–6, 2013.

D. C. Montgomery. *Desgin And Analysis Of Experiments*. 2001. ISBN 0-471-31469-0.

G. E. Moore and Others. Cramming more components onto integrated circuits, 1965.

M. D. Morris, T. J. Mitchell, and D. Ylvisaker. Bayesian design and analysis of computer experiments: Use of derivatives in surface prediction. *Technometrics*, 35(3):243–255, 1993. ISSN 15372723. doi: 10.1080/00401706.1993.10485320.

A. A. Mullur and A. Messac. Metamodeling using extended radial basis functions: A comparative approach. *Engineering with Computers*, 21(3):203–217, 2006. ISSN 01770667. doi: 10.1007/s00366-005-0005-7.

D. E. Myers. Co-kriging—new developments. In *Geostatistics for Natural Resources Characterization Part 1*, pages 295–305. Springer, 1984.

R. H. Myers and D. C. Montgomery. Response Surface Methodology: Process and Product in Optimization Using Designed Experiments. 1995. ISSN 0471581003.

A. Narayan, C. Gittelson, and D. Xiu. A stochastic collocation algorithm with multifidelity models. *SIAM Journal of Scientific Computing*, 36(2):495–521, 2014.

L. W. Ng and K. E. Willcox. Multifidelity approaches for optimization under uncertainty. *International Journal for Numerical Methods in Engineering*, (February):1102–1119, 2014. doi: 10.1002/nme. URL http://onlinelibrary.wiley.com/doi/10.1002/nme.3279/full.

R. M. Paiva, A. R. Carvalho, C. Crawford, and A. Suleman. Comparison of surrogate models in a multidisciplinary optimization framework for wing design. *AIAA Journal*, 48(5):995–1006, 2010. ISSN 00011452. doi: 10.2514/1.45790.

K. Palmer and K. Tsui. A minimum bias Latin hypercube design. *IIE Transactions (Institute of Industrial Engineers)*, 33(9):793–808, 2001. ISSN 0740817X. doi: 10.1080/07408170108936873.

S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 2010.

C. Park, R. T. Haftka, and N. H. Kim. Remarks on multi-fidelity surrogates. *Structural and Multidisciplinary Optimization*, 55(3):1029–1050, 2017. ISSN 16151488. doi: 10.1007/s00158-016-1550-y. URL http://dx.doi.org/10.1007/s00158-016-1550-y.

J. M. Parr, A. I. Forrester, A. J. Keane, and C. M. Holden. Enhancing infill sampling criteria for surrogate-based constrained optimization. *Journal of Computational Methods in Sciences and Engineering*, 12(1-2):25–45, 2012a. ISSN 14727978. doi: 10.3233/JCM-2012-0402.

J. M. Parr, A. J. Keane, A. I. Forrester, and C. M. Holden. Infill sampling criteria for surrogate-based optimization with constraint handling. *Engineering Optimization*, 44(10):1147–1166, 2012b. ISSN 0305215X. doi: 10.1080/0305215X.2011.637556.

P. Pedersen. *The integrated approach of FEM-SLP for solving problems of optimal design*. Technical University of Denmark, 1980.

B. Peherstorfer, T. Cui, Y. Marzouk, and K. Willcox. Multi-fidelity Monte Carlo: importance sampling method using high-fidelity and surrogate model. 7831:1–23, 2015.

B. Peherstorfer, K. Willcox, and M. Gunzburger. Optimal model management for multifidelity Monte Carlo estimation. *SIAM Journal on Scientific Computing*, 38(5):A3163–A3194, 2016. ISSN 10957200. doi: 10.1137/15M1046472.

B. Peherstorfer, K. Willcox, and M. Gunzburger. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *SIAM Review*, 60(3):550–591, 2018. ISSN 00361445. doi: 10.1137/16M1082469.

M. Penwarden, S. Zhe, A. Narayan, and R. M. Kirby. Multifidelity modeling for physics-informed neural networks (pinns). *Journal of Computational Physics*, 451:110844, 2022.

M. J. Powell. A fortran subroutine for of non-linear. 1968.

M. J. Powell. A new algorithm for unconstrained optimization. In *Nonlinear programming*, pages 31–65. Elsevier, 1970.

M. J. D. Powell. Convergence properties of a class of minimization algorithms. In *Nonlinear programming 2*, pages 1–27. Elsevier, 1975.

P. Z. Qian and C. F. Wu. Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments. *Technometrics*, 50(2):192–204, 2008. ISSN 00401706. doi: 10.1198/004017008000000082.

P. Z. Qian, B. Tang, and C. F. Jeff Wu. Nested space-filling designs for computer experiments with two levels of accuracy. *Statistica Sinica*, 19(1):287–300, 2009. ISSN 10170405.

P. Z. Qian, Y. Hwang, M. Ai, and H. Su. Asymmetric nested lattice samples. *Technometrics*, 56 (1):46–54, 2014. ISSN 00401706. doi: 10.1080/00401706.2013.800002.

N. V. Queipo, R. T. Haftka, W. Shyy, T. Goel, R. Vaidyanathan, and P. Kevin Tucker. Surrogate-based analysis and optimization. *Progress in Aerospace Sciences*, 41(1):1–28, 2005. ISSN 03760421. doi: 10.1016/j.paerosci.2005.02.001.

M. M. Rai and N. K. Madavan. Aerodynamic design using neural networks. *AIAA journal*, 38 (1):173–182, 2000. ISSN 00011452. doi: 10.2514/2.938.

D. Rajnarayan, A. Haas, and I. Kroo. A multifidelity gradient-free optimization method and application to aerodynamic design. *12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, MAO*, (September), 2008.

D. G. Rajnarayan. *Trading risk and performance for engineering design optimization using multifidelity analyses*. Stanford University, 2009. ISBN 1109240821.

J. Rasmussen. Accumulated approximation: A new method for structural optimization by iterative improvement. In *NASA. Langley Research Center, The Third Air Force (NASA Symposium on Recent Advances in Multidisciplinary Analysis and Optimization*, 1990.

J. E. Rayas-Sánchez. The Aggressive Space Mapping (ASM) Algorithm : Over Two Decades of Development and Engineering Applications. pages 1–18, 2015.

S. Razavi, B. A. Tolson, and D. H. Burn. Numerical assessment of metamodelling strategies in computationally intensive optimization. *Environmental Modelling and Software*, 34:67–86, 2012. ISSN 13648152. doi: 10.1016/j.envsoft.2011.09.010. URL http://dx.doi.org/10.1016/j.envsoft.2011.09.010.

M. Redhe and L. Nilsson. Optimization of the new Saab 9-3 exposed to impact load using a space mapping technique. *Structural and Multidisciplinary Optimization*, 27(5):411–420, 2004. ISSN 1615147X. doi: 10.1007/s00158-004-0396-x.

M. Redhe and L. Nilsson. A multipoint version of space mapping optimization applied to vehicle crashworthiness design. *Structural and Multidisciplinary Optimization*, 31(2):134–146, 2006. ISSN 1615147X. doi: 10.1007/s00158-005-0544-y.

F. Regazzoni, S. Pagani, A. Cosenza, A. Lombardi, and A. Quarteroni. A physics-informed multi-fidelity approach for the estimation of differential equations parameters in low-data or large-noise regimes. *Rendiconti Lincei*, 32(3):437–470, 2021.

R. Reichart, K. Tomanek, U. Hahn, and A. Rappoport. Multi-task active learning for linguistic annotations. *ACL-08: HLT - 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference*, (June):861–869, 2008.

P.-E. Rethore, P. Fugslang, G. C. Larsen, T. Buhl, T. J. Larsen, and H. A. Madsen. Multi-fidelity optimization of wind farms. *Wind Energy*, (November 2013), 2013. doi: 10.1002/we.

T. D. Robinson, M. S. Eldred, K. E. Willcox, and R. Haimes. Surrogate-based optimization using multifidelity models with variable parameterization and corrected space mapping. *AIAA Journal*, 46(11):2814–2822, 2008a. ISSN 00011452. doi: 10.2514/1.36043.

T. D. Robinson, M. S. Eldred, K. E. Willcox, and R. Haimes. Surrogate-based optimization using multifidelity models with variable parameterization and corrected space mapping. *AIAA Journal*, 46(11):2814–2822, 2008b. ISSN 00011452. doi: 10.2514/1.36043.

J. F. Rodriguez, J. E. Renaud, and L. T. Watson. Trust region augmented Lagrangian methods for sequential response surface approximation and optimization. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 80449, page V002T29A004. American Society of Mechanical Engineers, 1997. ISBN 0791880443.

J. F. Rodríguez, J. E. Renaud, and L. T. Watson. Convergence of trust region augmented Lagrangian methods using variable fidelity approximation data. *Structural Optimization*, 15 (3-4):141–156, 1998. ISSN 09344373. doi: 10.1007/bf01203525.

J. F. Rodríguez, V. M. Pérez, D. Padmanabhan, and J. E. Renaud. Sequential approximate optimization using variable fidelity response surface approximations. *Structural and Multidisciplinary Optimization*, 22(1):24–34, 2001. ISSN 1615147X. doi: 10.1007/s001580100122.

B. Rosenbaum. Efficient Global Surrogate Models for Responses of Expensive Simulations. *PhD Thesis*, page 171, 2013.

F. Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958. ISSN 1939-1471.

J. Roshanian, A. A. Bataleblu, and M. Ebrahimi. A novel metamodel management strategy for robust trajectory design of an expendable launch vehicle. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, 0(0):1–18, 2019. ISSN 20413025. doi: 10.1177/0954410019860638.

W. J. Roux, N. Stander, and R. T. Haftka. Response surface approximations for structural optimization. *6th Symposium on Multidisciplinary Analysis and Optimization*, 534(August 1997):565–578, 1996. doi: 10.2514/6.1996-4042.

G. Rozza, D. B. P. Huynh, and A. T. Patera. Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations. *Archives of Computational Methods in Engineering*, 15(3):1–47, 2007. ISSN 1134-3060. doi: 10.1007/bf03024948.

D. E. Rumelhart, B. Widrow, and M. A. Lehr. The Basic Ideas in Neural Networks. *Communications of the ACM*, 37(3):87–92, 1994. ISSN 15577317. doi: 10.1145/175247.175256.

J. Sacks, W. J. Welch, T. J. M. And, and H. P. Wynn. Design and analysis of computer experiments. *Statistical SCience*, 4(4):409–423, 1989. ISSN 0002-6018. doi: 10.1002/9781118842003.ch9.

T. J. Santner, B. J. Williams, W. I. Notz, and B. J. Williams. *The design and analysis of computer experiments*, volume 1. Springer, 2003.

M. Santos, B. Mattos, and R. Girardi. Aerodynamic coefficient prediction of airfoils using neural networks. In *46th AIAA aerospace sciences meeting and exhibit*, page 887, 2008.

M. J. Sasena. Flexibility and Efficiency Enhancements for Constrained Global Design Optimization with Kriging Approximations. *University of Michigan*, pages 1–237, 2002. doi: 10.1.1.2.4697. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.2.4697.

R. R. Schaller. Moore's law: past, present, and future. *IEEE Spectrum*, 34(6):52–55, 57, 1997. ISSN 00189235. doi: 10.1109/6.591665.

P. J. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, 2010. ISSN 14697645. doi: 10.1017/S0022112010001217.

L. A. Schmit and B. Farshl. Some approximation concepts for structural synthesis. *AIAA Journal*, 12(5):692–699, 1974. ISSN 00011452. doi: 10.2514/3.49321.

M. Schonlau. Computer Experiments and Global Optimization. 1997.

B. Settles. Computer Sciences Active Learning Literature Survey. (January), 2009.

H. S. Seung, M. Opper, and H. Sompolinsky. Query by committee. In *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory*, pages 287–294, 1992. ISBN 089791497X. doi: 10.1145/130385.130417.

M. C. Shewry and H. P. Wynn. Maximum entropy sampling. *Journal of applied statistics*, 14(2): 165–170, 1987. ISSN 0266-4763.

M. Shi, L. Lv, W. Sun, and X. Song. A multi-fidelity surrogate model based on support vector regression. *Structural and Multidisciplinary Optimization*, 2020. ISSN 16151488. doi: 10.1007/s00158-020-02522-6.

T. W. Simpson. Design and Analysis of Computer Experiments in Multidisciplinary Design Optimization: A Review of How Far We Have Come - or Not. *AIAA*, (September):1–21, 2008. doi: doi:10.2514/6.2008-5802.

T. W. Simpson, J. D. Peplinski, P. N. Koch, and J. K. Allen. Metamodels for computer-based engineering design: Survey and recommendations. *Engineering with Computers*, 17(2): 129–150, 2001. ISSN 01770667. doi: 10.1007/PL00007198.

G. Singh and R. V. Grandhi. Mixed-Variable Optimization Strategy Employing Multifidelity Simulation and Surrogate Models. *AIAA Journal*, 48(1):215–223, 2009. ISSN 0001-1452. doi: 10.2514/1.43469.

P. Singh, D. Deschrijver, and T. Dhaene. A BALANCED SEQUENTIAL DESIGN STRATEGY FOR GLOBAL SURROGATE MODELING. In *Winter Simulation Conference*, 2013.

A. Sóbester, S. J. Leary, and A. J. Keane. A parallel updating scheme for approximating and optimizing high fidelity computer simulations. *Structural and Multidisciplinary Optimization*, 27(5):371–383, 2004. ISSN 1615147X. doi: 10.1007/s00158-004-0397-9.

J. Sobieszczanski-Sobieski and R. T. Haftka. Multidisciplinary aerospace design optimization: Survey of recent developments. *Structural Optimization*, 14(1):1–23, 1997. ISSN 09344373. doi: 10.1007/BF01197554.

A. I. M. Sobol. On the Systematic Search in a Hypercube Published by : Society for Industrial and Applied Mathematics Stable URL : https://www.jstor.org/stable/2156633. 16(5):790–793, 1979.

I. M. Sobol. Distribution of points in a cube and approximate evaluation of integrals. *USSR Computational mathematics and mathematical physics*, 7:86–112, 1967.

E. Spyromitros-Xioufis, G. Tsoumakas, W. Groves, and I. Vlahavas. Multi-target regression via input space expansion: treating targets as inputs. *Machine Learning*, 104(1):55–98, 2016. ISSN 15730565. doi: 10.1007/s10994-016-5546-z.

R. B. Statnikov and J. B. Matusov. *Multicriteria optimization and engineering*. 2012.

J. Steck, K. Rokhsaz, J. Steck, and K. Rokhsaz. Some applications of artificial neural networks in modeling of nonlinear aerodynamics and flight dynamics. In *35th Aerospace Sciences Meeting and Exhibit*, page 338, 1997.

G. Sun, G. Li, M. Stone, and Q. Li. A two-stage multi-fidelity optimization procedure for honeycomb-type cellular materials. *Computational Materials Science*, 49(3):500–511, 2010. ISSN 0927-0256.

J. A. K. Suykens, J. P. L. Vandewalle, and B. L. de Moor. *Artificial neural networks for modelling and control of non-linear systems*. 1996. ISBN 1475724934.

K. Svanberg. The method of moving asymptotes—a new method for structural optimization. *International Journal for Numerical Methods in Engineering*, 24(2):359–373, 1987. ISSN 10970207. doi: 10.1002/nme.1620240207.

M. A. Taddy, H. K. Lee, G. A. Gray, and J. D. Griffin. Bayesian guided pattern search for robust local optimization. *Technometrics*, 51(4):389–401, 2009. ISSN 00401706. doi: 10.1198/TECH.2009.08007. URL https://doi.org/10.1198/TECH.2009.08007.

D. J. Toal and A. J. Keane. Non-stationary kriging for design optimization. *Engineering Optimization*, 44(6):741–765, 2012. ISSN 0305215X. doi: 10.1080/0305215X.2011.607816.

V. Toropov, F. Va, and V. Markine. Refinements in the multi-point approximation method to reduce the effects of noisy structural responses. In *6th Symposium on Multidisciplinary Analysis and Optimization*, page 4087, 1996.

V. V. Toropov. Simulation approach to structural optimization. *Structural Optimization*, 1(1): 37–46, 1989. ISSN 09344373. doi: 10.1007/BF01743808.

J. D. Toscano, V. Oommen, A. J. Varghese, Z. Zou, N. Ahmadi Daryakenari, C. Wu, and G. E. Karniadakis. From pinns to pikans: Recent advances in physics-informed machine learning. *Machine Learning for Computational Science and Engineering*, 1(1):1–43, 2025.

C. J. Turner, R. H. Crawford, and M. I. Campbell. Multidimensional sequential sampling for NURBs-based metamodel development. *Engineering with Computers*, 23(3):155–174, 2007. ISSN 01770667. doi: 10.1007/s00366-006-0051-9.

M. Tyan and J. W. Lee. Efficient multi-response adaptive sampling algorithm for construction of variable-fidelity aerodynamic tables. *Chinese Journal of Aeronautics*, 32(3):547–558, 2019. ISSN 10009361. doi: 10.1016/j.cja.2018.12.012. URL https://doi.org/10.1016/j.cja.2018.12.012.

W. C. van Beers and J. P. Kleijnen. Customized sequential designs for random simulation experiments: Kriging metamodeling and bootstrapping. *European Journal of Operational Research*, 186(3):1099–1113, 2008. ISSN 03772217. doi: 10.1016/j.ejor.2007.02.035.

E. R. Van Dam, B. Husslage, D. Den Hertog, and H. Melissen. Maximin Latin hypercube designs in two dimensions. *Operations Research*, 55(1):158–169, 2007. ISSN 0030364X. doi: 10.1287/opre.1060.0317.

V. Vapnik. *The nature of statistical learning theory*. 1999.

S. Venkataraman, R. Haftka, and T. Johnson. Design of shell structures for buckling using correction response surface approximations. In *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, page 4855, 1998.

G. Venter, R. T. Haftka, and J. H. Starnes. Construction of response surface approximations for design optimization. *AIAA Journal*, 36(12):2242–2249, 1998. ISSN 00011452. doi: 10.2514/2.333.

F. A. Viana and R. T. Haftka. Using multiple surrogates for minimization of the RMS error in meta-modeling. In *Proceedings of the ASME Design Engineering Technical Conference*, volume 1, pages 851–860, 2008. ISBN 9780791843253. doi: 10.1115/DETC2008-49240.

F. A. Viana, T. W. Simpson, V. Balabanov, and V. Toropov. Metamodeling in multidisciplinary design optimization: How far have we really come? *AIAA Journal*, 52(4):670–690, 2014. ISSN 00011452. doi: 10.2514/1.J052375.

R. Vitali, R. Haftka, and B. Sankar. Correction response surface approximations for stress intensity factors of a composite stiffened plate. In *39th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference and Exhibit*, page 2047, 1998.

R. Vitali, O. Park, R. T. Haftka, B. V. Sankar, and C. A. Rose. Structural optimization of a hat-stiffened panel using response surfaces. *Journal of aircraft*, 39(1):158–166, 2002. ISSN 0021-8669.

G. Wahba. *Spline models for observational data*. SIAM, 1990.

R. Wallach, B. S. de Mattos, and R. da Mota Girardi. Aerodynamic coefficient prediction of a general transport aircraft using neural network. *ICAS-Secretariat - 25th Congress of the International Council of the Aeronautical Sciences 2006*, 2(January):1199–1214, 2006. doi: 10.2514/6.2006-658.

G. G. Wang and S. Shan. Review of metamodeling techniques in support of engineering design optimization. *Journal of Mechanical Design, Transactions of the ASME*, 129(4):370–380, 2007. ISSN 10500472. doi: 10.1115/1.2429697.

H. Wang, F. Ye, E. Li, and G. Li. A comparative study of expected improvement-assisted global optimization with different surrogates. *Engineering optimization*, 48(8):1432–1458, 2016.

Y. Wang and C.-Y. Lai. Multi-stage neural networks: Function approximator of machine precision. *Journal of Computational Physics*, 504:112865, 2024.

E. W. Weisstein. Least squares fitting–polynomial, 2002. URL https://mathworld.wolfram.com/.

B. Widrow and M. Hoff. Adaptive Switching Circuits. Technical Report May, 1960.

S. M. Wild and C. Shoemaker. Global convergence of radial basis function trust-region algorithms for derivative-free optimization. *SIAM Review*, 55(2):349–371, 2013. ISSN 00361445. doi: 10.1137/120902434.

R. L. Winkler. Combining Probability Distributions from Dependent Information Sources. *Management Science*, 27(4):479–488, 1981. ISSN 0025-1909. doi: 10.1287/mnsc.27.4.479.

B. A. Wujek, J. E. Renaud, S. M. Batill, and J. B. Brockman. Concurrent subspace optimization using design variable sharing in a distributed computing environment. *Concurrent Engineering Research and Applications*, 4(4):361–376, 1996. ISSN 1063293X. doi: 10.1177/1063293x9600400405.

S. Xiong, P. Z. Qian, and C. F. Wu. Sequential design and analysis of high-accuracy and low-accuracy computer codes. *Technometrics*, 55(1):37–46, 2013. ISSN 00401706. doi: 10.1080/00401706.2012.723572.

Y. Xiong, W. Chen, D. Apley, and X. Ding. A non-stationary covariance-based Kriging method for metamodelling in engineering design. *International Journal for Numerical Methods in Engineering*, 2006. doi: 10.1002/nme. URL http://onlinelibrary.wiley.com/doi/10.1002/nme.3279/full.

S. Xu, H. Liu, X. Wang, and X. Jiang. A robust error-pursuing sequential sampling approach for global metamodeling based on voronoi diagram and cross validation. *Journal of Mechanical Design, Transactions of the ASME*, 136(7), 2014. ISSN 10500472. doi: 10.1115/1.4027161.

W. Yao, X. Chen, and W. Luo. A gradient-based sequential radial basis function neural network modeling method. *Neural Computing and Applications*, 18(5):477–484, 2009. ISSN 09410643. doi: 10.1007/s00521-009-0249-z.

K. Q. Ye. Orthogonal Column Latin Hypercubes and Their Application in Computer Experiments. *Journal of the American Statistical Association*, 93(444):1430–1439, 1998. ISSN 1537274X. doi: 10.1080/01621459.1998.10473803.

R. Yondo, E. Andrés, and E. Valero. A review on design of experiments and surrogate models in aircraft real-time and many-query aerodynamic analyses. *Progress in Aerospace Sciences*, 96(November 2017):23–61, 2018. ISSN 03760421. doi: 10.1016/j.paerosci.2017.11.003.

M. Y. Yoo and J. H. Choi. Probabilistic calibration of computer model and application to reliability analysis of elasto-plastic insertion problem. *Transactions of the Korean Society of Mechanical Engineers, A*, 37(9):1133–1140, 2013. ISSN 22885226. doi: 10.3795/KSME-A.2013.37.9.1133.

Y. Yue and K. Meerbergen. Accelerating optimization of parametric linear systems by model order reduction. *SIAM Journal of Optimisation*, 23(2):1344–1370, 2013.

M. K. Zahir and Z. Gao. Variable-fidelity optimization with design space reduction. *Chinese Journal of Aeronautics*, 26(4):841–849, 2013. ISSN 10009361. doi: 10.1016/j.cja.2013.06.002. URL http://dx.doi.org/10.1016/j.cja.2013.06.002.

L. E. Zerpa, N. V. Queipo, S. Pintos, and J. L. Salager. An optimization methodology of alkaline-surfactant-polymer flooding processes using field scale numerical simulation and multiple surrogates. *Journal of Petroleum Science and Engineering*, 47(3-4):197–208, 2005. ISSN 09204105. doi: 10.1016/j.petrol.2005.03.002.

L. Zhang, Q. J. Zhang, and J. Wood. Statistical neuro-space mapping technique for large-signal modeling of nonlinear devices. *IEEE Transactions on Microwave Theory and Techniques*, 56 (11):2453–2467, 2008. ISSN 00189480. doi: 10.1109/TMTT.2008.2004894.

M. Zhang, M. Tomac, C. Wang, and A. Rizzi. Variable fidelity methods and surrogate modeling of critical loads on x-31 aircraft. In *51st AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition*, page 1081, 2013.

Y. Zhang, T. N. Hoang, K. H. Low, and M. Kankanhalli. Near-Optimal active learning of multi-output Gaussian processes. *30th AAAI Conference on Artificial Intelligence, AAAI 2016*, pages 2351–2357, 2016. ISSN 2159-5399. doi: 10.1609/aaai.v30i1.10209.

Q. Zhou, X. Shao, P. Jiang, Z. Gao, H. Zhou, and L. Shu. An active learning variable-fidelity metamodelling approach based on ensemble of metamodels and objective-oriented sequential sampling. *Journal of Engineering Design*, 27(4-6):205–231, 2016. ISSN 14661837. doi: 10.1080/09544828.2015.1135236.

Q. Zhou, Y. Wang, S. K. Choi, P. Jiang, X. Shao, J. Hu, and L. Shu. A robust optimization approach based on multi-fidelity metamodel. *Structural and Multidisciplinary Optimization*, 57(2):775–797, 2018. ISSN 16151488. doi: 10.1007/s00158-017-1783-4.

Q. Zhou, J. Wu, T. Xue, and P. Jin. A two-stage adaptive multi-fidelity surrogate model-assisted multi-objective genetic algorithm for computationally expensive problems. *Engineering with Computers*, (0123456789), 2019. ISSN 14355663. doi: 10.1007/s00366-019-00844-8. URL https://doi.org/10.1007/s00366-019-00844-8.

D. Zimmerman, C. Pavlik, A. Ruggles, and M. P. Armstrong. An Experimental Comparison of Ordinary and Universal Kriging and Inverse Distance Weighting. *Mathematical Geology*, 31 (4), 1999. ISSN 00138703. doi: 10.1023/A.

Adaptive Multi-Fidelity Model Management 3

3.1 Multi-Stage Adaptive Correction . . . 128
3.1.1 Formulation . . . 134
3.1.2 Correction Methods . . . 134
3.1.3 Automated Surrogate Modelling using Artificial Neural Networks . . . 138
3.1.4 Numerical Tests . . . 143
3.2 Multi-Fidelity Model Management Framework for Outer-Loop Applications . 174
3.2.1 Initial Surrogate Generation . . . 174
3.2.2 Adaptive Sampling . . . 177
3.3 Chapter Summary . . . 178
3.4 Bibliography . . . 183

In many fields and applications across computational science and engineering, any particular system of interest is likely able to be represented by several alternative numerical models/-formulations. Such a collection of methods typically conforms to the inherent hierarchical trade-off between the cost of evaluation and the relative accuracy to which the underlying physical process can be represented.

Multi-Fidelity modelling and optimisation involves leveraging the response trends of one or more computationally inexpensive, lower quality numerical models (for example, obtained using a computer simulation based on coarsely-discretised structures of interest) to supplement a surrogate construction/refinement process in cooperation with a judiciously selected set of high-fidelity model evaluations, or indeed to entirely replace the original expensive function altogether.

Fundamentally, it is assumed that that lower-fidelity models can inexpensively supply func-

tional trends of the underlying process(es) with only occasional recourse to the high-fidelity model required to establish global/local accuracy, apply some correction/adjustment or guarantee outer-loop convergence (e.g. as part of a multi-fidelity optimisation scheme) (Ng and Willcox, 2014; Alexandrov et al., 2001; Koziel and Yang, 2011). By exploiting domain-specific knowledge in this manner, one may in essence attempt to mitigate the 'curse of dimensionality' normally associated with black-box surrogate modelling (Simpson, 2008; Keane and Nair, 2005).

The effectiveness of any multi-fidelity surrogate modelling technique largely depends on the correlation between the high- and low-fidelity responses (Bakr et al., 2002; Koziel and Leifsson, 2016; Shi et al., 2020a). Put simply, a particular strategy cannot be expected to produce a high-accuracy representation of a target function if the supporting lower-fidelity approximation is entirely arbitrary or irrelevant. In engineering and other physical-science based applications, a degree of correlation between alternative fidelity levels is largely expected given that each approximation ultimately aims to represent the same physical process. For instance, the same fundamental physics may have been used to model certain phenomena, or similar collections of data may have been used to calibrate any existing empirical aspects of the associated formulation (Allaire and Willcox, 2014).

Given that there is currently no single universally accepted/applicable multi-fidelity modelling technique, an interesting consideration is whether multiple, alternative methods, each with different strengths and weaknesses depending on the nature of the target problem (including the nature of the correlations between alternative fidelity levels), may be employed in cooperation rather than isolation within a given framework. For example, if a low-fidelity approximation can first be adaptively adjusted to maximise correlation with available high-fidelity data, the performance of a subsequent direct correction process could be numerically much simpler and/or more effective.

This section considers the exploitation of model correlation in multi-fidelity modelling. The fundamental concept is referred to as a *multi-stage* approach, in which a multi-fidelity surrogate modelling technique may be composed of one or more unique *stages* aimed at collaboratively improving the efficacy of the final adaptive correction by exploiting alternative aspects of the observed model correlation. Of note, this work primarily considers *adaptation* methods for multi-fidelity modelling (see Section 2.3.2), though it is acknowledged that certain aspects may also prove relevant to *fusion*-based approaches.

First, three of the most popular/prevalent multi-fidelity correction techniques are selected from the literature and formulated in a manner applicable to general datasets. Next, each technique is integrated into a series of multi-stage, compound processes, where the output of one stage feeds into the input to another. For completeness, this series includes each correction

technique individually (referred to as *single-stage* methods) in addition to processes comprised of two unique methods (i.e. *two-stage*) and finally all three (i.e. *three-stage*). Each process is then applied to a series of numerical benchmark functions, sourced from the literature, to assess their efficacy according to a number of generalised performance metrics.

Following the presentation and discussion of the results from the numerical tests, a novel multi-fidelity model management framework for optimisation problems is presented, incorporating the examined correction techniques alongside an adaptive sampling procedure and sensitivity-based validation criteria. The framework is designed in a modular format, where specific multi-fidelity correction and surrogate modelling methods can be independently swapped in/out. This allows a helpful level of flexibility as a research level code for testing different methods and approaches. For example, the multi-fidelity correction step can be mathematically independent from the surrogate modelling step, allowing each to be tested/examined in a more unbiased manner. Conversely, if the user wishes to implement a surrogate modelling technique that, for example, performs the multi-fidelity correction within the modelling process itself (see Section 3.1.3 for further discussion), the correction step can be switched off, or easily bypassed.

The work presented in this chapter can be found in the following peer-reviewed publications:

- B. Parsonage and C. A. Maddock. Multi-stage Multi-fidelity Information Correction for Artificial Neural Network Based Meta-modelling. *2020 IEEE Symposium Series on Computational Intelligence, SSCI 2020*, pages 950–957, 2020. doi: 10.1109/SSCI47803.2020.9308255
- B. Parsonage and C. Maddock. A multi-fidelity model management framework for multi-objective aerospace design optimisation. *Frontiers in Aerospace Engineering*, 2 (February):1–21, 2023. doi: 10.3389/fpace.2023.1046177

3.1 Multi-Stage Adaptive Correction

Multi-stage (or mixed-modelling) approaches to multi-fidelity response correction has received some attention in the literature. In most cases, one or more unique correction methods are implemented to *precondition* a low-fidelity approximation into a more suitable form amenable to adaptive correction, thus improving the attainable prediction quality of the final surrogate. Individual techniques typically exploit a specific aspect or feature of the available data, thus a multi-stage approach is in essence a *collaborative* effort.

Arguably, an earlier recognition of collaborative multi-fidelity correction is reflected in the *comprehensive* approach for surrogate adaptation. Recall from Section 2.3.2.1, the comprehens-

ive correction approach utilises both additive and multiplicative discrepancies between the low- and high-fidelity responses to generate a corrected, multi-fidelity surrogate. A typical formulation could be:

$$\hat{y}_{HF} = \rho(\mathbf{x})y_{LF}(\mathbf{x}) + \delta(\mathbf{x}) \tag{3.1}$$

where ρ and δ represent multiplicative and additive correction functions respectively. The multiplicative term is often supplied as a constant value (see for example (Keane, 2012) and (Perdikaris et al., 2015)), though non-constant methods have been applied also (Qian and Wu, 2008). The additive term is usually applied via some suitable surrogate modelling technique to create an approximating response surface of the additive discrepancy recorded at all high-fidelity sample locations. An alternative form of comprehensive correction was offered by Gano et al. (2005):

$$\hat{y}_{HF} = w(\mathbf{x})\rho(\mathbf{x})y_{LF}(\mathbf{x}) + (1 - w(\mathbf{x}))[y_{LF}(\mathbf{x}) + \delta(\mathbf{x})] \tag{3.2}$$

where $w(\mathbf{x})$ is a weighting function that may be adaptively updated. This approach is employed, for example, in the works (Zheng et al., 2013) and (Fischer et al., 2018).

However, whilst the benefits of collaborative action have been well reported within this context, correction terms are usually evaluated simultaneously and thus the overall approach remains distinct from multi-stage methods. An important early example of a more inherently collaborative approach is the comprehensive Bayesian calibration method of Kennedy and O'Hagan (Kennedy and O'Hagan, 2001a). The proposed Kriging-based framework employs a calibration-based GP surrogate modelling approach (Section 2.3.2.2), where calibration parameters are treated as non-physical hyper-parameters. As argued by Fernández-Godino et al. (2016), model calibration on it's own cannot strictly be deemed a multi-fidelity approach given it's incorporation of only one source of data. However, Kennedy and O'Hagan (2001a) employ their calibration-based approach in collaboration with a comprehensive correction incorporating multi-fidelity data. The cross-interaction between the low-fidelity and discrepancy function models is used to enhance the GP-model. Related applications of this methodology can be found in (Qian and Wu, 2008) and (Biehler et al., 2015).

The superior flexibility of this framework relative to the original Bayesian calibration approach (Kennedy and O'Hagan, 2000) is tempered by the increased complexity resulting from the large number of hyper-parameters and general model parameters. The simultaneous determination of which can prove challenging and/or impractical (Park et al., 2017). As such, the authors offer methods for sequential group-wise parameter estimation (Kennedy and O'Hagan, 2001b; Bayarri et al., 2007).

Methods employing cooperation between multiplicative scale factors and discrepancy functions have since become more common in application (Fernández-Godino et al., 2016; Park et al.,

2017; Zhou et al., 2018). The collaborative correction concept was investigated in more detail in (Park et al., 2018, 2017). The authors report that the use of low-fidelity scaling coefficients to reduce the "bumpiness[1]" (an integral of the square of the second derivative) of the additive discrepancy function can increase the prediction accuracy of the final surrogate. The given justification is that smoother, less variable functions are easier to model numerically, thus reducing the variation of discrepancy with respect to the target response mitigates the effects of fitting error. The authors apply such a calibration step as a precursor to a Kriging based correction function within a Bayesian update framework.

Building upon the comprehensive GP Bayesian framework (Kennedy and O'Hagan, 2001a), the concept of reducing the variability of cross-fidelity discrepancy prior to adaptive correction (Park et al., 2018) was adopted in the work (Zhang et al., 2022). In this case, the selection of appropriate values for a series of multiplicative scale factors (each corresponding to one of an ensemble of non-hierarchical low-fidelity responses) is formulated into an optimisation problem. The objective is to minimise the second derivative of a GP model constructed to predict the discrepancy between the high-fidelity response and the ensemble-averaged[2] low-fidelity response. In this manner, as reported by (Park et al., 2018), the discrepancy model reduces in complexity with a corresponding reduction in modelling error of the final approximation. This approach is termed the *Non-Hierarchical Low-Fidelity* (NHLF) co-Kriging approach.

In (Zhou et al., 2020), a multi-fidelity fusion-based approach is proposed for *Generalised Hierarchical co-Kriging* (GCK) surrogate modelling. This method again utilises a comprehensive GP Bayesian framework similar to that originally introduced by (Kennedy and O'Hagan, 2001a). In this case, a parameter-calibrated low-fidelity Kriging model is aggregated with a stochastic Kriging model of the function discrepancy. The latter enables the final GCK surrogate to consider the predictive uncertainty from the low-fidelity Kriging model at locations where only high-fidelity samples are available. This in turn enables the estimation of model parameters separately for both nested and non-nested sample sets.

The hybrid variable-fidelity global approximation modelling method presented by Zheng et al. (2013) combines the use of a two-parameter linear scaling function with a Kriging correction surrogate. The objective here is to determine the parameters that minimise the discrepancy between the calibrated surrogate and the high-fidelity sample set. First, a low-fidelity response is approximated using a radial basis function surrogate to reduce the computational cost of further evaluation. Next, the explicit RBF surrogate is tuned to the high-fidelity samples using a two-parameter linear scaling function, in order to capture the global trends of the high-fidelity function. Finally, the tuned surrogate is modified using a Kriging-based correction model built

[1] also referred to as "roughness", introduced by (Duchon, 1977; Gutmann, 2001).
[2] see Section 2.3.3.1.

upon the additive discrepancy between low- and high-fidelity samples.

A particularly well known approach utilising collaborative multiplicative and additive discrepancies is *Space Mapping* (SM). In (Koziel and Leifsson, 2016), the authors describe a multi-point Space Mapping method as a generalised update to *Output Space-Mapping* (OSM) and a form of parametric model calibration. In this method, the parameters of the low-fidelity model correction are determined using high-fidelity data at several data-points, rather than just one (as for standard Space Mapping (Koziel and Leifsson, 2016)). The authors report that the multi-point modification improves the generalisation capabilities of the final surrogate at the expense of more complex model formulation.

However, a problem with this approach, and indeed with any parametric method assuming a simple functional form of the calibration, is that such simple mappings may not be capable of providing sufficient accuracy (Koziel and Leifsson, 2016). Furthermore, any relatively simplistic model calibration, particularly if independent of the model input variables, may only provide a limited modification of the range of the low-fidelity model. Finally, given the finite number of calibration parameters, which are extracted simultaneously for the entire region of interest, the surrogate can be more generally described as a regression model. As such, the modelling error might not decrease below certain, problem-dependent, non-zero bounds, even if the number of samples is allowed to increase without limit (Koziel and Bandler, 2006). Of note, an attempt to mitigate these issues was investigated by Koziel et al. (2006c). In this work, a separate parameter extraction operation was performed for each evaluation of the surrogate model. Whilst the authors found that this approach increased the overall accuracy compared to the standard approach, the significant increase in evaluation time was largely prohibitive and reduced the potential applicability of the method.

The limitations of standard Space Mapping (and derivatives thereof) can be somewhat alleviated through multi-stage collaboration with alternative techniques. Indeed, the authors of (Koziel et al., 2006b, 2008; Koziel and Leifsson, 2016) acknowledge that Space Mapping, particularly OSM, has found an established relevance as a supplementary low-fidelity model correction/enhancement layer, applied sequentially amongst other types of corrections. An example of this is the enhanced (mixed-modelling) Space Mapping technique presented in (Koziel and Leifsson, 2016), of the general form:

$$\mathbf{s}(\mathbf{x}) = \mathbf{s}_{SM}(\mathbf{x}) + \tilde{\mathbf{s}}(\mathbf{x}) \tag{3.3}$$

where $\mathbf{s}(\mathbf{x})$ is a standard Space Mapping model and $\tilde{\mathbf{s}}(\mathbf{x})$ is a function approximation model. The former can be considered a trend function, whereas the latter is often modelled on the remaining discrepancy (in the form of residual modelling errors) between the trend function and the high-fidelity samples (Koziel and Bandler, 2007a). The authors list a number of advantages of this approach:

- High modelling accuracy can be achieved with limited high-fidelity samples, particularly if using underlying physics-based low-fidelity models.
- The resulting surrogate can be as, if not more, computationally inexpensive as the low-fidelity model by virtue of the ability to exploit analytical formulae within the function approximation layer.
- Modelling accuracy is not limited relative to the number of high-fidelity samples. That is, any amount of available high-fidelity model data can be utilised, thus the resulting approximation can reach any desired accuracy level provided sufficient high-fidelity samples are supplied.

This framework has been adopted by several works. For instance, where the function approximation model $\tilde{\mathbf{s}}(\mathbf{x})$ is implemented using Radial Basis Function interpolation (Koziel and Bandler, 2007a), fuzzy systems (Koziel and Bandler, 2007b) and Kriging interpolation (Koziel et al., 2012). In each case, the authors report that the staged application of both methods performs higher (i.e. to a higher accuracy) than implementing either method in isolation (assuming that the same amount of high-fidelity data is available).

This section offers a further investigation into multi-stage formulations for generic multi-fidelity surrogate model generation. A generalisation criteria is considered where the physics-based models of interest are assumed to be black-box functions, therefore the nature of the correlation between fidelity levels is not explicitly known *a priori*. Adopting the ideology of the aforementioned methods, three distinct stages representing fundamental features of commonly used model adaptation techniques from the literature are defined in a general manner.

The first employs a parametric calibration, most similar in operation to Space Mapping (Section 2.3.2.3), where the parameter extraction process presented in (Koziel and Leifsson, 2012) is used to determine two linear tuning parameters representing a comprehensive (multiplicative and additive) correction. This approach exploits the functional form of the low-fidelity output response, which is scaled and translated such to minimise the error between low- and high-fidelity responses at the observed sample points. The second utilises a non-parametric correction technique based upon the Response Correction methods of Koziel and Leifsson (2013b), where a weighted response of relative response vectors is used to infer the corresponding high-fidelity response across available samples. This method again exploits the general form of the low-fidelity response, though this time performance may be inferred by the reduction in variation of the discrepancy between the low- and high-fidelity responses. The final included method utilises an additive bridge function. In this case, an ANN surrogate is used to model and predict the additive discrepancy between the low- and high-fidelity responses. This

Correction Stage	Abbrv.	Type	Info source	Driver
Parametric tuning	P	parametric	LF response/sample-point error	sample-point error
Relative response mapping	R	non-parametric	LF response	LF/HF correlation
Additive discrepancy model	A	non-parametric/surrogate-based	sample-point error	sample-point error

Table 3.1: Summary of the selected multi-fidelity adaptive correction techniques.

approach includes an additional normalised multi-source performance metric and automated training and hyper-parameter selection procedure. This method exploits only the information relating to the error between low- and high-fidelity responses at the observed sample locations. The functional form of this discrepancy is estimated through the surrogate modelling process, thus estimating the high-fidelity response at observed locations. Table 3.1 summarises the three selected correction techniques with respect to the information exploited from the low- and high-fidelity models and the manner by which corrective performance is measured.

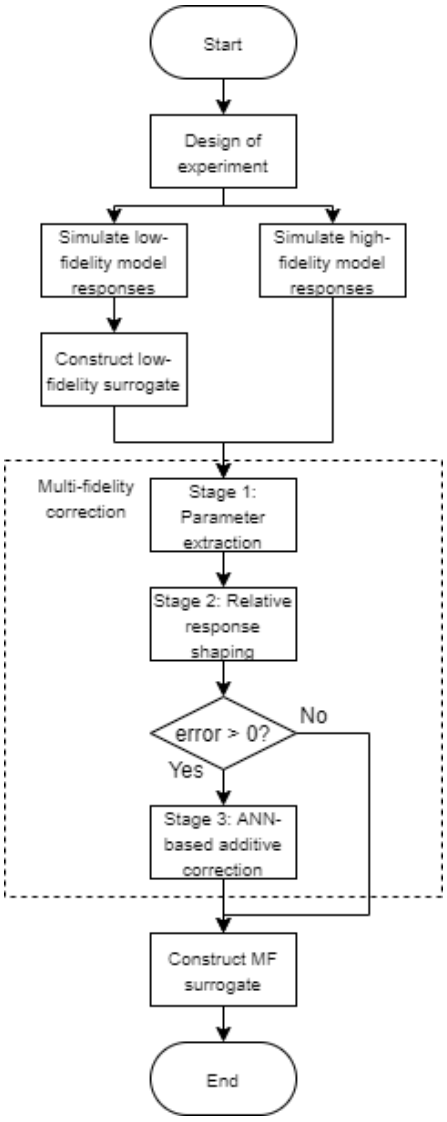


FIGURE 3.1: Flowchart for the proposed multi-fidelity correction.

3.1.1 Formulation

A generic model is defined as a function $\mathbf{f} : \mathbf{X} \rightarrow \mathbf{Y}$ that for a given input $\mathbf{x} \in \mathbf{X}$ gives an output $\mathbf{y} \in \mathbf{Y}$ where $\mathbf{X} \subseteq \mathbb{R}^{d_{in}}$ is the input domain with dimension $d_{in} \subset \mathbb{N}$ and $\mathbf{Y} \subset \mathbb{R}^{d_{out}}$ is the input domain with dimension $d_{out} \subset \mathbb{N}$. In this context, a function $\mathbf{f}(\mathbf{x})$ may represent the direct computational evaluation of a numerical model (i.e., a mathematical description and the concomitant numerical analysis) or the evaluation of an appropriate surrogate model built upon a sample set of input/output correspondence (e.g., obtained using physical experiments, analytical solutions, or expert analysis). Thus a high-fidelity model is notated simply as $\mathbf{f}_H : \mathbf{X} \rightarrow \mathbf{Y}$, a low-fidelity model as $\mathbf{f}_L : \mathbf{X} \rightarrow \mathbf{Y}$ and a surrogate model as $\mathbf{f}_S : \mathbf{X} \rightarrow \mathbf{Y}$. Let $\mathbf{x}^{(i)}, i = 1, \ldots, M$ denote the set of M design points. Directly collocated high- and low-fidelity samples is assumed, although it is possible to obtain such a correspondence via surrogate estimation (Forrester et al., 2007). The set of N high-fidelity samples is thus defined as $\mathbf{x}_H^{(i)}$, $i = 1, \ldots, N$ where $\mathbf{x}_H \in \mathbf{x}$.

For ease of notation, discussion of the selected corrective techniques is limited to the case of correcting a low-fidelity function $\mathbf{f}_L$ via selective samples of a high-fidelity function $\mathbf{f}_H$. In reality, sequential application of stages implies the low-fidelity input for one is simply the multi-fidelity output from the preceding stage.

3.1.2 Correction Methods

3.1.2.1 Parametric Tuning

A parametric approach (or *calibration* method, such as OSM, see Section 2.3.2.3) utilises an explicit formulation where the relationship between a low- and high-fidelity model is defined using a quantifiable set of tuning parameters $\mathbf{p} = [\mathbf{a}, \mathbf{b}]$, often given in the form (Koziel and Leifsson, 2016):

$$\mathbf{f}_S(\mathbf{x}, \mathbf{p}) = \mathbf{a} * \mathbf{f}_L(\mathbf{x}) + \mathbf{b} \tag{3.4}$$

where $\mathbf{a}$ and $\mathbf{b}$ are unbounded $d_{in} \times 1$ vectors. This is in contrast to non-parametric methods, with which the functional relationship between model responses is inferred from input/output correspondence rather than any explicit formula (Koziel and Leifsson, 2016). A simplified visualisation of the Parametric approach is shown in Figure 3.2. For a more complete discussion on the distinction between parametric and non-parametric methods, the reader is referred to (Koziel and Leifsson, 2016).

In this work, the vector of scaling parameters $\mathbf{p}$ is found by formulating (3.4) into the following parameter extraction problem:

$$\mathbf{p} = \arg\min_{\mathbf{q}} \frac{1}{\sigma} \sqrt{\frac{1}{N} \sum_{k=1}^{N} \left[\mathbf{f}_S(\mathbf{x}_H^{(k)}, \mathbf{q}) - \mathbf{f}_H(\mathbf{x}_H^{(k)}) \right]^2} \tag{3.5}$$

Where $\mathbf{q}$ is the optimisation vector and σ is the standard deviation of the N observed high-fidelity responses. Equation (3.5) is solved using the MATLAB® gradient based local optimiser, `fmincon`. A sequential quadratic programming algorithm is employed with a function tolerance of $1 \times 10^{-}9$ and maximum number of function evaluations of 1×10^{3}. An initial guess for $\mathbf{p}$ is given as:

$$\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}_{d_{in}} \tag{3.6}$$

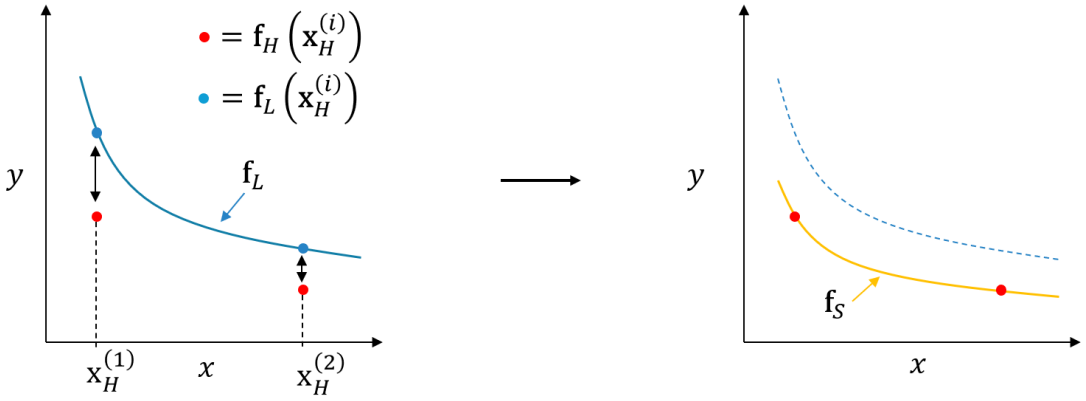


FIGURE 3.2: Visualisation of the Parametric tuning (P) correction stage.

3.1.2.2 Relative Response Mapping

The second technique selected for this study is a non-parametric correction based on the Response Correction approaches presented by Koziel and Leifsson (2014, 2016). In contrast to parametric methods, non-parametric methods do not define the relationship between low- and high-fidelity via some explicit formula, but rather directly extract correspondence from the model responses themselves (Koziel and Leifsson, 2016). Though often more complex in formulation, and more restrictive in assumptions regarding their applicability, non-parametric methods have been characterised by strong generalisation capability compared with parametric techniques. In this case, a weighted set of relative response vectors is used to infer the corresponding high-fidelity response across available samples. This approach again exploits the general form of the low-fidelity response, but evaluates performance via the reduction in the variation of the discrepancy between the low- and high-fidelity responses.

First, the change in low-fidelity model response between each pair-wise combination of design point and high-fidelity sample point defines the translation matrix $\mathbf{T}$, where:

$$\mathbf{t}_{i,j} = \mathbf{f}_L(\mathbf{x}^{(i)}) - \mathbf{f}_L(\mathbf{x}_H^{(j)}) \tag{3.7}$$

such that,

$$\mathbf{T} = \begin{bmatrix} \mathbf{t}_{1,1} & \mathbf{t}_{1,2} & \cdots & \mathbf{t}_{1,N} \\ \mathbf{t}_{2,1} & \mathbf{t}_{2,2} & \cdots & \mathbf{t}_{2,N} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{t}_{M,1} & \mathbf{t}_{M,2} & \cdots & \mathbf{t}_{M,N} \end{bmatrix}$$

where M is the number of design points $\mathbf{x}^{(i)}$ and N is the number of high-fidelity samples $\mathbf{x}_H^{(j)}$. The high-fidelity model response at each design is predicted by mapping the translation matrix $\mathbf{T}$ across the observed high-fidelity responses:

$$\mathbf{f}_S(\mathbf{x}^{(i)}) = \frac{\sum[(\mathbf{f}_H(\mathbf{x}_H^{(j)}) + \mathbf{t}_{i,j}) * \mathbf{w}_{i,j}]_{j=1,\ldots,N}}{\sum[\mathbf{w}_{i,j}]_{j=1,\ldots,N}} \tag{3.8}$$

The weighting factor $\mathbf{w}_{i,j}$ is determined according to the relative distance between points:

$$\mathbf{w}_{i,j} = 1 - \mathbf{d}_{i,j} \tag{3.9}$$

where $\mathbf{d}$ is the matrix of Euclidean norms between each design point and each high-fidelity sample, where $\hat{\mathbf{x}}$ denotes the scaled form of input vector $\mathbf{x}$ conforming to $\hat{\mathbf{x}}^{(i,k)} \in [0,1]$ for all $k = 1, \ldots, d_{in}$:

$$\mathbf{d}_{i,j} = \sqrt{\sum_{k=1}^{d_{in}} \left[\hat{\mathbf{x}}^{(i,k)} - \hat{\mathbf{x}}_H^{(j,k)}\right]^2} \tag{3.10}$$

where d_{in} is the design vector dimensionality. The weight factor $\mathbf{w}$ is constrained to the bounds $\mathbf{w}_{i,j} \in [0,1]$ to impose a maximum area of influence for any particular high-fidelity sample. A simplified visualisation of this approach is shown in Figure 3.3.

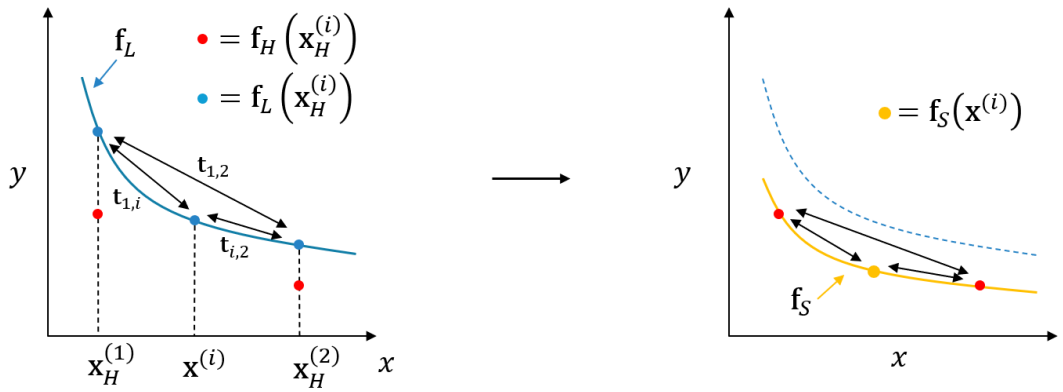


FIGURE 3.3: Visualisation of the Relative Response Mapping (R) correction stage.

An important feature of this technique is the independence from any existing correlation between low- and high-fidelity samples points. Consider the circumstance in which a pre-processing stage such as that presented in (Koziel and Leifsson, 2012) or (Zheng et al., 2013) is applied to a problem in which $M \gg N$. In the extreme case that the initial parametric

correction should correct the low-fidelity samples to such a degree that there is negligible error at all observed points, the subsequent additive discrepancy based correction would fail to produce any noticeable improvement. This problem is circumvented here by utilising correction vectors dependant only on the *change* in response of the low-fidelity model. As such, when mapped across the existing high-fidelity samples, even if the current surrogate approximation has zero modelling error relative to observed samples, the functional response can still be shaped to more closely match that of the true response. Furthermore, the included weighting coefficients by definition do not impose an exact matching criteria at the location of each observed high-fidelity sample. Thus the final multi-fidelity response, while still exhibiting a quantifiable correction, may now contain sufficient errors at the high-fidelity sample points to construct an effective discrepancy function. The intent here is that the response shaping performed in this manner reduces the variability of the discrepancy function as in (Park et al., 2017), allowing a subsequent discrepancy correction stage to notably improve the final surrogate prediction quality.

3.1.2.3 Additive Bridge Function

The final included method is a standard additive bridge function (see Section 2.3.2.1). In this case, the direct discrepancy between the low- and high-fidelity responses at all observed high-fidelity points is used to generate an approximating surrogate of the discrepancy function. The matrix of translation vectors t may be given as:

$$\mathbf{t}_j = \mathbf{f}_L(\mathbf{x}_H^{(j)}) - \mathbf{f}_H(\mathbf{x}_H^{(j)})$$

An important consideration in this case is the relationship between surrogate training performance and number of training samples. In this work, should the number of high-fidelity samples (and therefore the number of direct translation vectors that may be calculated) be less than some predefined value, a data augmentation routine is applied in order to increase the likelihood of an effective training procedure. From recommendation and practice (Heath, 2020), the ideal number of samples is defined as $n_{desired} = (30 d_{in}/\phi_{trn})$, where the ratio of training data to input data $\phi_{trn} = 0.7$. This ensures sufficient samples are available for the training set after the dataset divisions. The augmented dataset of translation vectors $\tilde{\mathbf{y}}$ is defined as:

$$\tilde{\mathbf{y}} = [\mathbf{y}, \mathbf{A}] \tag{3.11}$$

where

$$\mathbf{A}_j = \mathbf{B}_j + w\boldsymbol{\theta}_j, \; j = 1, \ldots, n_{aug} \tag{3.12}$$

with $\mathbf{B} \subset \mathbf{y}$, the scalar weighting factor w chosen in this case as 1%, and $\boldsymbol{\theta}$ being a vector of uniformly distributed[3] random values $\boldsymbol{\theta} \in [-1, 1]$. The total number of data augments is $n_{aug} = n_{desired} - n_s$.

[3] Chosen to give a larger, more even spread of the resulting augmentations around each data point.

The resulting set of translations is used to train a set of ANN surrogate models according to a generalised bi-level procedure (described in more detail in Section 3.1.3). Networks are trained using Bayesian Regularisation, given the improved generalisation when considering higher-dimensional or highly non-linear responses. The target MSE for the training process is:

$$\mathrm{MSE}_{target} = \frac{\frac{1}{N}\sum_{i=1}^{N}(\mathbf{y}^{(i)} - \bar{\mathbf{y}})^2}{N_{trneq}} \tag{3.13}$$

with a minimum target gradient:

$$\mathbf{g}_{min} = \frac{\mathrm{MSE}_{target}}{100} \tag{3.14}$$

Training accuracy is measured using the normalised mean square error (3.23) and the corresponding R^2 value (3.24). Following the convergence of each network, performance is additionally examined by evaluating the prediction accuracy of the resulting multi-fidelity dataset obtained by applying the network of translation vectors to the original response data:

$$\hat{y}^{(i)} = y^{(i)} - \hat{\mathbf{t}}_i \tag{3.15}$$

A simplified visualisation of this approach is shown in Figure 3.4. Evaluating the R^2 of $\hat{y}$ relative to the original training set gives an independent evaluation of the translation network. On completion of the bi-level training process, the network exhibiting the strongest correction quality is selected and the complete multi-fidelity response is generated according to (3.15). Should this criteria contain more than one network, a further selection filter is applied by selecting that exhibiting the strongest training accuracy.

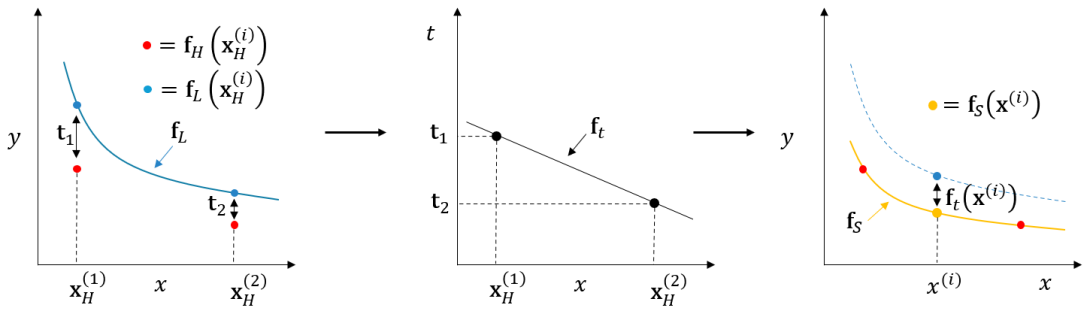


FIGURE 3.4: Visualisation of the Additive Bridge Function (A) correction stage.

3.1.3 Automated Surrogate Modelling using Artificial Neural Networks

Considering the generation of surrogate models for use within optimisation and optimal control problems, the goal is often to provide a global representation of the high-fidelity response meeting the requirements of zero- and first-order consistency (Alexandrov et al., 2001) to an acceptable degree within a reduced computational budget. The latter point is as compared

with constructing a surrogate of similar accuracy using exclusive high-fidelity sampling. First order consistency is desirable (though not always achievable) in this context given that, in conventional mathematics, if a surrogate and its derivative matches the target function, convergence to the same solution can be guaranteed (Alexandrov et al., 2001). Specifically considering continuous optimal control problems, it is highly desirable that system models are continuously differentiable given that a continuous/smooth representation of the dynamic system state is often required to reach feasible solutions.

Common methods of generating continuously differentiable surrogates suited to multi-fidelity applications are Kriging (Sacks et al., 1989) and *Response Surface Models* (RSMs) (Myers and Montgomery, 1995). However, these methods often require the user-specification of hyper-parameters such as (in the case of Kriging) regression and correlation functions (including an initial guess of the correlation function parameters), which may require *a priori* knowledge of the function under consideration. Whilst for physics-based subsystem models a certain degree of prior knowledge may be assumed, the most appropriate choice of hyper-parameters may not necessarily remain constant across separate subsystems. This work aims to implement a more generalised methodology, suitable for multiple models of unknown functional form within multiple subsystems.

In this study, surrogate model generation is accomplished using ANNs. Recall from Section 2.1.3, an ANN is a collection of simple computational units (neurons) arranged into an inter-connecting parallel processing structure (Beale et al., 2017). A network may be *trained* according to a finite set of corresponding input/target vectors. A suitably trained ANN may perform tasks such as pattern recognition, identification, classification, system control and function approximation (non-linear regression) (Beale et al., 2017). In aerospace engineering, ANNs have been used for the estimation of aerodynamic coefficients (Santos et al., 2008), space vehicle design and trajectory optimisation (Roshanian et al., 2019; Minisci and Vasile, 2013), turbo-machinery blade optimisation (Mengistu and Ghaly, 2008), wing design (Paiva et al., 2010), flow control, aeroelasticity, and interpolation of wind tunnel data (Yondo et al., 2018).

ANNs have found particular application in multi-fidelity modelling. Bakr et al. (2000) present a Space Mapping based neuro-modelling optimisation technique in which the high-fidelity optimum is found by replacing the low-fidelity model at each iteration with an adaptively-sized ANN trained using the high-fidelity samples from the previous iteration. Leary et al. (2003) introduced the *Knowledge-Based Neural Network* (KBNN) technique as a means of incorporating low-fidelity data into the training procedure of a limited set of high-fidelity samples via scaling/translation based parameter mapping. Minisci and Vasile (2013) estimate aerodynamic coefficients using a multi-fidelity ANN first trained on a global set of samples given by a simplified analytical model, then iteratively refined with CFD data scheduled via evolution control. Low-fidelity samples are sequentially discarded should they fall outwith a

predefined applicability region relative to new high-fidelity points.

More recently, advances in *Machine Learning* (ML)-based fields have prompted a huge increase in research devoted to incorporating multi-fidelity information within the structure of the ANN itself. A prominent example is that of multi-fidelity PINNs (Toscano et al., 2025). Generating a PINN can typically take one of three forms: transfer learning, simultaneous training and consecutive training (Howard et al., 2023). Transfer learning approaches first train a network on low-fidelity data before a correction term is subsequently applied (Song and Tartakovsky, 2022; De et al., 2020). Simultaneous training involves learning the linear/non-linear correlations between the low and high-fidelity networks by constructing and maintaining separate networks of each (Meng and Karniadakis, 2020). Consecutive training operates similarly, in that several networks representing the low and high-fidelity data can be trained first, with an additional network added to learn the correlation between the two (Liu and Wang, 2019). In this scenario however, the final network is trained after the others are completed, rather than all at once.

Other examples of multi-fidelity PINNs include the work by Regazzoni et al. (2021) which extends the methodology presented by (Meng and Karniadakis, 2020) to Bayesian networks in order to help quantify the uncertainty in the resulting prediction. The authors first employ an ANN to approximate the low-fidelity data, then use an additional network to learn the correction term using both the high-fidelity data and a physics-informed loss. Wang and Lai (2024) introduces a multi-stage ANN that divided the training process into sequential stages, where each involved a new network fit to the residuals of the previous, and Penwarden et al. (2022) offers an approach to combine different ANNs of varying degrees of fidelity by exploiting their low-rank structure.

Whilst several of the principles described in these works are incorporated into the work presented in this thesis (for instance, the staged approach to multi-fidelity learning), it was of interest to the current study to decouple the incorporation of multi-fidelity data from the surrogate modelling process where possible. This is primarily for to two reasons: 1) to allow an unbiased examination of the various presented multi-fidelity techniques in a more generalised manner of application, and 2) to allow the complete multi-fidelity design framework designed as part of this work to incorporate a modular format, whereby specific multi-fidelity techniques and surrogate modelling methods can be swapped out independently from each other. Furthermore, were multiple methods to be applied simultaneously for a given problem (for example, if employing an *Ensemble of Metamodels*-based approach), this modular format easily allows different multi-fidelity techniques to be paired with different surrogate modelling approaches. For these reasons, standard shallow-ANNs are employed as the surrogate modelling process throughout this work, where any multi-fidelity correction/tuning of data occurs outwith the structure of the network itself.

Some relevant advantages of ANNs for generalised multi-fidelity surrogate modelling applications are as follows:

- Automated and computationally efficient optimisation of expensive computational models remains an open problem (Koziel and Leifsson, 2016). Thus, surrogate modelling techniques supporting higher levels of automation, such as ANNs, are highly desirable to support such applications.
- ANNs are non-parametric in the sense that they do not require any *a priori* specification of assumed functional relationships or model hyper-parameters by the user. This is in contrast to other, designer-centric modelling techniques such as Kriging.
- Related closely to the previous point, ANNs can be applied in a general manner. That is, they can operate with minimal, or in the absence of, *a priori* knowledge regarding the target function structure.

A generic bi-level training methodology for multi-layer feed-forward neural networks is presented, largely adapted from the recommendations of Heath (2020). This formulation is an attempt to automatically determine the optimal number of neurons within the single hidden layer of a standard MLP architecture (chosen due to an expected better generalisation in regions potentially far from the training set), thus enabling a more automated surrogate construction process applicable to general non-linear regression problems. The method utilises both Bayesian Regularisation and early stopping criteria, in conjunction with a normalised performance metric, implemented using the MATLAB® Neural Network Toolbox (Beale et al., 2017). Designed analogous to a black-box function, sample sets are provided as input with the corresponding surrogate model(s) produced as output. A standard training process is conducted via 4 basic steps:

1. Assemble training data
2. Initialise the network weights and biases
3. Train the network
4. Validate the network response to new inputs

Let $\mathbf{x} \in \mathbb{R}$ denote the sampling plan such that

$$\mathbf{x}^{(i)} \mapsto \mathbf{y}^{(i)} = f(\mathbf{x}^{(i)}), \quad i = 1, 2, \ldots, n_s \tag{3.16}$$

where

$$\mathbf{x} = [\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n_s)}]^T$$
$$\mathbf{y} = [y(\mathbf{x}^{(1)}), \ldots, y(\mathbf{x}^{(n_s)})]^T$$

In this work, initial sample sets are generated using LHS to ensure that space-filling criteria can be met across the entire input domain for any sample size n_s. Within the proposed procedure, the input data is subsequently distributed amongst training, validation and testing sets, where the number of samples in each, n_{trn}, n_{val} and n_{tst} respectively, conforms to the following ratios:

$$\begin{bmatrix} n_{trn} \\ n_{val} \\ n_{tst} \end{bmatrix} = n_s \begin{bmatrix} \phi_{trn} \\ \phi_{val} \\ \phi_{tst} \end{bmatrix} = n_s \begin{bmatrix} 0.7 \\ 0.15 \\ 0.15 \end{bmatrix} \tag{3.17}$$

The number of samples n_s for an input vector of dimension d_{in} is chosen such that:

$$10d_{in} \leq n_{trn} \leq 30d_{in} \tag{3.18}$$

The initialisation process allocates random initial weights and biases to the network, and performs a division of the available input data into training, validation and testing sets according to the specified division function and ratios. This random initialisation coupled with the gradient descent method used in back-propagation, as discussed earlier, means that the network solution may differ between any given number of random initialisations. While some authors advocate the practise of re-training the network multiple times, or averaging the response from multiple networks, these methods still ultimately rely on a certain degree of user discretion.

The number of unknown network parameters (weights and biases) N_w is given by:

$$N_w = H(d_{in} + 1) + d_{out}(H + 1) \tag{3.19}$$

where H is the number of neurons in the hidden layer. The number of training equations required for a given training set is given by:

$$N_{trneq} = n_{trn} * d_{out} \tag{3.20}$$

where the number of samples in the training set N_{trn} is determined by the specified data division ratios. To avoid over-fitting,

$$N_w \leq N_{trneq} \tag{3.21}$$

By equating equations (3.19) and (3.20) it is clear that (3.21) will always be true provided the number of neurons H is less than or equal to an upper bound H_{ub}, given by:

$$H_{ub} = \left\lceil \frac{d_{out}(n_{trn} - 1)}{d_{in} + d_{out} + 1} \right\rceil \tag{3.22}$$

In this work, a bi-level strategy for network training and parameter selection is employed. In the outer level, the number of neurons in the single hidden layer is varied in $n_{HL} = 10$ near-equally spaced divisions between a lower bound of 0 (representing a linear response) and the upper bound H_{ub}. An inner level then performs $n_{init} = 10$ random initialisations of the network weights, biases and data divisions[4], before sequentially training each network to convergence and saving to a database.

In each case, the objective is to determine the network parameters that yield an acceptable ratio of mean-square-error to mean-target-variance for random subsets of non-training data (Heath, 2020). Training performance is tracked via the network errors recorded for the independent testing set of size n_{tst} (3.17). A normalised performance criteria relating N_t predictions $\boldsymbol{\alpha}$ and targets $\mathbf{t}$ facilitates network selection:

$$\text{NMSE} = \frac{\frac{1}{N}\sum_{i=1}^{N}(t_i - \alpha_i)^2}{\frac{1}{N}\sum_{i=1}^{N}(t_i - \bar{t})^2} \tag{3.23}$$

The resulting global evaluation criteria coefficient of determination R^2 is calculated:

$$R^2 = 1 - \text{NMSE} \tag{3.24}$$

To reduce the risk of a poor generalisation (overfitting), it is of interest to minimise the number of neurons contained within the hidden layer. A minimum performance criteria R_{target} is defined as the successful modelling of 99% of the target subset variance ($R^2 \geq 0.99$). Therefore, the network that achieves this criteria for a minimum number of hidden layer neurons is selected.

In summary, the proposed automated ANN surrogate modelling process is dependant on the following list of user-definable hyperparameters given in Table 3.2:

Parameter	Definition	Default value
$[\phi_{trn}, \phi_{val}, \phi_{tst}]$	Training, validation and testing set proportions	$[0.7, 0.15, 0.15]$
n_{HL}	Number of divisions between the minimum and maximum number of hidden layers to be tested	10
n_{init}	Number of random initialisations per tested network	10
R_{target}	Minimum R^2 of the final surrogate	99

Table 3.2: List of user-specifiable hyperparameters for the proposed automated ANN surrogate modelling process.

3.1.4 Numerical Tests

Despite the continued development of a wide variety of methods for multi-fidelity surrogate modelling, there currently exists little in the way of standardised approaches to compare and

[4]Whilst the random division of input data in (3.17) means that the resulting subsets cannot be thus assumed space filling themselves, the variation of subset assignment within the inner-level loop is intended to minimise any resulting bias.

rigorously assess alternative methods (Fernández-Godino et al., 2016; Peherstorfer et al., 2015; Mainini et al., 2022). An attempt was made by Beran et al. (2020) to classify numerical test cases for multi-fidelity methods into three general categories:

- *L1 problems*: computationally inexpensive/efficient analytical functions with exact solutions.
- *L2 problems*: simplified engineering application problems that can be executed with a reduced computational expense.
- *L3 problems*: more complex engineering use-cases, typically incorporating multi-physics coupling.

With reference to these classifications, Mainini et al. (2022) compile a comprehensive suite of *L1*-class numerical testing functions for multi-fidelity surrogate modelling techniques. Collected from the literature, these benchmark functions represent purely analytical problems, with no explicit association to real-life applications. However, the premise for selection was to broadly exemplify certain mathematical characteristics often encountered in simulation-based optimisation problems and related real-world applications. Of particular interest is behaviour that challenges the accurate modelling of the underlying process (which may in turn hinder the search and identification of optimal solutions if employed within an optimisation framework, for example). Such challenges include issues associated with scalability and dimensionality, the existence of localised, multi-modal, and discontinuous behaviour and the existence of noise.

This work selects several numerical test functions based on the above recommendations, for the purpose of assessing and comparing the described multi-fidelity correction techniques, including their applicability as constituents of a multi-stage response adaptation process. Testing functions are selected to assess weaknesses and strengths of multi-fidelity methods with regards to the mathematical challenges described above.

The numerical testing strategy employed in this section is as follows. The multi-fidelity correction techniques listed in Section 3.1.2 are used to define a set of single- and multi-stage correction processes. The performance of each process is evaluated by measuring the obtainable prediction accuracy for each selected benchmark function when the number of available high-fidelity samples is sequentially increased towards an appropriate upper bound of sample density. To provide two representative performance baselines, results are compared with the predictions obtained via a single-fidelity surrogate modelling technique employing only the available high- and low-fidelity data in turn.

It is acknowledged that true global accuracy generally cannot be practically achieved using high-fidelity simulations owing to the prohibitive computational expense (Mainini et al., 2022). However, it is hoped that the qualities and behaviours of the included methods within the context of global accuracy can be adequately assessed using the proposed analytical benchmarks. The following section details the formulations of the selected benchmark functions, sampling strategy and validation metrics.

3.1.4.1 Benchmark Functions

The following numerical test functions are intended to represent a broad range of potential applications of multi-fidelity methodology, capturing fundamental mathematical characteristics and properties that mimic real-world engineering problems (Mainini et al., 2022). This behaviour could include:

- *Atypical local behaviour* such as multi-modality and discontinuities that cannot be neglected *a priori* (particularly when considering real-world problems, where the representation of the complete variable domain could be especially important).
- *Function scalability*, with respect to both the parameterisation of the function and the range of available fidelities. The latter may be used to demonstrate the available improvement from modifying the fidelity level.
- The *type of discrepancy* between fidelities. Quantification of this type allows a deeper assessment of the multi-fidelity methods, since the number of fidelities available must be correlated to the associated discrepancy type. For example, a linear discrepancy is easier to model than a non-linear discrepancy.
- The existence of *noise*. In real-world problems, embedded noise may be inescapable, particularly in systems that can exhibit abrupt changes within the output domain. The effects of noise on a system response are typically difficult to directly identify, let alone eliminate, thus the ability of multi-fidelity methods to model noisy functions is of practical interest.

A function may be generally represented as $f(\mathbf{x}) \in \mathbb{R}$ with input vector $\mathbf{x} \in \mathbb{R}^{d_i n}$, where d_{in} is the dimensionality of the parameter space. This definition can also be applied to include problem constituent system/subsystem models requiring approximation, or the objective function(s) of an outer-loop optimisation process. With regards to the overarching multi-fidelity context, each benchmark function is presented with a number of available fidelity levels. In each case, $f_1(\mathbf{x})$ is taken as the highest fidelity level, with all possible L alternatives considered the cheaper, less accurate approximations, such that $f_L(\mathbf{x})$ represents the lowest

available fidelity level. Whilst for some functions more than two fidelity levels are available, for the purposes of comparison only two levels are considered for each test case. The selected test functions are summarised below in Table 3.3.

Test function	d	behaviour	scalability	discrepancy	noise	LF correlation R^2	LF correlation r^2
Forrester (Forrester et al., 2007)	1	local, multi-modal	parametric	(non)linear	no	-120.9613	0.5233
Sasena (Sasena, 2002)	1	local	-	linear	no	-28.6617	0.6682
Currin (Currin et al., 1991)	2	local	parametric	nonlinear	no	-13.2682	0.9657
Rosenbrock (Mainini et al., 2022)	2	local, uni-modal	parametric	nonlinear	no	-1.21e+04	0.9898
Rastrigin (shifted) (Mainini et al., 2022)	2	multi-modal	parametric/fidelity	nonlinear	no	31.1097	0.8779
Heterogeneous (Clark et al., 2016)	2	local/multi-modal	parametric	nonlinear	no	-953.58	0.8831
Hartman (Hartman, 1973)	6	multi-modal	-	nonlinear	no	-26.4120	0.6125
Paciorek (Toal, 2015)	2	multi-modal	fidelity	nonlinear	yes	-34.6552	0.1398
Borehole (Morris et al., 1993)	8	local	fidelity	nonlinear	no	71.8684	0.9972
Park2 (Park, 1991)	4	local	parametric	nonlinear	no	65.2837	0.9965

Table 3.3: Numerical test function summary, including the correlation between the included low- and high-fidelity approximations in terms of R^2 and Pearson's correlation coefficient r^2 values.

3.1.4.2 Sampling

Given the importance of sampling criteria within the context of surrogate modelling in general (see Section 2.2), to ensure a fair comparison it is of interest to examine the selected correction techniques with as little bias given to sampling strategy as possible. While by this definition it may be functionally most effective to employ a classical DoE technique such as *Full Factorial Design* (FFD) or *Central Composite Design* (CCD), the overriding context of multi-fidelity modelling requires the consideration of cases in which high-fidelity function evaluations are minimised. Depending on dimensionality, this is likely to compromise the space-filling properties of most classical methods necessary to ensure unbiased distributions. Therefore, this work adopts a standardLHS approach (see Section 2.2). Recall, LHS is popular in the surrogate modelling community largely due to two significant properties; (1) space-filling criteria can be met across the entire input domain, allowing the representation of underlying function structure with a minimum number of function evaluations (Yondo et al., 2018), and (2) the structural distribution of samples exhibits a non-collapsing property, i.e. the latin-hypercube definition can still be met with a low number of samples (Liu et al., 2018a). The latter implies that sample size is largely independent of problem dimensionality, a particularly useful property when considering costly analyses/simulations (Tyan and Lee, 2019).

The sampling plan attempts to achieve two compromises: balancing computational cost with accuracy, and balancing consistency with low bias (Mainini et al., 2022). Despite the advantages of LHS, if one assumes no prior knowledge regarding the target function to be approximated, it can be difficult to estimate the optimal, or even appropriate, initial sampling size such that these compromises can be achieved and computational resource is not unduly wasted.

Furthermore, standard LHS may result in non-uniform distributions[5]. For example, samples allocated along the design space diagonal satisfy both conditions (1) and (2). Accordingly, each technique is applied over a series of sequentially increasing LHS-generated sample sets of high-fidelity data, where individual performance is assessed at each computation. The design spaces corresponding to the (inexpensive) analytical benchmark functions (for all available levels of pseudo-fidelity) are exhaustively sampled to allow a more rigorous assessment of method accuracy and performance.

Real-world design problems are constrained by the limited time and computing resources available to conduct numerical/experimental analyses of alternative solutions. This motivates a recommendation to conduct the relevant analyses at specified computational budgets assigned to the overall modelling task, rather than prescribing a maximum number of allowed iterations, for example. For the purposes of the prescribed numerical tests, computational budget is defined by a maximum allowable number of high-fidelity samples for each benchmark. The termination condition is thus the complete exhaustion of the computational budget ascribed to each.

The included correction techniques, surrogate modelling methodology and sampling strategy contains several sources of stochastic variability. Importantly then, to ensure statistical significance throughout the described numerical testing procedure, all computations are performed across 100 independent runs, each initialised according to a unique *Random Number Generator* (RNG) seeding. Reported results are thus the averaged values across all runs.

3.1.4.3 Validation Metrics

Multi-fidelity validation metrics can be described in two ways (Mainini et al., 2022): goal-sensitive and goal-insensitive. Goal-sensitive metrics are concerned with the outer-loop result of the overriding application, e.g. evaluating the accuracy or location of the optima computed using multi-fidelity approximations[6]. On the other hand, goal-insensitive metrics are more concerned with the global accuracy of multi-fidelity approximations across the design space, thus are more general to both outer-loop and many-query applications. Goal-insensitive metrics should exhibit the following properties (Mainini et al., 2022):

- High consistency: all methods should be evaluated in the same way.
- Low bias: the evaluation strategy should not be biased.
- High utility: the evaluation of metrics should be highly accurate and informative.

[5] Numerous modifications to LHS have been proposed (Beachkofski and Grandhi, 2002; Leary et al., 2003; Ye, 1998; Palmer and Tsui, 2001) to address this issue and provide more uniform sampling distributions.

[6] Importantly, goal-sensitive metrics generally require incorporating information regarding the goal itself (e.g. the problem optima), assumed to be known *a priori*.

- Affordable cost: computing metrics should be affordable.

Based on these criteria, the following performance metrics are selected to offer a comprehensive framework capable of comparing a broad spectrum of alternative multi-fidelity modelling methods regardless of specific features or idiosyncratic behaviour:

- the *Relative Maximum Absolute Error* (RMAE)

$$\text{RMAE} = \frac{1}{\sigma} \max |y^{(i)} - \hat{y}^{(i)}|, i = 1, \ldots, N$$

- the *Relative Root Mean Square Error* (RRMSE)

$$\text{RRMSE} = \frac{1}{\sigma} \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y^{(i)} - \bar{y})^2}$$

- the global evaluation criteria coefficient of determination R^2

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y^{(i)} - \hat{y}^{(i)})^2}{\sum_{i=1}^{N} (y^{(i)} - \bar{y})^2}$$

- the square of Pearson's correlation coefficient r^2

$$r^2 = \left(\frac{\sum_{i=1}^{N} (y_H^{(i)} - \bar{y_H})(y_L^{(i)} - \bar{y_L})}{\sqrt{\sum_{i=1}^{N} (y_H^{(i)} - \bar{y_H})^2} \sqrt{\sum_{i=1}^{N} (y_L^{(i)} - \bar{y_L})^2}} \right)^2$$

where each is evaluated with respect to sample size N. The model response at test point i is represented by $y^{(i)}$, the average model response across all test points is $\bar{y}$, and $\hat{y}$ is the surrogate response. Standard deviation σ is given by:

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (y^{(i)} - \bar{y})^2} \tag{3.25}$$

The square of Pearson's correlation coefficient r^2 (Toal, 2015; Shi et al., 2020a) describes the correlation between the high- and low-fidelity responses. Of significance is the invariance of this coefficient under separate changes in location and scale between the two functions. An r^2 value of 1 represents models related via a simple linear equation, and thus exhibiting a flatter error surface profile. The implication here is that even for an approximate model with relatively poor prediction capability, indicated by a lower R^2 or higher average error values, a strong geometric correlation between it and the 'true' response may ultimately still produce a strong multi-fidelity approximation via relatively simple correction techniques.

One final consideration is the *Cross-Validation* (CV) error (see Section 2.1.6). The commonly used *Leave-One-Out* (LOO) CV error can prove prohibitively expensive to evaluate given only one sample is omitted at a time. Alternatively, samples may be omitted in groups. K-fold CV is a popular validation metric for cases in which the generation of data is computationally expensive or difficult (Forrester and Keane, 2009; Eason and Cremaschi, 2014). The observed sample set is randomly divided into K equal sized subsets. Then, K surrogates are trained, where one subset is omitted from the training data each time and used as validation data. The sum of squared error is computed for each point when it is in the validation subset. The sum of squared CV errors for the full data set is used as an indicator of the prediction quality. Here, K-fold CV is included as an additional validation metric, with $k = 5$. Of note, whilst the CV error performs well for selecting surrogates (Viana and Haftka, 2008), Park et al. (2017) find that this does not carry over to multi-fidelity frameworks.

Most often the primary objective for incorporating multi-fidelity modelling is to reduce the overall computational cost, and therefore simulation time, required to achieve a satisfactory representation of the process of interest. Of critical importance then is the accurate reporting of relevant cost metrics, including time savings estimates, in a general manner such that results and conclusions may be reliably compared across the literature (Fernández-Godino et al., 2016, 2019). However, it should be noted that the performance of many multi-fidelity surrogate generation methods is inherently problem dependant. Indeed, Fernández-Godino et al. (2016) conclude that without adopting a similar problem structure, reported cost savings for any given methodology may vary widely between studies. Two important metrics are therefore included in the analysis of each test case. The first is the percentage reduction in RRMSE between the single-fidelity and multi-fidelity methods for several distinct, fixed levels of computational cost (in this case, number of high-fidelity samples). That is, *accuracy for fixed-cost*. The second is the number of extra high-fidelity evaluations needed for the single-fidelity approach to match or exceed the RRMSE acheived by the multi-fidelity approach at each distinct level of computational cost. That is, *cost for fixed-accuracy*.

A further consideration can be given to the computational effort required to construct the multi-fidelity surrogates in the first place, in relation to their associated cost benefits. Whilst for the following numerical tests, the allocation of representative computational costs between fidelities is purely arbitrary, the computation time to generate each surrogate is tracked in each case. Furthermore, the computation time to perform the multi-fidelity correction (which for some processes includes the generation of intermediate bridging surrogates) is also tracked for completeness.

3.1.4.4 Results

This section presents the results obtained for the numerical tests. For each benchmark function, the formulation is presented for the "high-fidelity" response and at least one corresponding "low-fidelity" response. For lower-dimensional cases, the high- and low-fidelity responses are depicted graphically. Plots of the RRMSE, RMAE and determination coefficient R^2 are presented. Case-specific results are summarised in a table including the best percentage reduction in RRMSE demonstrated by any of the included multi-fidelity methods over the single-fidelity approach, and the corresponding cost savings (in terms of both percentage reduction and number of high-fidelity samples). Finally, the progression of *K-Fold Cross Validation Error* (kCVE) is shown for all multi-fidelity methods. Section 3.1.4.5 then includes a summary discussion and analysis of all presented results.

Sasena Test Function The one-dimensional sinusoidal test function originally presented in (Sasena, 2002) is given by:

$$y_H = -\sin x - \exp(x/100) + 10 \tag{3.26}$$

Huang et al. (2006) and Zhou et al. (2017) have employed a low-fidelity counterpart of the form:

$$y_L = -\sin x - \exp(x/100) + 10.3 + 0.03 * (x-3)^2 \tag{3.27}$$

where $x \in [0, 10]$ for both y_H and y_L. Equations (3.26) and (3.27) exhibit a fairly strong spatial correlation with matching stationary points, shown in Figure 3.5, but with opposing locations of the global minimum.

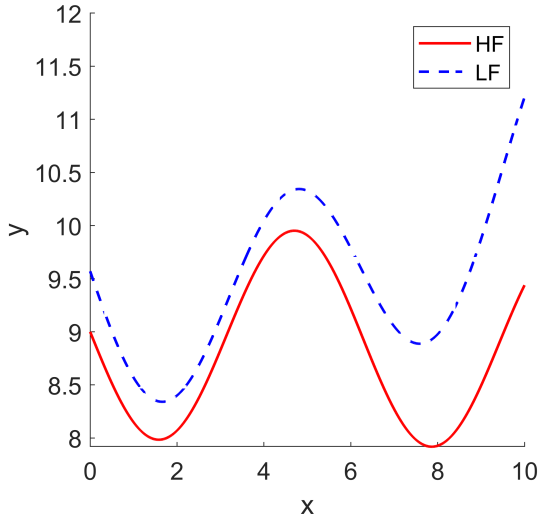


FIGURE 3.5: Sasena test function, including high-fidelity response (solid line) and low-fidelity response shown (dashed line).

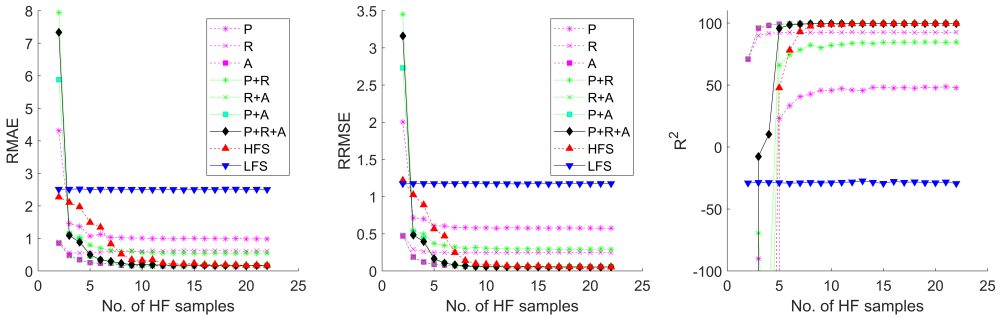
Summary results are included in Table 3.4. The best observed RRMSE values obtained using any multi-fidelity method are presented at progressively increasing levels of computational budget. For each level, the corresponding performance of the single-fidelity is included for reference. The percentage improvement in RRMSE over the single-fidelity approach is then given, as well as the cost savings considering the number of extra high-fidelity samples required for the single-fidelity approach to reach an equivalent or superior level of accuracy.

computational budget (n_{HF}/d)	5	10	15	20
Best MFM RRMSE	0.083 (0.035) (RA)	0.052 (0.016) (RA)	0.052 (0.013) (RA)	0.052 (0.017) (PRA)
HFS RRMSE	0.567 (0.130)	0.097 (0.049)	0.059 (0.021)	0.057 (0.021)
% RRMSE reduction	85.318	46.788	11.901	10.213
% cost saving (n_{HF}/d for equivalent RRMSE)	54.545 (11)	>54.55 (>22)	>31.82 (>22)	>9.09 (>22)

*calculated as $(1 - A/B) * 100$

Table 3.4: Summary results for the Sasena test case ($d = 1$), where n_{HF} is the number of high-fidelity samples. Standard deviations of the RRMSE values are included in brackets.

Figure 3.6 shows the progression of RRMSE, RMAE and R^2 for each multi-fidelity method relative to the number of high-fidelity samples. Included for reference are the values pertaining to the single-fidelity surrogate constructed using only the available high-fidelity and low-fidelity samples in turn.



(a) RMAE (b) RRMSE (c) R^2

FIGURE 3.6: Performance metrics of the Parametric (P), Relative (R) and Additive (A) corrective stages applied individually and in sequential combinations to the Sasena numerical test case relative to the number of high-fidelity evaluations, where (A) is the RMAE (B) is the RRMSE and (C) is R^2. Included for reference are the error values of the low-fidelity model, and the single-fidelity surrogate trained only on the available high-fidelity samples at each iteration.

Finally, Figure 3.7 details the progression of kCVE (where $k = 5$) for each multi-fidelity method relative to the number of high-fidelity samples.

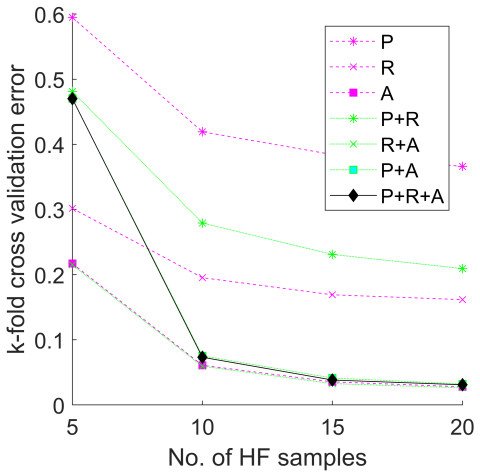


FIGURE 3.7: kCVEs for the Sasena test case.

Forrester Test Function Forrester et al. (2007) present a simple multi-modal function with one global minimum, one local minimum and a zero-gradient inflection point (Figure 3.8):

$$f_H(x) = (6x - 2)^2 \sin(12x - 4) \tag{3.28}$$

The alternative low-fidelity approximation is given by:

$$f_L(x) = Af_H(x) + B(x - 0.5) + C \tag{3.29}$$

where $x \in [0, 1]$ for both y_H and y_L. The constants A, B and C can be varied to modify the correlation between the low- and high-fidelity function, with default values $A = 0.5$, $B = 10$ and $C = -5$.

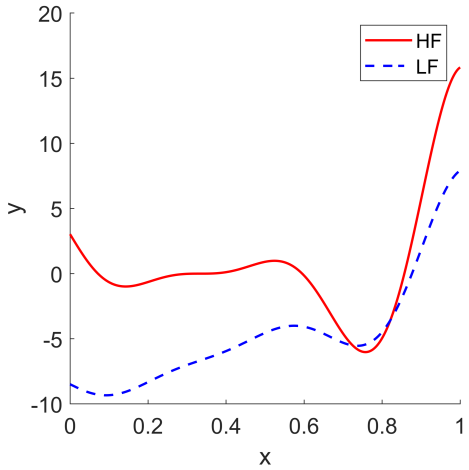


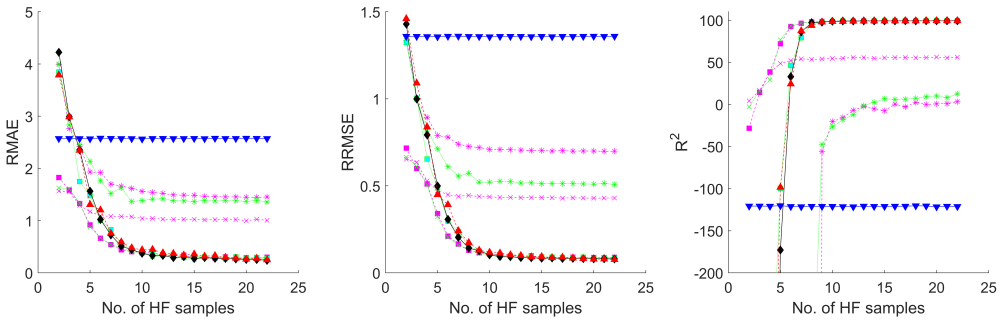
FIGURE 3.8: Forrester test function, including high-fidelity response (solid line) and low-fidelity response shown (dashed line).

As for the previous case, summary results are given in Table 3.5. Plots of RRMSE, RMAE and R^2 are given in Figure 3.9 and Figure 3.10 shows the kCVE.

computational budget (n_{HF}/d)	5	10	15	20
Best MFM RRMSE	0.323 (0.126) (RA)	0.104 (0.027) (PRA)	0.082 (0.014) (RA)	0.080 (0.010) (PRA)
HFS RRMSE	0.510 (0.295)	0.113 (0.030)	0.084 (0.023)	0.080 (0.020)
% RRMSE reduction	36.637	7.746	2.329	-0.354
% cost saving (n_{HF}/d for equivalent RRMSE)	28.57 (7)	16.67 (12)	21.05 (19)	0 (20)

*calculated as $(1 - A/B) * 100$

Table 3.5: Summary for the Forrester test case ($d = 1$), where n_{HF} is the number of high-fidelity samples. Standard deviations of the RRMSE values are included in brackets.



(a) RMAE (b) RRMSE (c) R^2

FIGURE 3.9: Performance metrics of the Parametric (P), Relative (R) and Additive (A) corrective stages applied individually and in sequential combinations to the Forrester numerical test case relative to the number of high-fidelity evaluations, where (A) is the RMAE (B) is the RRMSE and (C) is R^2. Included for reference are the error values of the low-fidelity model, and the single-fidelity surrogate trained only on the available high-fidelity samples at each iteration.

Currin Exponential Function The following 2-dimensional function was presented by Currin et al. (1991):

$$f_H(\mathbf{x}) = \left[1 - \exp\left(-\frac{1}{2x_2}\right)\right] * \frac{2300x_1^3 + 1900x_1^2 + 2092x_1 + 60}{100x_1^3 + 500x_1^2 + 4x_1 + 20} \tag{3.30}$$

where $x_i \in [0, 1]$ for $i = 1, 2$. For the purpose of multi-fidelity simulation, Xiong et al. (2013) define a low-fidelity counterpart as:

$$f_L(\mathbf{x}) = \frac{1}{4}[f_H(x_1 + 0.05, x_2 + 0.05) + f_H(x_1 + 0.05, \max(0, x_2 - 0.05))] \\ * \frac{1}{4}[f_H(x_1 - 0.05, x_2 + 0.05) + f_H(x_1 - 0.05, \max(0, x_2 - 0.05))] \tag{3.31}$$

As for the previous cases, summary results are given in Table 3.6. Plots of RRMSE, RMAE and R^2 are given in Figure 3.12 and Figure 3.13 shows the kCVE.

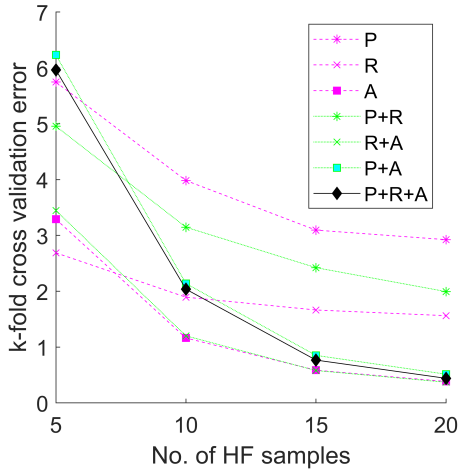


FIGURE 3.10: kCVEs for the Forrester test case.

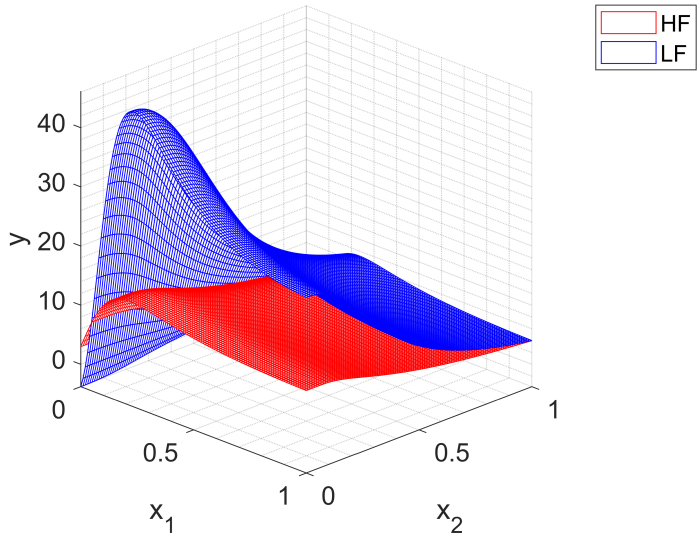


FIGURE 3.11: Currin exponential test function, including high-fidelity response (red) and low-fidelity response shown (blue).

Park Test Function The 4-dimensional function presented by Park (1991) has been used as an illustrative example in several papers (Cox et al., 2001; Xiong et al., 2013) outlining methods for adaptively calibrating models using variable fidelity information:

$$f_H(\mathbf{x}) = \frac{2}{3}\exp(x_1 + x_2) - x_4 \sin(x_3) + x_3 \tag{3.32}$$

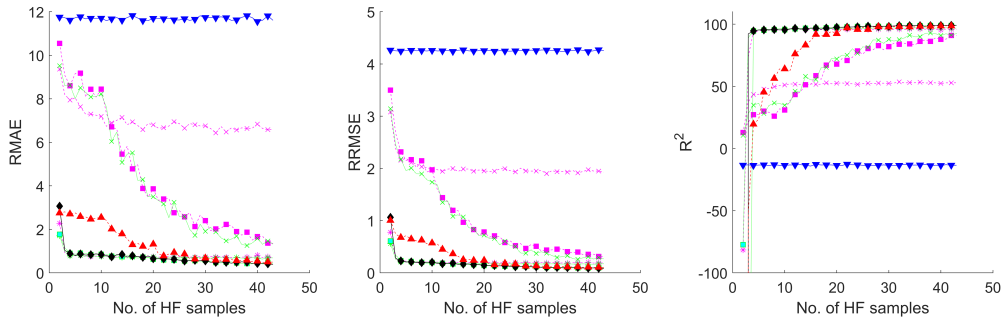
where $x_i \in [0, 1]$ for all $i = 1, \ldots, 4$. The low-fidelity counterpart is given by (Xiong et al., 2013):

$$f_L(\mathbf{x}) = 1.2 f(\mathbf{x}) - 1 \tag{3.33}$$

computational budget (n_{HF}/d)	5	10	15	20
Best MFM RRMSE	0.188 (0.042) (PR)	0.140 (0.057) (PA)	0.115 (0.058) (PRA)	0.091 (0.029) (PRA)
HFS RRMSE	0.554 (0.169)	0.225 (0.139)	0.118 (0.043)	0.108 (0.045)
% RRMSE reduction	66.072	37.795	2.477	15.593
% cost saving (n_{HF}/d for equivalent RRMSE)	50 (10)	20 (12.5)	21.05 (19)	>6.98 (>21.5)

*calculated as $(1 - A/B) * 100$

Table 3.6: Summary for the Currin Exponential test case ($d = 2$), where n_{HF} is the number of high-fidelity samples. Standard deviations of the RRMSE values are included in brackets.



(a) RMAE (b) RRMSE (c) R^2

FIGURE 3.12: Performance metrics of the Parametric (P), Relative (R) and Additive (A) corrective stages applied individually and in sequential combinations to the Currin numerical test case relative to the number of high-fidelity evaluations, where (A) is the RMAE (B) is the Relative RRMSE and (C) is R^2. Included for reference are the error values of the low-fidelity model, and the single-fidelity surrogate trained only on the available high-fidelity samples at each iteration.

As for the previous cases, summary results are given in Table 3.7. Plots of RRMSE, RMAE and R^2 are given in Figure 3.14 and Figure 3.15 shows the kCVE for each method.

computational budget (n_{HF}/d)	5	10	15	20
Best MFM RRMSE	0.074 (0.012) (PA)	0.069 (0.010) (RA)	0.065 (0.010) (A)	0.064 (0.011) (PA)
HFS RRMSE	0.124 (0.036)	0.087 (0.019)	0.075 (0.015)	0.071 (0.014)
% RRMSE reduction	40.571	20.824	13.696	9.736
% cost saving (n_{HF}/d for equivalent RRMSE)	69.231 (16.25)	>53.49 (>21.5)	>30.23 (>21.5)	>6.98 (>21.5)

*calculated as $(1 - A/B) * 100$

Table 3.7: Summary for the Park test case ($d = 4$), where n_{HF} is the number of high-fidelity samples. Standard deviations of the RRMSE values are included in brackets.

The Rosenbrock Function The Rosenbrock function is a well-known D-dimensional optimisation benchmark problem described by (Madsen et al., 2004; Fletcher, 2013; Bryson

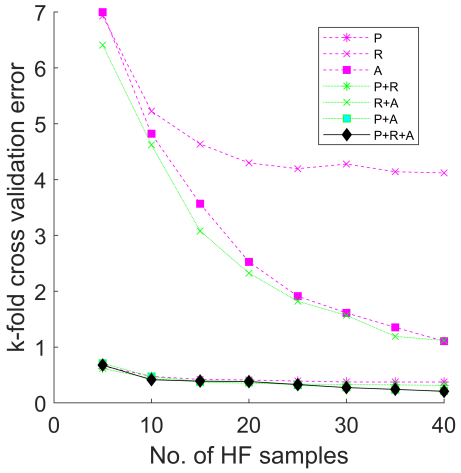
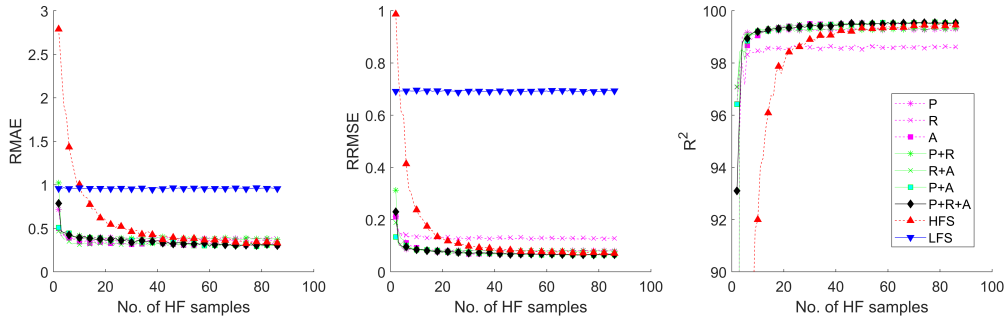


FIGURE 3.13: kCVEs for the Currin test case.



(a) RMAE (b) RRMSE (c) R^2

FIGURE 3.14: Performance metrics of the Parametric (P), Relative (R) and Additive (A) corrective stages applied individually and in sequential combinations to the Park numerical test case relative to the number of high-fidelity evaluations, where (A) is the RMAE (B) is the RRMSE and (C) is R^2. Included for reference are the error values of the low-fidelity model, and the single-fidelity surrogate trained only on the available high-fidelity samples at each iteration.

and Rumpfkeil, 2016a):

$$f_1(\mathbf{x}) = \sum_{i=1}^{D-1} 100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2 \tag{3.34}$$

The function is uni-modal and contains one local minimum at $\mathbf{x} = \{-1, 1, \ldots, 1\}^T \text{for} 4 \leq D \leq 7$, in addition to the global minimum $\mathbf{x}^* = \{1, \ldots, 1\}^T$ located within a narrow, parabolic valley. The variable domain is defined, for this case, as $-2 \leq x_i \leq 2 \text{for} i = 1, \ldots, D$. Two alternative fidelity levels (denoted medium-fidelity (Ficini et al., 2021) and low-fidelity (Bryson

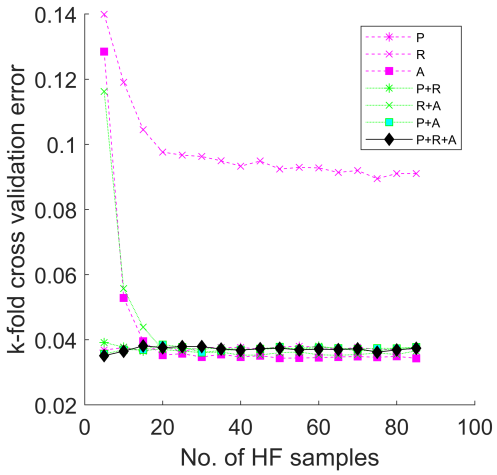


FIGURE 3.15: kCVEs for the Park test case.

and Rumpfkeil, 2016b) respectively) can be expressed as:

$$f_2(\mathbf{x}) = \sum_{i=1}^{D-1} 50(x_{i+1} - x_i^2)^2 + (-2 - x_i)^2 - \sum_{i=1}^{D} 0.5x_i \tag{3.35}$$

$$f_3(\mathbf{x}) = \frac{f_1(\mathbf{x}) - 4 - \sum_{i=1}^{D} 0.5x_i}{10 + \sum_{i=1}^{D} 0.25x_1} \tag{3.36}$$

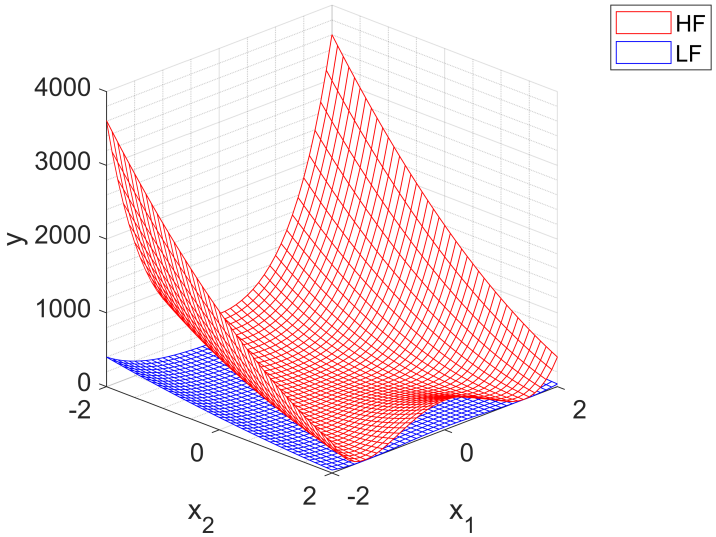


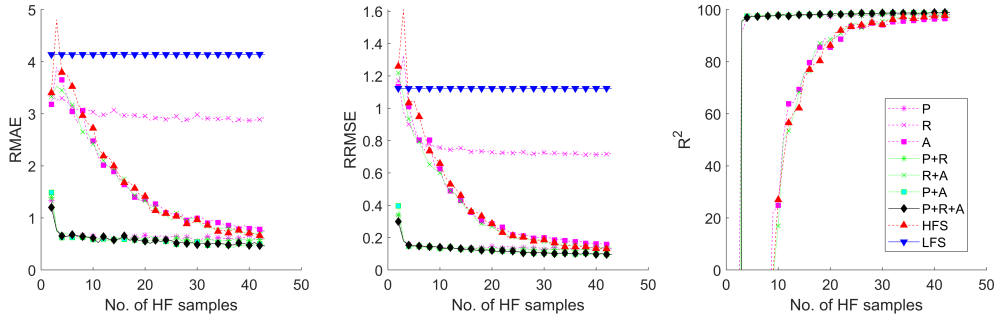
FIGURE 3.16: Rosenbrock test function, including high-fidelity response (red) and low-fidelity response shown (blue).

As for the previous case, summary results are given in Table 3.8. Plots of RRMSE, RMAE and R^2 are given in Figure 3.17 and Figure 3.18 shows the kCVE for each method.

computational budget (n_{HF}/d)	5	10	15	20
Best MFM RRMSE	0.130 (0.036) (PR)	0.120 (0.033) (PA)	0.104 (0.029) (PRA)	0.097 (0.025) (PRA)
HFS RRMSE	0.705 (0.288)	0.269 (0.111)	0.179 (0.079)	0.131 (0.045)
% RRMSE reduction	81.61	55.30	42.01	26.46
% cost saving (n_{HF}/d for equivalent RRMSE)	>76.74 (>21.5)	>53.49 (>21.5)	>30.23 (>21.5)	>6.98 (>21.5)

*calculated as $(1 - A/B) * 100$

Table 3.8: Summary for the Rosenbrock test case ($d = 2$), where n_{HF} is the number of high-fidelity samples. Standard deviations of the RRMSE values are included in brackets.



(a) RMAE (b) RRMSE (c) R^2

FIGURE 3.17: Performance metrics of the Parametric (P), Relative (R) and Additive (A) corrective stages applied individually and in sequential combinations to the Rosenbrock numerical test case relative to the number of high-fidelity evaluations, where (A) is the RMAE (B) is the RRMSE and (C) is R^2. Included for reference are the error values of the low-fidelity model, and the single-fidelity surrogate trained only on the available high-fidelity samples at each iteration.

The Rastrigin Function A shifted-rotated Rastrigin function can represent a multi-modal landscape likely to be seen in real-world engineering optimisation problems (Mainini et al., 2022). The function is shifted to change the position of the minimum and rotated to change the properties of the function within the variable space. The function may be written as follows:

$$f_1(\mathbf{z}) = \sum_{i=1}^{D} (z_i^2 + 1 - \cos(10\pi z_i)) \tag{3.37}$$

where

$$\mathbf{z} = R(\theta)(\mathbf{x} - \mathbf{x}^*), \quad R(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{3.38}$$

where R is a 2D rotation matrix[7], rotation angle $\theta = 0.2$ and $-0.1 \leq x_i \leq 0.2$ for $i = 1, \ldots, D$. Alternative fidelity levels are defined in (Wang et al., 2018) using a resolution error e_r:

$$e_r(\mathbf{z}, \phi) = \sum_{i=1}^{D} a(\phi) \cos^2(w(\phi) z_i + b(\phi) + \pi) \tag{3.39}$$

[7]The included 2D rotation matrix can be extended to arbitrary dimension using the Aguilera-Perez algorithm (Aguilera and Pérez-Aguila, 2004).

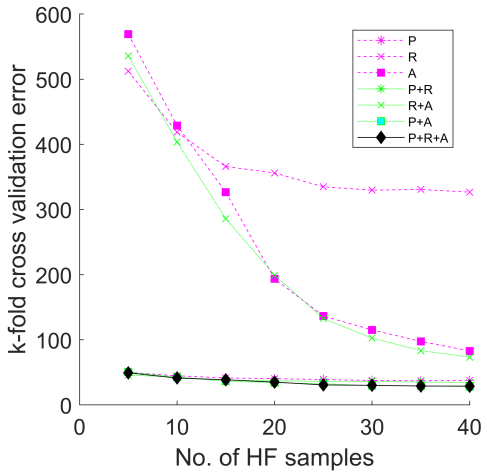


FIGURE 3.18: kCVEs for the Rosenbrock test case.

where

$$a(\phi) = \Theta(\phi) \tag{3.40}$$

$$w(\phi) = 10\pi\Theta(\phi) \tag{3.41}$$

$$b(\phi) = 0.5\pi\Theta(\phi) \tag{3.42}$$

$$\Theta(\phi) = 1 - 0.0001\phi \tag{3.43}$$

Fidelity levels may thus be created using:

$$f_i(\mathbf{z}, \phi_i) = f_1(\mathbf{z}) + e_r(\mathbf{z}, \phi_i) \text{ for } i = 1, 2, 3 \tag{3.44}$$

where $\phi_1 = 10000$ (high-fidelity), $\phi_2 = 5000$ (medium-fidelity) and $\phi_3 = 2500$ (low-fidelity).

As for the previous case, summary results are given in Table 3.9. Plots of RRMSE, RMAE and R^2 are given in Figure 3.20 and Figure 3.21 shows the kCVE for each method.

computational budget (n_{HF}/d)	5	10	15	20
Best MFM RRMSE	0.385 (0.042) (R)	0.363 (0.038) (PR)	0.219 (0.118) (PRA)	0.167 (0.087) (PA)
HFS RRMSE	1.252 (0.327)	0.382 (0.211)	0.204 (0.055)	0.162 (0.051)
% RRMSE reduction	69.217	4.952	-7.176	-3.125
% cost saving (n_{HF}/d for equivalent RRMSE)	52.38 (10.5)	4.76 (10.5)	1 (15)	-8.11 (18.5)

*calculated as $(1 - A/B) * 100$

Table 3.9: Summary for the Rastrigin (shifted and rotated) test case ($d = 2$), where n_{HF} is the number of high-fidelity samples. Standard deviations of the RRMSE values are included in brackets.

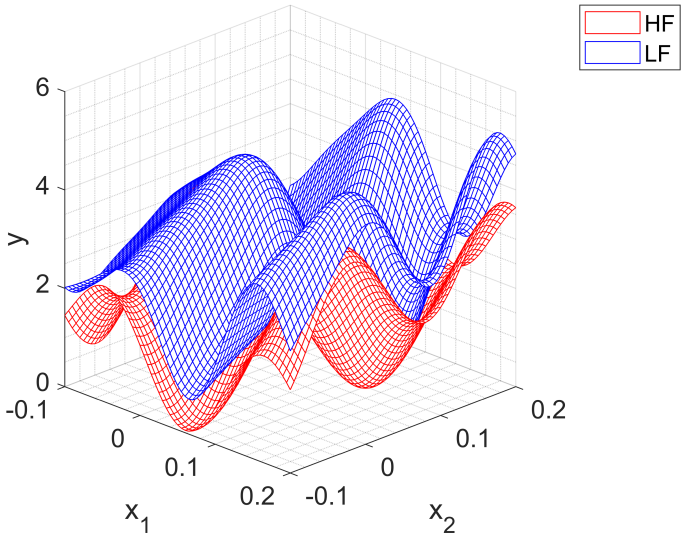
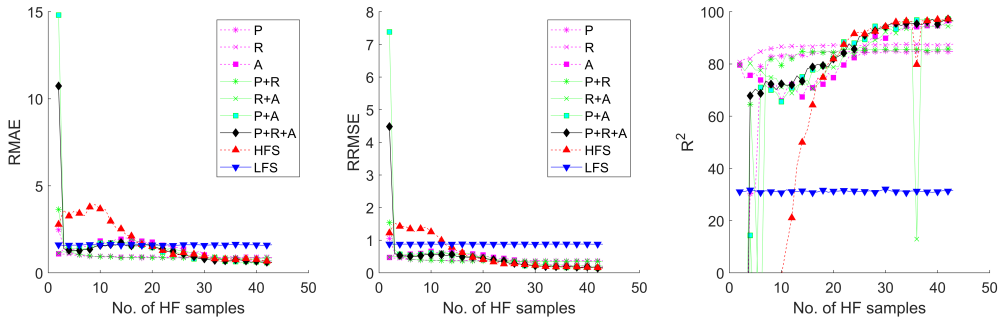


FIGURE 3.19: Rastrigin (shifted and rotated) test function, including high-fidelity response (red) and low-fidelity response shown (blue).



(a) RMAE (b) RRMSE (c) R^2

FIGURE 3.20: Performance metrics of the Parametric (P), Relative (R) and Additive (A) corrective stages applied individually and in sequential combinations to the Rastrigin numerical test case relative to the number of high-fidelity evaluations, where (A) is the RMAE (B) is the RRMSE and (C) is the R^2. Included for reference are the error values of the low-fidelity model, and the single-fidelity surrogate trained only on the available high-fidelity samples at each iteration.

The Heterogeneous Function The following heterogeneous, non-polynomial analytic functions are defined on unit hypercubes $0 \leq x_i \leq 1$for$i = 1, \ldots, D$, in up to 3 dimensions[8].

[8]the 2D case is taken from (Clark et al., 2016).

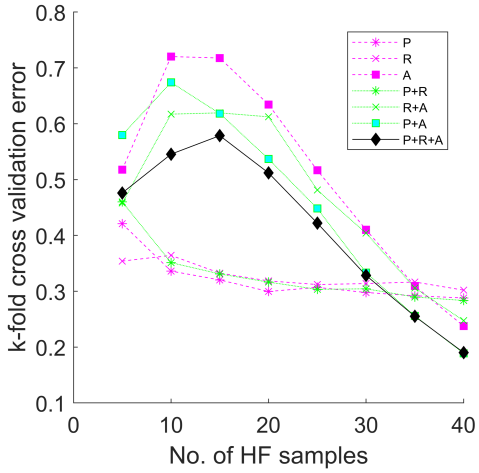


FIGURE 3.21: kCVEs for the Rastrigin test case.

The high-fidelity function is written as:

$$f_1 = \begin{cases} f_1(x) = \sin[30(x-0.9)^4]\cos[2(x-0.9)] + (x-0.9)/2 & \text{for } D = 1 \\ f_1(\mathbf{x}) = \sin[21(x_1-0.9)^4]\cos[2(x_1-0.9)] & \\ \quad +(x_1-0.7)/2 + \sum_{i=2}^{D} i x_i^i \sin(\Pi_{j=1}^{i} x_j) & \text{for } D \geq 2 \end{cases} \tag{3.45}$$

Low-fidelity versions are obtained by applying linear additive and multiplicative bridge functions:

$$f_2 = \begin{cases} f_2(x) = (f_1(x) - 1 + x)/(1 + 0.25x) & \text{for } D = 1 \\ f_2(\mathbf{x}) = (f_1(\mathbf{x}) - 2 + \sum_{i=1}^{D} x_i) & \\ /(5 + \sum_{i=1}^{2} 0.25 i x_i - \sum_{i=3, D>2}^{D} 0.25 i x_i) & \text{for } D \geq 2 \end{cases} \tag{3.46}$$

As for the previous case, summary results are given in Table 3.10. Plots of RRMSE, RMAE and R^2 are given in Figure 3.23 and Figure 3.24 shows the kCVE for each method.

The Paciorek Function The Paciorek equation (Toal, 2015) exhibits localised multi-modal properties. A normally distributed random noise parameter is added to model function noise. High- and low-fidelity versions are defined as:

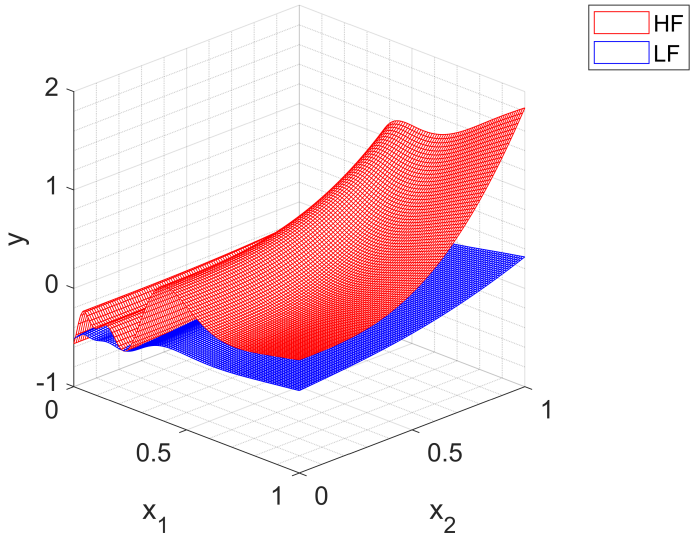


FIGURE 3.22: Heterogeneous test function, including high-fidelity response (red) and low-fidelity response shown (blue).

computational budget (n_{HF}/d)	5	10	15	20
Best MFM RRMSE	0.339 (0.038) (PR)	0.270 (0.098) (PA)	0.168 (0.047) (PA)	0.137 (0.047) (PRA)
HFS RRMSE	0.617 (0.272)	0.252 (0.120)	0.161 (0.060)	0.125 (0.032)
% RRMSE reduction	45.051	-7.303	-3.970	-10.150
% cost saving (n_{HF}/d for equivalent RRMSE)	37.50 (8)	-5.26 (9.5)	-3.45 (14.5)	-8.11 (18.5)

*calculated as $(1 - A/B) * 100$

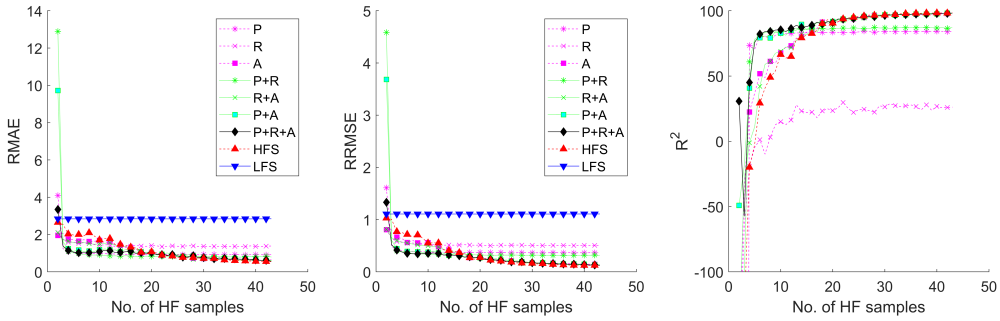
Table 3.10: Summary for Heterogeneous test case ($d = 2$), where n_{HF} is the number of high-fidelity samples. Standard deviations of the RRMSE values are included in brackets.

$$f_1(\mathbf{x}) = \sin\left[(\Pi_{i=1}^{D} x_i)^{-1}\right] + \mathcal{N}(0, \alpha_1) \tag{3.47}$$

$$f_2(\mathbf{x}) = f_1(\mathbf{x}) - 9A^2 \cos\left[(\Pi_{i=1}^{D} x_i)^{-1}\right] + \mathcal{N}(0, \alpha_2) \tag{3.48}$$

where the parameter $0 \leq A \leq 1$ represents the relative error between fidelities and $\mathcal{N}$ represents random Gaussian noise. In this case, $A = 0.5$. The coefficient α_i defines the level of noise, where the high-fidelity noise is set at $\alpha_1 = 0.0125$ (corresponding to approximately 5% of the response interval), and the low-fidelity noise is set at $\alpha_2 = 0.075$ (corresponding to approximately 10% of the response interval). The input is taken as $0.3 \leq x_i \leq 1, \forall i$.

As for the previous case, summary results are given in Table 3.11. Plots of RRMSE, RMAE and R^2 are given in Figure 3.26 and Figure 3.27 shows the kCVE for each method.



(a) RMAE (b) RRMSE (c) R^2

FIGURE 3.23: Performance metrics of the Parametric (P), Relative (R) and Additive (A) corrective stages applied individually and in sequential combinations to the Heterogeneous numerical test case relative to the number of high-fidelity evaluations, where (A) is the RMAE (B) is the RRMSE and (C) is the R^2. Included for reference are the error values of the low-fidelity model, and the single-fidelity surrogate trained only on the available high-fidelity samples at each iteration.

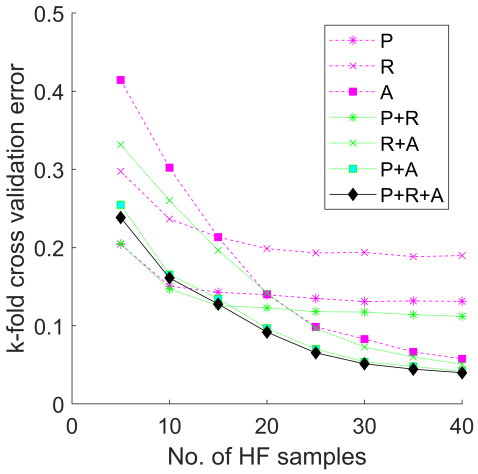


FIGURE 3.24: K-fold CV Errors for the Heterogeneous test case.

The Hartman Function The rescaled 6-dimensional formulation of the Hartman function (Hartman, 1973) used in (Park et al., 2018, 2017) is defined as:

$$f_H(\mathbf{x}) = -\frac{1}{1.94}\left[2.58 + \sum_{i=1}^{4} \alpha_i \exp\left(-\sum_{j=1}^{6} A_{ij}(x_j - P_{ij})^2\right)\right] \tag{3.49}$$

where $x_i \in [0.1, 1]$ for all $i = 1, \ldots, 6$ and:

$$\boldsymbol{\alpha} = \begin{bmatrix} 1 & 1.2 & 3 & 3.2 \end{bmatrix}^T$$

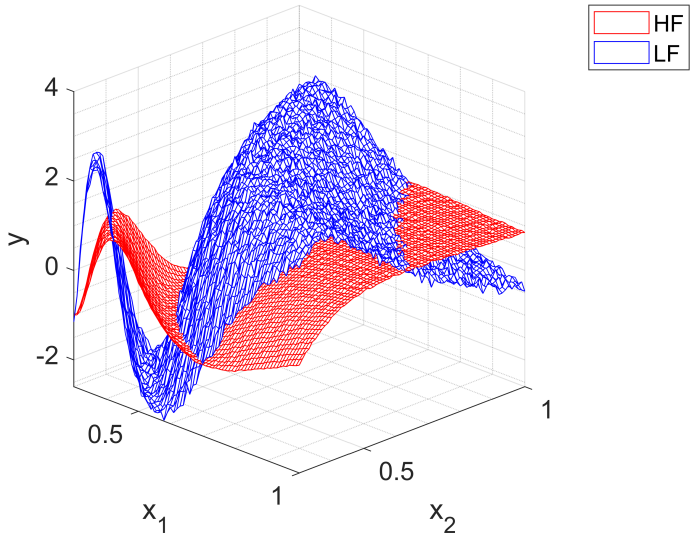


FIGURE 3.25: Paciorek test function, including high-fidelity response (red) and low-fidelity response shown (blue).

computational budget (n_{HF}/d)	5	10	15	20
Best MFM RRMSE	0.567 (0.215) (PA)	0.348 (0.188) (PRA)	0.301 (0.175) (PRA)	0.261 (0.162) (PA)
HFS RRMSE	0.601 (0.191)	0.373 (0.181)	0.263 (0.147)	0.228 (0.129)
% RRMSE reduction	5.58	6.70	-14.22	-14.42
% cost saving (n_{HF}/d for equivalent RRMSE)	9.09 (5.5)	4.76 (10.5)	-30.44 (11.5)	-21.21 (16.5)

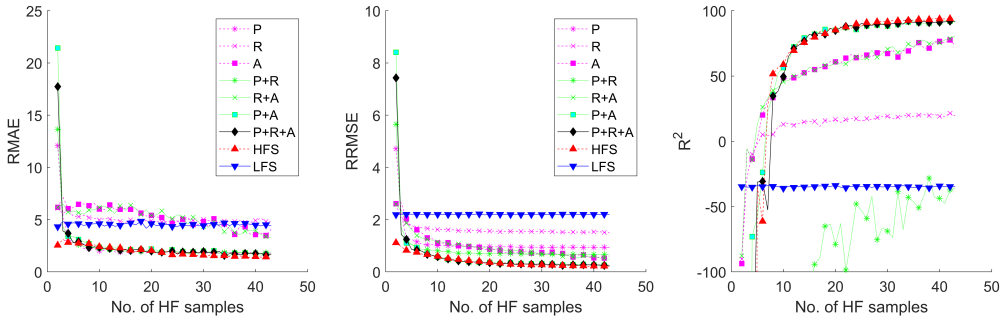
*calculated as $(1 - A/B) * 100$

Table 3.11: Summary for the Paciorek test case ($d = 2$), where n_{HF} is the number of high-fidelity samples. Standard deviations of the RRMSE values are included in brackets.

$$\mathbf{A} = \begin{bmatrix} 10 & 3 & 17 & 3.5 & 1.7 & 8 \\ 0.05 & 10 & 17 & 0.1 & 8 & 4 \\ 3 & 3.5 & 1.7 & 10 & 17 & 8 \\ 17 & 8 & 0.05 & 10 & 0.1 & 14 \end{bmatrix}$$

$$\mathbf{P} = 10^{-4} \begin{bmatrix} 1312 & 1696 & 5569 & 124 & 8283 & 5886 \\ 2329 & 4135 & 8307 & 3736 & 1004 & 9991 \\ 2348 & 1451 & 3522 & 2883 & 3047 & 6650 \\ 4047 & 8828 & 8732 & 5743 & 1091 & 0381 \end{bmatrix}$$

In this form, the Hartman function contains 6 local-minima, a mean value of zero and a variance of one. Park et al. (2018, 2017) present a low-fidelity approximation of the 6D Hartman function



(a) RMAE (b) RRMSE (c) R^2

FIGURE 3.26: Performance metrics of the Parametric (P), Relative (R) and Additive (A) corrective stages applied individually and in sequential combinations to the Pciorek numerical test case relative to the number of high-fidelity evaluations, where (A) is the RMAE (B) is the RRMSE and (C) is the R^2. Included for reference are the error values of the low-fidelity model, and the single-fidelity surrogate trained only on the available high-fidelity samples at each iteration.

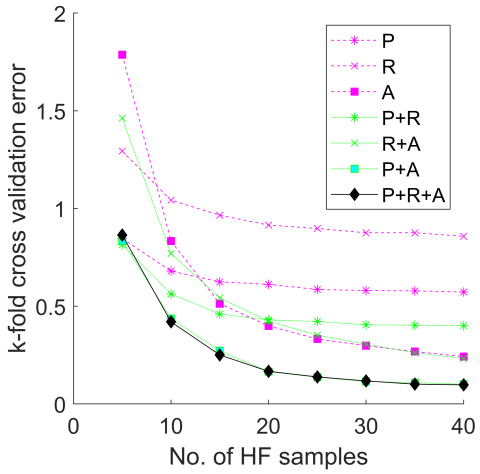


FIGURE 3.27: kCVEs for the Paciorek test case.

as:

$$f_L(\mathbf{x}) = -\frac{1}{1.94}\left[2.58 + \sum_{i=1}^{4} \beta_i f_{exp}\left(-\sum_{j=1}^{6} A_{ij}(x_j - P_{ij})^2\right)\right] \tag{3.50}$$

where $\boldsymbol{\beta} = \begin{bmatrix} 0.5 & 0.5 & 2 & 4 \end{bmatrix}^T$ and

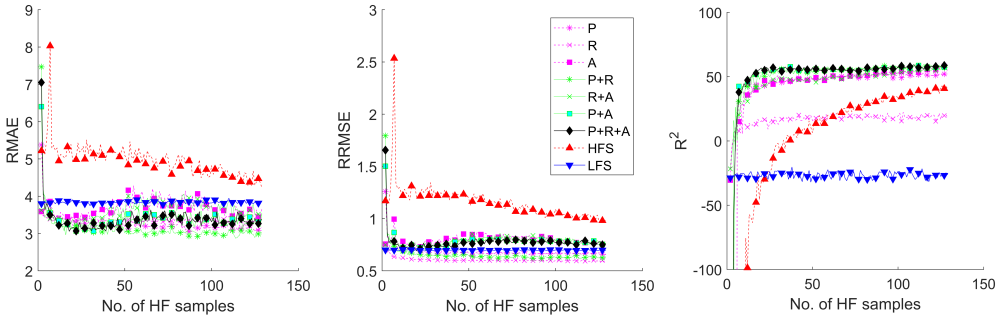
$$f_{exp}(x) = \left[\exp\left(\frac{-4}{9}\right) + \exp\left(\frac{-4}{9}\right)\frac{x+4}{9}\right]^9 \tag{3.51}$$

As for the previous case, summary results are given in Table 3.12. Plots of RRMSE, RMAE and R^2 are given in Figure 3.28 and Figure 3.29 shows the kCVE for each method.

computational budget (n_{HF}/d)	5	10	15	20
Best MFM RRMSE	0.613 (0.044) (R)	0.598 (0.047) (R)	0.597 (0.041) (R)	0.593 (0.039) (R)
HFS RRMSE	1.250 (0.355)	1.190 (0.259)	1.091 (0.247)	1.024 (0.224)
% RRMSE reduction	50.933	49.755	45.257	42.098
% cost saving (n_{HF}/d for equivalent RRMSE)	>76.74 (<21.5)	>53.49 (<21.5)	>30.23 (<21.5)	>6.98 (<21.5)

*calculated as $(1 - A/B) * 100$

Table 3.12: Summary for Hartman test case ($d = 6$), where n_{HF} is the number of high-fidelity samples. Standard deviations of the RRMSE values are included in brackets.



(a) RMAE (b) RRMSE (c) R^2

FIGURE 3.28: Performance metrics of the Parametric (P), Relative (R) and Additive (A) corrective stages applied individually and in sequential combinations to the Hartman numerical test case relative to the number of high-fidelity evaluations, where (A) is the RMAE (B) is the RRMSE and (C) is R^2. Included for reference are the error values of the low-fidelity model, and the single-fidelity surrogate trained only on the available high-fidelity samples at each iteration.

The Borehole model function The Borehole function is an 8-dimensional empirical approximation of the water flow rate through a borehole, developed by Morris et al. (1993). The Borehole function, in this case assumed to be the high-fidelity model, approximates the water flow rate (m^3/yr) according to:

$$y_H = \frac{2\pi T_u (H_u - H_l)}{\ln(R/R_w)\left(1 + \frac{2LT_u}{\ln(R/R_w)R_w^2 K_w} + \frac{T_u}{T_l}\right)} \tag{3.52}$$

where the input variables and bounds are as shown in Table 3.13. The low-fidelity approximation is obtained from the literature (Xiong et al., 2013; Zhou et al., 2016; Park et al., 2018):

$$y_L = \frac{5T_u (H_u - H_l)}{\ln(R/R_w)\left(1.5 + \frac{2LT_u}{\ln(R/R_w)R_w^2 K_w} + \frac{T_u}{T_l}\right)} \tag{3.53}$$

As for the previous case, summary results are given in Table 3.14. Plots of RRMSE, RMAE and R^2 are given in Figure 3.30 and Figure 3.31 shows the kCVE for each method.

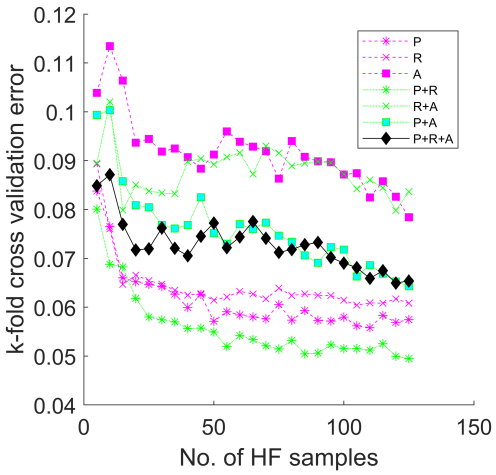


FIGURE 3.29: kCVEs for the Hartman test case.

Input	Bounds	Unit	Description
R_w	$[0.05, 0.15]$	m	Borehole radius
L	$[1120, 1680]$	m	Borehole length
K_w	$[9855, 12045]$	m/yr	Hydraulic conductivity of borehole
R	$[100, 50000]$	m	Radius of aquifer influence
T_u	$[63070, 115600]$	m^2/yr	Transmissivity of upper aquifer
H_u	$[990, 1110]$	m	Potentiometric head of upper aquifer
T_l	$[63.1, 116]$	m^2/yr	Transmissivity of lower aquifer
H_l	$[700, 820]$	m	Potentiometric head of lower aquifer

Table 3.13: Input variables for Borehole function.

computational budget (n_{HF}/d)	5	10	15	20
Best MFM RRMSE	0.070 (0.011) (PA)	0.063 (0.010) (PRA)	0.059 (0.010) (PRA)	0.058 (0.008) (PRA)
HFS RRMSE	0.093 (0.022)	0.069 (0.010)	0.065 (0.012)	0.064 (0.011)
% RRMSE reduction	25.145	8.301	8.798	8.447
% cost saving (n_{HF}/d for equivalent RRMSE)	50.62 (10.125)	38.93 (16.375)	>30.23 (>21.5)	>6.98 (>21.5)

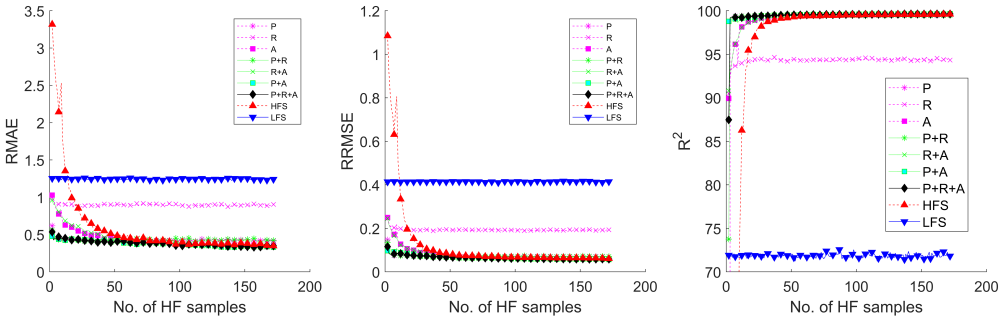
*calculated as $(1 - A/B) * 100$

Table 3.14: Summary for Borehole test case ($d = 8$), where n_{HF} is the number of high-fidelity samples. Standard deviations of the RRMSE values are included in brackets.

3.1.4.5 Summary

The obtained RRMSE results for the complete suite of benchmark test cases are summarised in Table 3.15. Similar tables displaying the corresponding RMAE, R^2, r^2 and surrogate generation times can be found in the Appendices.

As reported in Tables 3.4 - 3.14, computational cost savings over the single-fidelity modelling



(a) RMAE (b) RRMSE (c) R^2

FIGURE 3.30: Performance metrics of the Parametric (P), Relative (R) and Additive (A) corrective stages applied individually and in sequential combinations to the Borehole numerical test case relative to the number of high-fidelity evaluations, where (A) is the RMAE (B) is the RRMSE and (C) is R^2. Included for reference are the error values of the low-fidelity model, and the single-fidelity surrogate trained only on the available high-fidelity samples at each iteration.

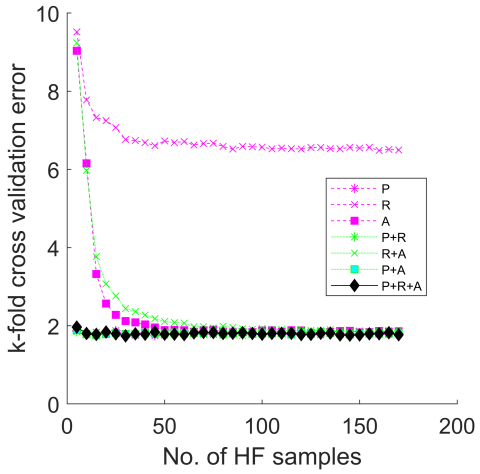


FIGURE 3.31: kCVEs for the Borehole test case.

approach are demonstrated in all cases by at least one multi-fidelity process at some high-fidelity sample size N. In 6/10 cases, the reported cost savings include estimates due to the single-fidelity approach reaching the maximum allocated budget of high-fidelity samples without having attained an equal or superior RRMSE value compared to the best multi-fidelity process. The performance of individual processes is generally seen to be dependant on dimensionality, sample size N and low/high-fidelity model correlation.

Considering model correlation, recall Table 3.18 (Section 3.1.4), showing the Pearson's correlation and R^2 values between the high- and low-fidelity test functions. A general trend for the prediction quality of the multi-fidelity surrogate to improve proportionally to Pearson's correlation coefficient can be observed. An interesting clarification is that this is not necessarily

Test functions	d	cost*	model									
			P	R	A	PR	RA	PA	PRA	HFS	LFS	
Forrester	1	5	0.789 (0.144)	0.472 (0.064)	0.341 (0.125)	0.712 (0.349)	**0.323** (0.126)	0.489 (0.328)	0.499 (0.4)	0.524 (0.308)	1.357 (0.027)	
		10	0.71 (0.025)	0.437 (0.024)	0.109 (0.026)	0.521 (0.075)	**0.104** (0.025)	0.109 (0.031)	**0.104** (0.027)	0.113 (0.03)	1.358 (0.027)	
		15	0.703 (0.017)	0.433 (0.015)	0.091 (0.016)	0.513 (0.036)	**0.082** (0.014)	0.092 (0.018)	0.084 (0.016)	0.087 (0.025)	1.359 (0.028)	
		20	0.7 (0.019)	0.429 (0.015)	0.088 (0.016)	0.509 (0.028)	0.082 (0.014)	0.085 (0.012)	**0.08** (0.01)	**0.08** (0.02)	1.357 (0.03)	
Sasena	1	5	0.607 (0.039)	0.249 (0.056)	0.086 (0.038)	0.372 (0.088)	**0.083** (0.035)	0.155 (0.068)	0.168 (0.071)	0.586 (0.136)	1.175 (0.019)	
		10	0.581 (0.015)	0.245 (0.026)	0.058 (0.018)	0.307 (0.036)	**0.052** (0.016)	0.061 (0.018)	0.06 (0.017)	0.085 (0.048)	1.172 (0.019)	
		15	0.578 (0.016)	0.247 (0.017)	0.053 (0.016)	0.293 (0.025)	**0.052** (0.013)	0.059 (0.016)	0.055 (0.016)	0.063 (0.026)	1.171 (0.02)	
		20	0.576 (0.017)	0.248 (0.016)	0.056 (0.017)	0.293 (0.02)	**0.052** (0.015)	0.056 (0.016)	**0.052** (0.017)	0.057 (0.021)	1.175 (0.02)	
Currin	2	5	0.209 (0.037)	1.991 (0.229)	1.973 (0.835)	**0.188** (0.042)	1.737 (0.62)	0.2 (0.051)	0.203 (0.056)	0.607 (0.204)	4.255 (0.111)	
		10	0.195 (0.031)	1.952 (0.209)	0.784 (0.581)	0.18 (0.033)	0.723 (0.562)	**0.14** (0.057)	0.145 (0.054)	0.216 (0.147)	4.264 (0.098)	
		15	0.192 (0.029)	1.949 (0.205)	0.509 (0.373)	0.177 (0.029)	0.373 (0.234)	0.117 (0.051)	**0.115** (0.058)	0.118 (0.043)	4.253 (0.097)	
		20	0.202 (0.039)	1.974 (0.179)	0.354 (0.238)	0.167 (0.03)	0.306 (0.222)	0.096 (0.028)	**0.091** (0.029)	0.108 (0.045)	4.254 (0.099)	
Heterogeneous	2	5	0.376 (0.037)	0.524 (0.053)	0.56 (0.199)	**0.339** (0.038)	0.5 (0.131)	0.378 (0.125)	0.354 (0.081)	0.587 (0.265)	1.108 (0.024)	
		10	0.371 (0.033)	0.514 (0.035)	0.271 (0.162)	0.323 (0.03)	0.281 (0.131)	**0.27** (0.098)	0.277 (0.138)	0.292 (0.181)	1.108 (0.023)	
		15	0.365 (0.032)	0.507 (0.031)	0.173 (0.048)	0.318 (0.028)	0.169 (0.049)	0.168 (0.047)	0.174 (0.069)	**0.157** (0.061)	1.108 (0.023)	
		20	0.364 (0.03)	0.506 (0.028)	0.147 (0.034)	0.319 (0.027)	0.141 (0.04)	0.145 (0.049)	0.137 (0.047)	**0.125** (0.032)	1.109 (0.023)	
Rastrigin(shifted)	2	5	0.388 (0.048)	**0.385** (0.042)	0.66 (0.16)	0.39 (0.062)	0.597 (0.155)	0.612 (0.157)	0.553 (0.139)	1.219 (0.42)	0.882 (0.064)	
		10	0.368 (0.037)	0.376 (0.041)	0.55 (0.181)	**0.363** (0.038)	0.504 (0.177)	0.439 (0.17)	0.444 (0.152)	0.426 (0.235)	0.88 (0.062)	
		15	0.36 (0.033)	0.376 (0.047)	0.306 (0.171)	0.354 (0.03)	0.264 (0.111)	0.226 (0.103)	0.219 (0.118)	**0.204** (0.055)	0.881 (0.065)	
		20	0.361 (0.035)	0.377 (0.042)	0.207 (0.127)	0.354 (0.034)	0.219 (0.111)	0.167 (0.087)	0.186 (0.107)	**0.161** (0.053)	0.881 (0.065)	
Rosenbrock	2	5	0.141 (0.035)	0.757 (0.067)	0.626 (0.177)	**0.13** (0.036)	0.6 (0.172)	0.137 (0.032)	0.141 (0.037)	0.613 (0.241)	1.123 (0.027)	
		10	0.146 (0.047)	0.729 (0.049)	0.285 (0.111)	0.132 (0.035)	0.268 (0.11)	**0.12** (0.033)	0.121 (0.037)	0.287 (0.171)	1.123 (0.027)	
		15	0.128 (0.031)	0.726 (0.046)	0.2 (0.072)	0.123 (0.035)	0.176 (0.064)	0.105 (0.035)	**0.104** (0.029)	0.179 (0.079)	1.122 (0.028)	
		20	0.144 (0.039)	0.714 (0.041)	0.162 (0.06)	0.124 (0.031)	0.142 (0.047)	0.099 (0.04)	**0.097** (0.025)	0.131 (0.045)	1.123 (0.028)	
Paciorek	2	5	1.001 (0.129)	1.619 (0.193)	1.109 (0.402)	0.808 (0.191)	1.15 (0.409)	**0.567** (0.215)	0.578 (0.225)	0.58 (0.189)	2.192 (0.18)	
		10	0.963 (0.057)	1.562 (0.198)	0.867 (0.407)	0.693 (0.099)	0.88 (0.496)	0.364 (0.184)	**0.348** (0.188)	0.373 (0.181)	2.204 (0.196)	
		15	0.947 (0.058)	1.553 (0.191)	0.733 (0.455)	0.68 (0.072)	0.671 (0.461)	0.315 (0.174)	0.301 (0.175)	**0.263** (0.147)	2.199 (0.194)	
		20	0.941 (0.038)	1.533 (0.174)	0.538 (0.361)	0.67 (0.064)	0.598 (0.477)	0.261 (0.162)	0.266 (0.169)	**0.239** (0.158)	2.192 (0.181)	
Park2	4	5	0.082 (0.016)	0.133 (0.022)	0.075 (0.013)	0.082 (0.015)	**0.074** (0.012)	0.078 (0.013)	0.077 (0.013)	0.125 (0.038)	0.692 (0.039)	
		10	0.083 (0.015)	0.129 (0.019)	**0.069** (0.013)	0.081 (0.016)	0.07 (0.013)	**0.069** (0.01)	0.072 (0.014)	0.089 (0.022)	0.694 (0.041)	
		15	0.081 (0.014)	0.129 (0.018)	0.068 (0.012)	0.078 (0.015)	0.069 (0.012)	0.066 (0.011)	**0.065** (0.01)	0.075 (0.015)	0.688 (0.042)	
		20	0.083 (0.016)	0.127 (0.018)	0.066 (0.01)	0.08 (0.012)	**0.064** (0.011)	0.066 (0.011)	0.066 (0.011)	0.073 (0.013)	0.691 (0.039)	
Hartman	6	5	0.688 (0.11)	**0.613** (0.044)	0.762 (0.207)	0.645 (0.079)	0.712 (0.157)	0.73 (0.127)	0.748 (0.186)	1.303 (0.434)	0.705 (0.04)	
		10	0.664 (0.088)	**0.598** (0.047)	0.857 (0.249)	0.651 (0.109)	0.817 (0.211)	0.804 (0.181)	0.794 (0.161)	1.216 (0.304)	0.698 (0.038)	
		15	0.665 (0.07)	**0.597** (0.041)	0.786 (0.16)	0.634 (0.069)	0.82 (0.156)	0.757 (0.109)	0.776 (0.138)	1.046 (0.217)	0.703 (0.042)	
		20	0.667 (0.078)	**0.593** (0.039)	0.755 (0.134)	0.639 (0.086)	0.755 (0.131)	0.762 (0.119)	0.766 (0.123)	0.973 (0.166)	0.696 (0.043)	
Borehole	8	5	0.072 (0.012)	0.193 (0.014)	0.074 (0.013)	0.073 (0.013)	0.081 (0.014)	**0.07** (0.011)	**0.07** (0.011)	0.09 (0.021)	0.415 (0.014)	
		10	0.071 (0.012)	0.196 (0.017)	0.065 (0.011)	0.069 (0.011)	0.066 (0.013)	0.064 (0.01)	**0.063** (0.01)	0.069 (0.01)	0.414 (0.015)	
		15	0.072 (0.013)	0.191 (0.016)	0.062 (0.012)	0.07 (0.012)	0.061 (0.01)	**0.059** (0.008)	**0.059** (0.01)	0.065 (0.012)	0.415 (0.012)	
		20	0.072 (0.012)	0.191 (0.015)	0.062 (0.01)	0.069 (0.012)	0.06 (0.01)	0.06 (0.009)	**0.058** (0.008)	0.064 (0.011)	0.416 (0.014)	

Table 3.15: RRMSE values for each multi/single-fidelity method applied to each benchmark function. Reported are the values at distinct levels of computational cost relative to problem dimensionality. All values are averaged over 100 independent runs, with standard deviation included in brackets.

Test functions	d	cost*	model									
			P	R	A	PR	RA	PA	PRA	HFS	LFS	
Forrester	1	5	1.932 (0.796)	1.173 (0.238)	0.918 (0.408)	2.134 (1.485)	**0.878** (0.408)	1.479 (1.268)	1.563 (1.517)	1.604 (1.147)	2.57 (0.09)	
		10	1.56 (0.213)	1.041 (0.12)	0.398 (0.177)	1.386 (0.399)	0.394 (0.169)	0.381 (0.18)	**0.373** (0.169)	0.432 (0.17)	2.575 (0.091)	
		15	1.486 (0.095)	1.024 (0.078)	0.294 (0.122)	1.367 (0.193)	**0.257** (0.107)	0.314 (0.138)	0.276 (0.125)	0.318 (0.166)	2.574 (0.097)	
		20	1.448 (0.053)	0.997 (0.074)	0.287 (0.12)	1.37 (0.172)	0.273 (0.107)	0.27 (0.106)	**0.255** (0.091)	0.276 (0.134)	2.569 (0.098)	
Sasena	1	5	1.074 (0.251)	0.565 (0.217)	0.256 (0.143)	0.795 (0.26)	**0.251** (0.137)	0.462 (0.23)	0.505 (0.216)	1.56 (0.467)	2.521 (0.053)	
		10	1.013 (0.136)	0.599 (0.091)	0.173 (0.08)	0.577 (0.085)	**0.148** (0.067)	0.181 (0.089)	0.198 (0.093)	0.306 (0.232)	2.505 (0.066)	
		15	1.006 (0.097)	0.624 (0.068)	**0.155** (0.077)	0.545 (0.06)	**0.155** (0.065)	0.175 (0.079)	0.179 (0.09)	0.215 (0.142)	2.504 (0.065)	
		20	0.982 (0.052)	0.605 (0.063)	0.162 (0.079)	0.548 (0.046)	**0.145** (0.066)	0.16 (0.075)	0.161 (0.079)	0.193 (0.113)	2.501 (0.067)	
Currin	2	5	0.79 (0.235)	7.189 (1.135)	8.446 (3.839)	**0.756** (0.251)	8.217 (3.449)	0.853 (0.276)	0.842 (0.265)	2.633 (0.888)	11.651 (0.941)	
		10	0.742 (0.214)	6.803 (1.219)	3.86 (3.287)	0.728 (0.211)	3.496 (3.278)	**0.627** (0.317)	0.65 (0.292)	1.037 (0.754)	11.785 (0.911)	
		15	0.738 (0.189)	6.755 (1.252)	2.398 (2.246)	0.717 (0.189)	1.912 (1.497)	0.58 (0.392)	**0.554** (0.335)	0.581 (0.309)	11.711 (0.92)	
		20	0.81 (0.278)	6.839 (1.079)	1.666 (1.566)	0.678 (0.193)	1.579 (1.52)	0.469 (0.219)	**0.428** (0.206)	0.543 (0.379)	11.774 (0.904)	
Heterogeneous	2	5	0.974 (0.155)	1.427 (0.312)	1.687 (0.615)	**0.904** (0.176)	1.503 (0.425)	1.183 (0.443)	1.087 (0.34)	1.83 (0.903)	2.839 (0.213)	
		10	0.945 (0.148)	1.415 (0.218)	1.013 (0.659)	**0.831** (0.16)	1.081 (0.484)	1.004 (0.372)	1.037 (0.444)	1.152 (0.687)	2.839 (0.212)	
		15	0.935 (0.134)	1.369 (0.212)	0.712 (0.239)	0.826 (0.145)	0.697 (0.26)	0.736 (0.251)	0.764 (0.327)	**0.681** (0.265)	2.84 (0.211)	
		20	0.923 (0.142)	1.367 (0.225)	0.634 (0.22)	0.816 (0.136)	0.645 (0.263)	0.672 (0.254)	0.647 (0.294)	**0.582** (0.208)	2.845 (0.212)	
Rastrigin(shifted)	2	5	0.94 (0.205)	0.974 (0.226)	1.829 (0.531)	**0.933** (0.164)	1.69 (0.503)	1.709 (0.513)	1.581 (0.476)	3.619 (1.456)	1.61 (0.242)	
		10	0.874 (0.134)	0.927 (0.235)	1.763 (0.651)	**0.867** (0.12)	1.627 (0.615)	1.458 (0.684)	1.536 (0.609)	1.507 (0.811)	1.607 (0.19)	
		15	0.855 (0.139)	0.925 (0.236)	1.028 (0.622)	0.855 (0.135)	0.884 (0.39)	0.862 (0.434)	**0.809** (0.416)	0.867 (0.341)	1.592 (0.184)	
		20	0.876 (0.156)	0.922 (0.238)	0.73 (0.407)	0.852 (0.14)	0.74 (0.31)	**0.638** (0.317)	0.666 (0.384)	0.7 (0.303)	1.59 (0.185)	
Rosenbrock	2	5	0.653 (0.302)	3.035 (0.575)	2.478 (0.836)	**0.578** (0.269)	2.424 (0.775)	0.592 (0.255)	0.615 (0.281)	2.515 (0.989)	4.142 (0.408)	
		10	0.667 (0.364)	2.958 (0.493)	1.361 (0.701)	0.578 (0.253)	1.33 (0.649)	**0.555** (0.279)	**0.555** (0.296)	1.377 (0.971)	4.145 (0.41)	
		15	0.555 (0.246)	2.984 (0.462)	0.994 (0.488)	0.541 (0.285)	0.953 (0.549)	0.542 (0.318)	**0.487** (0.251)	0.925 (0.552)	4.134 (0.408)	
		20	0.667 (0.323)	2.874 (0.387)	0.796 (0.449)	0.568 (0.275)	0.727 (0.423)	0.49 (0.353)	**0.473** (0.229)	0.688 (0.38)	4.138 (0.412)	
Paciorek	2	5	2.008 (0.602)	4.968 (1.547)	6.014 (3.378)	**2.214** (0.645)	6.179 (3.231)	2.358 (0.95)	2.35 (0.967)	2.378 (0.86)	4.252 (1.387)	
		10	1.948 (0.534)	4.888 (1.486)	5.413 (3.389)	**1.945** (0.425)	5.407 (3.824)	2.18 (1.437)	1.988 (1.283)	1.973 (1.138)	4.602 (1.804)	
		15	1.828 (0.504)	5.018 (1.502)	4.818 (3.695)	1.895 (0.388)	4.317 (3.772)	1.985 (1.448)	1.942 (1.571)	**1.618** (1.178)	4.49 (1.867)	
		20	1.818 (0.4)	4.915 (1.517)	3.584 (3.289)	1.846 (0.431)	3.935 (3.612)	1.693 (1.483)	1.75 (1.439)	**1.523** (1.211)	4.355 (1.62)	
Park2	4	5	0.385 (0.139)	0.378 (0.083)	0.344 (0.123)	0.395 (0.124)	**0.325** (0.115)	0.364 (0.134)	0.363 (0.122)	0.558 (0.247)	0.955 (0.062)	
		10	0.393 (0.133)	0.356 (0.08)	0.33 (0.136)	0.394 (0.147)	0.332 (0.129)	**0.328** (0.107)	0.358 (0.135)	0.432 (0.144)	0.972 (0.068)	
		15	0.391 (0.127)	0.362 (0.074)	0.332 (0.126)	0.378 (0.13)	0.327 (0.12)	0.308 (0.098)	**0.293** (0.104)	0.356 (0.148)	0.962 (0.06)	
		20	0.394 (0.138)	0.36 (0.072)	0.3 (0.101)	0.387 (0.134)	**0.297** (0.103)	0.307 (0.12)	0.314 (0.102)	0.346 (0.124)	0.956 (0.059)	
Hartman	6	5	3.251 (0.675)	3.421 (0.72)	3.404 (1.183)	**3.054** (0.652)	3.222 (0.838)	3.085 (0.642)	3.137 (1.046)	5.358 (1.835)	3.856 (0.711)	
		10	3.204 (0.697)	3.318 (0.684)	4.129 (1.966)	**3.032** (0.813)	3.809 (1.756)	3.458 (1.113)	3.396 (1.009)	5.12 (1.726)	3.808 (0.636)	
		15	3.149 (0.763)	3.352 (0.629)	3.597 (1.363)	**3.045** (0.664)	3.835 (1.239)	3.174 (0.685)	3.334 (0.991)	4.539 (1.472)	3.89 (0.686)	
		20	3.145 (0.658)	3.332 (0.66)	3.484 (1.065)	**3.093** (0.727)	3.489 (1.036)	3.279 (0.927)	3.308 (1.039)	4.23 (1.343)	3.842 (0.658)	
Borehole	8	5	0.415 (0.175)	0.906 (0.202)	0.437 (0.168)	0.44 (0.195)	0.502 (0.19)	**0.396** (0.164)	0.413 (0.149)	0.547 (0.27)	1.246 (0.177)	
		10	0.394 (0.158)	0.906 (0.205)	0.393 (0.158)	0.397 (0.149)	0.394 (0.183)	0.377 (0.15)	**0.373** (0.142)	0.403 (0.174)	1.239 (0.179)	
		15	0.447 (0.186)	0.913 (0.234)	0.372 (0.161)	0.409 (0.152)	0.359 (0.144)	0.345 (0.124)	**0.339** (0.14)	0.373 (0.172)	1.234 (0.166)	
		20	0.431 (0.176)	0.891 (0.22)	0.353 (0.147)	0.406 (0.161)	0.343 (0.132)	0.344 (0.131)	**0.336** (0.13)	0.378 (0.158)	1.258 (0.206)	

Table 3.16: RMAE values for each multi/single-fidelity method applied to each benchmark function. Reported are the values at distinct levels of computational cost relative to problem dimensionality. All values are averaged over 100 independent runs, with standard deviation included in brackets.

dependant on the original prediction quality of the low-fidelity approximation. A notable example here is the Currin exponential function, exhibiting poor predictive performance of the low-fidelity model, but owing to the high Pearson's correlation, an R^2 value greater than 90% can be achieved with less than 4 high-fidelity samples, compared to the 16 required for the single-fidelity approach. Perhaps intuitively then, it can be said that a strong geometric correlation between low- and high-fidelity responses gives a greater potential for multi-fidelity model correction. In a similar fashion, both the higher-dimensional Park and Borehole model functions, exhibiting sub-optimal R^2 values but high Pearson's correlation for their respective low-fidelity approximations, facilitate a multi-fidelity prediction quality of $R^2 > 90\%$ with very few (3) high-fidelity samples.

To examine the performance of the multi-fidelity correction methods themselves in more detail, first consider each individual stage applied in isolation. The Parametric (P) stage, based on a simple linear scaling relationship (see Section 3.1.2.1), is seen to be particularly effective when the Pearson's correlation between the low- and high-fidelity data is high (refer to Table 3.3). That is, the relative functional form of the low-fidelity approximation more closely matches that of the high-fidelity data, even if displaced in output space (i.e. giving a large and/or nonlinear absolute error). This includes the Currin, Rosenbrock, Borehole and Park functions (all of which demonstrate a low-fidelity/high-fidelity $r^2 > 90\%$). However, the rather simplistic nature of the defining parametric operation limits the effectiveness of this stage in cases exhibiting a more complicated/nonlinear discrepancy. For example, the Forrester, Sasena and Paciorek functions. Whilst in these cases the Parametric stage may still initially offer a promising reduction in error at lower values of N, the effectiveness of successive samples exponentially decreases.

A similar limitation is seen for the Relative (R) stage, itself fundamentally constrained by the complexity of the low-fidelity model (see Section 3.1.2.2). More specifically, given that the low-fidelity model by definition can only approximate the true function, even if represented in sufficient resolution it can be expected that the R-stage will exhibit a comparable limit in quality. This behaviour (i.e. an initially rapid increase in correction quality but with swift plateauing as N increases) is observed across all of the examined test cases.

Addressing the plateauing effects of the Parametric and Relative methods requires a more detailed consideration of sample-point discrepancy. The Additive (A) stage represents the only individual operation that expressly aims to fully eliminate sample point discrepancy in a non-destructive manner. That is, the discrepancy between observed samples is explicitly built into the nonlinear bridging surrogate that is then applied across the entire domain. It is reasonable to assume then that correction quality will increase proportional to the high-fidelity sample size N, which indeed matches the observed behaviour for the majority of the examined test cases. However, the converse is naturally seen to be true at lower values of N,

with particularly poor performance in the Currin and Rosenbrock cases. Strong performance can be seen for the lower dimensional test cases, namely the Forrester and Sasena functions, highlighting the efficiency of ANNs for nonlinear regression problems (see Section 2.1.3).

Applying stages in sequential combination (e.g. PR, RA, and PA) seems to alleviate some of the limitations of single-stage processes, namely instability and early convergence. For example, in the Borehole test case, preconditioning the low-fidelity response with the P-stage both improves the converged error values of the subsequently applied R-stage and similarly, the correction quality/stability of the A-stage. In the Heterogeneous, Paciorek and Rosenbrock test cases, the PA method exhibits the initial rapid improvement in prediction quality expected of the P-stage but without the corresponding plateauing. Instead, the PA method steadily reduces the error in accordance with the increasing sample size N, as the additive discrepancy is gradually represented in higher detail.

In the majority of 2-dimensional cases (Currin, Heterogeneous, Rosenbrock and Paciorek), each two-stage method performs better than their constituent single-stage methods individually. However, in a select few examples, the addition of a preconditioning stage actually reduces the overall performance. For instance, PR performs worse than R for the Forrester, Sasena and Rastrigin test cases. Furthermore, several examples are observed of the RA stage exhibiting almost identical performance to the A stage on it's own, particularly for the lower dimensional/well-correlated cases (e.g. Forrester, Sasena, Currin, Rosenbrock, Paciorek and Borehole). This can be attributed to the R-stage primarily aiming to increase the functional correlation, rather than decrease discrepancy. Thus when the correlation is already high, only marginal improvement is gained. This is supported by noting that in these cases, the PA stage performs notably better than the RA stage.

With this in mind, the most consistently reliable sequence is found to include all 3 stages (PRA). For most cases, the PRA process exhibits both high prediction accuracy and fast convergence/rapid error reduction with increasing N. As observed for the two-stage approaches, the three-stage method is seen to exhibit the best features of each individual stage, compensating for deficiencies present when in isolated application. A good example of this is the Heterogeneous test case, in which the PRA method demonstrates the quick initial improvement in prediction quality of the P/R-stages without the corresponding plateauing effects, as well as the progressive reduction in error of the A-stage, without the initial instability.

The single-fidelity approach generally offers a poorer quality, less stable prediction, particularly at lower values of N. This is in line with expectation given the limited available training data. However, conforming to the conjecture that a suitably trained ANN may represent any nonlinear function given sufficient training data (Rumelhart et al., 1994) (see Section 2.1.3), it can be seen that for most (8/10) examples there exists a value of N after which the single-

fidelity surrogate is either roughly equivalent or superior to the multi-fidelity methods in terms of average accuracy. This behaviour is again in line with expectation. The multi-fidelity model, by definition, always incorporates some less accurate, lower-fidelity data, therefore ultimately cannot reach the same level of accuracy as a pure high-fidelity surrogate assuming the latter has sufficient samples and that low-fidelity data is not being actively discarded[9]. The exact number of required samples is likely to be generally dependant on factors such as problem dimensionality, model correlation and the sampling strategy employed. Concerning the latter, an adaptive sampling strategy that refines the multi-fidelity approximation by exploiting areas of high predicted improvement and/or maximum discrepancy to improve local accuracy may further increase the attainable prediction quality relative to the overall number of samples. This naturally also implies potential improvement in the single-fidelity modelling approach, though it may still be conjectured that the decline in cost advantage relative to pure high-fidelity surrogate modelling may still be tempered. This requires further investigation outwith the scope of the current work.

However, given that the objective with multi-fidelity modelling is largely the reduction of cost to achieve a desired accuracy, the existence of a critical number of high-fidelity samples N_{crit} does not devalue the performance of multi-fidelity methods at $N < N_{crit}$. For example, recall from Section 2.1.3 that trained neural networks are highly sensitive to the initial weights, biases and available training data (generally low-quantity for pure high-fidelity modelling). Thus it may be expected that in addition to an exponentially poorer prediction mean at lower values of N, similar behaviour will be observed for prediction variance. Indeed, increased variance is observed for each method utilising a neural network (i.e. non-deterministic) stage. Whilst the collaborative effects of each method on prediction quality variance seem to follow the same general behaviour as previously discussed, of particular interest is that even after N_{crit} has been reached, there still exists notable variation in the prediction quality of the single-fidelity surrogate, likely due to the combination of limited available samples and a lack of any deterministic preconditioning stage(s). In many cases, the variance of the best multi-fidelity process remains superior to that of the single-fidelity approach until well past N_{crit}. This suggests that incorporating low-fidelity data in this manner also improves the *stability* of the resulting prediction relative to single-fidelity modelling, particularly at lower values of N. This may have far-reaching consequences when considering the effects of this behaviour on an outer-loop process such as optimisation, in which the final obtained result may be particularly dependant on the design space exploration completed during earlier iterations. Thus if a multi-fidelity prediction is both more accurate *and* more stable between iterations, the overall outer-loop process may exhibit compounded improvements in terms of solution quality and convergence speed.

[9] Which indeed is the premise behind methods such as Trust Region Model Management (Alexandrov et al., 2001), see Section 2.4.3.1.

3.2 Multi-Fidelity Model Management Framework for Outer-Loop Applications

An effective model management strategy must efficiently distribute available computational resource (in the form of model evaluations) between the available levels of fidelity whilst preserving the numerical accuracy and/or convergence of the outer-loop quantity of interest (e.g. model quality or objective function value) that would be obtained if only the high-fidelity model was used (Peherstorfer et al., 2018). Recall from Section 2.4, a complete *Multi-Fidelity Model Management* (MFMM) framework for outer-loop applications consists of an amalgamation of several constituent processes:

1. Initial sampling of the associated design space(s).
2. Generation of initial surrogates using some appropriate modelling technique(s).
3. Optimisation using the constructed surrogates.
4. New points at which to evaluate data are identified according to some specified infill criteria.
5. Surrogates are updated with new information.
6. Evaluation of accuracy/convergence criteria. If not met, return to step 4.

A graphical representation of the complete framework developed as part of this work is shown in Figure 3.32. The framework consists of an outer-level management loop which successively calls a hierarchy of generalised subroutines for surrogate model generation and multi-fidelity correction. An advantage of this arrangement is the flexibility allowed during problem formulation. For example, with generalised surrogate generation and correction subroutines, any number of alternative/competing models may be specified for a particular subsystem. Models themselves may have non-equal (or overlapping) input domains, and may be time-phase specific or invariant. Sample points at which models are to be initially evaluated may be supplied, or autonomously selected during the presented surrogate generation procedure. Additionally, existing data may be supplied to augment the model generation process as sets of corresponding input/output arrays. The framework itself is designed to be modular where appropriate, such that alternative surrogate generation techniques and multi-fidelity correction methods may be easily substituted.

3.2.1 Initial Surrogate Generation

To begin, the above correction techniques are successively applied to a low-fidleity sample set selected using LHS and supplemented with a minimum of 2 samples of the expensive

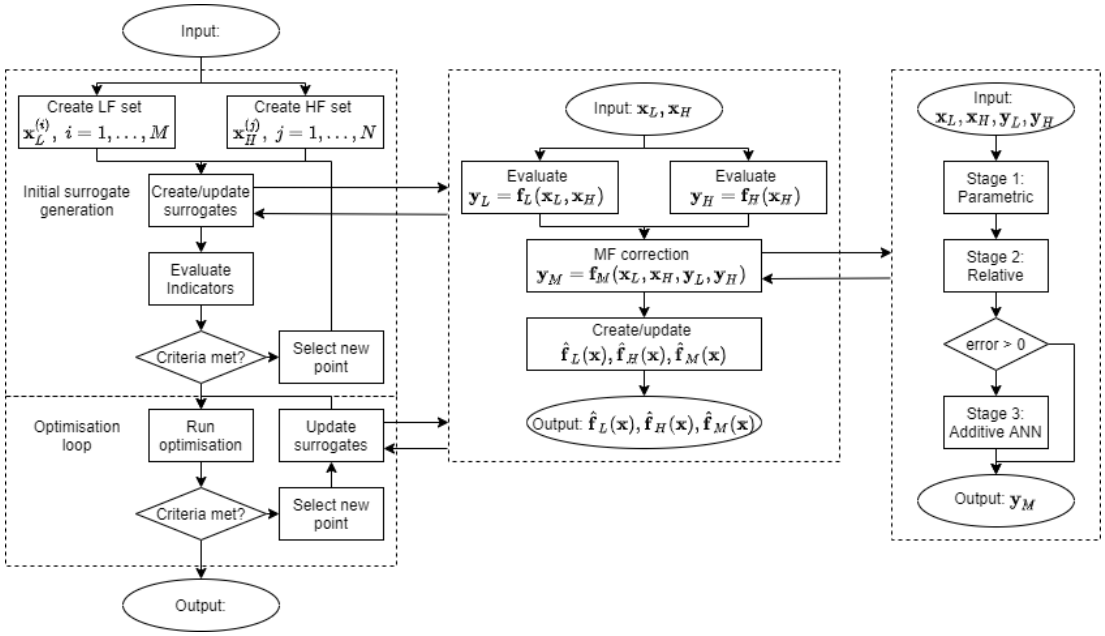


FIGURE 3.32: The complete integrated *Multi-Fidelity Model Management* (MFMM) framework. The left-most section represents the outer-loop, including an initial pre-processing step where surrogates are adaptively constructed to meet an acceptable accuracy criteria, and the optimisation loop, where surrogates are iteravely updated as the optimisation process progresses. The middle section represents the multi-fidelity surrogate generation subroutine, where the create/update action is itself a generic subroutine that accepts only lists of correlated input/output data. Finally, the right-most section represents the multi-fidelity correction procedure, where in this case the 3-stage approach is implemented. The framework is designed to be modular such that alternative surrogate generation techniques and multi-fidelity correction methods may be easily substituted.

high-fidelity function (such that an output gradient may be observed, required for the multi-fidelity correction procedure to operate), taken at the boundary values of the input vector. Boundary values are chosen with the assumption of a lower resulting likelihood of undesirable extrapolation, given limited/no prior knowledge of the function response.

During each subsequent iteration, 'current' surrogate models are generated/updated using the resultant datasets. To adaptively select new points for high-fidelity evaluation, the sensitivity of the current multi-fidelity approximation is considered. The total sensitivity S_{Ti} represents the contribution of each input element x_i to the output variance $V(Y)$ (Guenther et al., 2015):

$$S_{Ti} = E_{X_{-i}}(V_{X_i}(Y|X_{-i}))/V(Y) \tag{3.54}$$

where $V_{X_i}(Y|X_{-i})$ is the variance of the function $f(x) = y$, over all values of $X_i = x$ with all other variables $X_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k)$ held constant; and $E_{X_{-i}}(V_{X_i}(Y|X_{-i}))$ is the expectation of that variance over all values of $X_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k)$.

Taking matrices A and $\mathbf{B}$ as independent draws of P random samples from the input domain,

matrix $\mathbf{A}_B$ is defined as:

$$\mathbf{A}_B^{(i)} = [\mathbf{A}[,1], \ldots, \mathbf{A}[,i-1], \mathbf{B}[,i], \mathbf{A}[,i+1], \ldots, \mathbf{A}[,k]] \tag{3.55}$$

where k is the number of independent variables x_i, $\mathbf{A}[,j]$ with $j = (1, \ldots, i-1, i+1, \ldots, k)$ represents the j^{th} column of matrix $\mathbf{A}$, and $\mathbf{B}[,i]$ represents the i^{th} column of matrix $\mathbf{B}$. The total sensitivity is evaluated using the Monte Carlo approximation presented by (Jansen, 1999; Saltelli et al., 2010):

$$\mathbf{S}_{Ti} \approx \left[\frac{1}{2P} \sum_{j=1}^{P} \left(f(\mathbf{A})_j - f(\mathbf{A}_B^{(i)})_j \right)^2 \right] / V(Y) \tag{3.56}$$

$$V(Y) = \frac{1}{P} \sum_{j=1}^{P} (Y_j - \bar{Y})^2 \tag{3.57}$$

Global exploration is promoted by defining the following proximity-based relationship, with bias determined by the model sensitivity relative to each input element:

$$\mathbf{d} = \sqrt{\sum_{i=1}^{k} \mathbf{S}_T^{(i)} \left(\mathbf{x}_H^{(i)} - \mathbf{x}_c^{(i)} \right)^2} \tag{3.58}$$

where the subscript c denotes a candidate point. Equation (3.58) is incorporated into the following optimisation problem:

$$\begin{aligned} &\arg\max_{\mathbf{x}_c} \mathbf{d}(\mathbf{x}) \\ s.t. \quad &\mathbf{x} \in [-1, 1]^k \end{aligned} \tag{3.59}$$

Thus at each iteration, a new sample may be selected from within the input domain at the location furthest from each existing high-fidelity sample, with bias given to input vector elements exhibiting a higher influence on the output response.

Particularly in problems where model evaluations are computationally expensive, stopping criteria for preliminary sampling procedures are often defined by a specific budget of function evaluations (Liu et al., 2018b). A more reliable method is to measure the successive relative improvement (SRI) in model approximation via some error criteria (e.g. CV error, jackknifing variance, relative absolute error (Kim et al., 2009) etc). In this study, the convergence of the initial multi-fidelity surrogate is indicated by a series of behavioural tests, adapted from the work of (Guenther et al., 2015):

- **Indicator 1**: Convergence is suggested by the stabilisation of sensitivity variability and trend. If the difference between the mean sensitivities of the preceding iterations and the

average of the mean sensitivities of the current iteration is within 3 times the value of averaged standard error, variability has converged. Trend convergence is achieved when the gradient term of a linear regression shows a change below a predefined tolerance over a certain number of iterations, with an absolute value greater than 0.003.

- **Indicator 2**: To suggest that models of previous iterations are similar to the current approximation, the average mean standard deviation should be greater than or equal to a multiple (in this case, 3) of the standard deviation of the standard deviations.
- **Indicator 3**: Accurate representation of the desired function is suggested by the mean absolute error of adaptively sampled points falling below a specified tolerance, in this case $< 10\%$, of the observed output range.
- **Indicator 4**: Convergence is suggested if 98.5% of the standardised errors of the adaptively sampled points over previous iterations are within [-3,3].

3.2.2 Adaptive Sampling

Following the successful generation of initial multi/high-fidelity surrogates, the optimisation process begins. The problem solution at each iteration determines the subsequent point to be sampled, with the aim to maximise the improvement in surrogate prediction quality while minimising the necessary number of samples (and thus cost). The proposed methodology is functionally similar to a homogeneous *Query-By-Committee* (QBC) formulation (Seung et al., 1992) (see Section 2.2.3.2), in that sampling is driven by response variances estimated by alternative (competing) surrogates as a representation of prediction error. A symmetric multi-response classification (Section 2.2.5.1) is adopted for the handling of multiple outputs from a single function.

At each iteration, the state profile of the current solution is used to deconstruct the multi-fidelity input domain into a series of discretised element specific hypercubes, focusing the input-space towards the current solution. Each region is populated with a set of candidate points selected via LHS (thus ensuring candidates are well-spaced, helping to avoid excessive local exploitation (Gramacy and Lee, 2009)). As in the work of (Han et al., 2010), it is assumed that competing surrogates will eventually converge to the same result given a sufficient number of samples. Thus relative discrepancy is chosen as the primary driver. Each candidate point is evaluated using the available surrogates and ranked according to:

$$\begin{aligned} \hat{\mathbf{x}} &\leftarrow \max \frac{1}{k} \sum_{k=1}^{n_m} (\mathbf{y}^{(k)} - \bar{\mathbf{y}})^2 \\ s.t. \quad & d_i \geq \beta * \bar{\mathbf{d}} \\ & d_i = |\mathbf{x}_i - \mathbf{x}| \end{aligned} \tag{3.60}$$

where n_m is the number of alternative models and the additional distance penalty coefficient $\beta = 0.2$ (as in (Zhou et al., 2016)), is imposed to further reduce the risk of sample clustering.

3.3 Chapter Summary

The overarching goal of this work is to contribute in part to a generalised framework supporting multi-fidelity optimisation and optimal control problems. Accordingly, this work proposes a complete MFMM framework in which surrogate generation and data correction methods are implemented within a modular outer-level structure. Each subroutine is designed to be applicable in a general sense, to enhance the flexibility and applicability of the framework. For example, multiple alternative/competing models may be specified for a particular subsystem. Models themselves may have non-equal (or overlapping) input domains, and may be time-phase specific or shared across multiple phases. Sample points at which models are to be initially evaluated may be supplied, or autonomously selected via the proposed surrogate generation procedure. Additionally, existing data may be supplied to augment the model generation process as sets of corresponding input/output arrays.

Test functions	d	cost*	model								
			P	R	A	PR	RA	PA	PRA	HFS	LFS
Forrester	1	5	-4.64e3 (3.63e4)	48.589 (19.387)	72.073 (39.95)	-7.92e2 (2.52e3)	**76.652** (31.24)	-99.795 (5.19e2)	-1.73e2 (1.02e3)	-2.22e2 (1.11e3)	-1.21e2 (6.906)
		10	-20.349 (79.78)	54.586 (6.665)	98.647 (0.687)	-26.521 (137.591)	**98.749** (0.665)	98.545 (1.152)	98.729 (0.763)	98.571 (0.819)	-121.047 (6.845)
		15	-7.468 (27.584)	55.262 (4.4)	99.1 (0.344)	6.758 (34.25)	**99.282** (0.266)	99.082 (0.396)	99.223 (0.323)	99.164 (0.492)	-121.339 (6.927)
		20	0.623 (13.224)	56.164 (4.178)	99.161 (0.312)	10.091 (24.172)	99.275 (0.27)	99.231 (0.227)	**99.319** (0.183)	99.307 (0.378)	-121.471 (6.464)
Sasena	1	5	22.998 (37.871)	92.336 (3.409)	99.115 (0.83)	65.947 (29.057)	**99.167** (0.746)	96.473 (4.058)	95.725 (5.035)	42.902 (43.476)	-28.845 (5.362)
		10	45.791 (9.624)	92.716 (1.609)	99.624 (0.233)	82.054 (6.848)	**99.697** (0.188)	99.572 (0.282)	99.592 (0.226)	99.004 (1.276)	-29.085 (6.147)
		15	48.261 (6.907)	92.699 (1.095)	99.678 (0.187)	84.464 (4.039)	**99.7** (0.152)	99.611 (0.203)	99.649 (0.22)	99.523 (0.395)	-27.614 (5.085)
		20	47.753 (5.083)	92.59 (1.091)	99.643 (0.214)	84.648 (3.12)	**99.7** (0.164)	99.636 (0.203)	99.693 (0.196)	99.613 (0.29)	-28.891 (5.556)
Currin	2	5	95.296 (1.824)	51.114 (4.024)	30.939 (22.224)	**96.251** (1.664)	34.798 (21.712)	95.53 (2.526)	95.481 (2.418)	60.755 (30.205)	-12.833 (3.415)
		10	95.9 (1.434)	52.553 (3.493)	67.924 (24.661)	96.598 (1.428)	72.015 (24.675)	**97.625** (2.303)	97.475 (2.167)	92.575 (14.574)	-13.157 (3.292)
		15	95.991 (1.283)	52.42 (4.174)	81.99 (16.435)	96.695 (1.037)	87.79 (11.845)	98.343 (1.723)	98.269 (2.93)	**98.369** (1.354)	-13.56 (3.819)
		20	95.417 (2.019)	52.402 (3.31)	87.856 (15.307)	97.049 (1.119)	90.904 (11.051)	98.961 (0.662)	**99.052** (0.672)	98.548 (1.563)	-13.21 (3.14)
Heterogeneous	2	5	82.653 (5.39)	15.068 (52.004)	67.74 (22.815)	83.801 (6.551)	68.964 (22.048)	82.777 (12.329)	**85.437** (7.007)	62.178 (33.964)	-9.54e2 (61.316)
		10	83.54 (3.597)	23.536 (23.621)	91.377 (8.922)	86.318 (4.441)	89.613 (11.727)	**91.789** (6.093)	90.979 (9.604)	89.34 (12.378)	-9.51e3 (56.229)
		15	84.205 (3.805)	26.409 (20.516)	96.693 (1.927)	86.82 (2.707)	96.791 (1.99)	96.84 (2.098)	96.413 (3.192)	**97.187** (2.517)	-9.51e2 (56.031)
		20	84.316 (3.028)	26.722 (17.501)	97.631 (1.314)	86.777 (2.778)	97.773 (1.622)	97.642 (1.997)	97.859 (1.784)	**98.269** (0.952)	-9.59e2 (58.244)
Rastrigin	2	5	83.007 (5.004)	**86.625** (2.665)	65.898 (18.032)	79.445 (21.717)	71.197 (14.518)	65.536 (17.893)	72.401 (12.813)	-13.195 (64.996)	31.106 (9.431)
		10	84.283 (4.517)	**87.188** (2.62)	74.787 (14.393)	84.427 (4.333)	78.756 (11.671)	82.18 (11.667)	81.931 (12.661)	78.218 (27.266)	31.17 (9.682)
		15	84.99 (3.5)	87.221 (2.675)	89.861 (10.704)	85.568 (2.828)	92.329 (5.588)	93.989 (6.142)	94.284 (6.757)	**95.577** (2.378)	31.313 (9.264)
		20	84.877 (3.144)	87.183 (2.583)	94.611 (6.341)	85.351 (3.606)	94.462 (5.073)	96.396 (4.456)	95.253 (5.653)	**97.219** (1.966)	31.662 (9.823)
Rosenbrock	2	5	97.744 (1.27)	-557.289 (611.36)	24.872 (66.755)	**98.053** (1.269)	16.997 (96.249)	97.882 (1.143)	97.775 (1.209)	26.891 (81.034)	-1.21e4 (8.83e2)
		10	97.33 (2.711)	-421.975 (366.577)	85.6 (21.593)	98.02 (1.183)	88.616 (10.956)	**98.369** (0.944)	98.255 (1.327)	86.822 (16.699)	-1.21e4 (8.04e2)
		15	98.142 (1.087)	-416.323 (311.124)	94.148 (8.499)	98.24 (1.352)	95.537 (4.617)	98.639 (1.196)	**98.753** (0.802)	95.189 (5.909)	-1.19e4 (8.85e2)
		20	97.516 (1.946)	-373.452 (279.175)	96.426 (3.737)	98.243 (0.97)	97.365 (2.615)	98.704 (1.796)	**98.931** (0.687)	97.793 (1.871)	-1.20e4 (9.82e2)
Paciorek	2	5	-3.30e4 (2.74e5)	13.026 (21.545)	48.79 (19.774)	-2.00e2 (5.15e2)	45.629 (21.119)	56.455 (56.173)	49.536 (66.231)	**60.082** (35.936)	-35.263 (23.551)
		10	-5.38e3 (3.83e4)	16.724 (14.423)	60.813 (20.728)	-78.956 (1.93e2)	60.435 (23.392)	**85.048** (14.186)	84.81 (16.968)	84.332 (12.851)	-34.344 (23.345)
		15	-1.56e3 (5.60e3)	19.304 (14.288)	67.201 (22.772)	-68.761 (186.82)	70.545 (24.592)	88.317 (11.652)	89.277 (11.508)	**91.486** (8.839)	-35.008 (23.441)
		20	-6.45e2 (9.01e2)	19.193 (14.846)	76.558 (21.942)	-43.602 (99.067)	75.396 (23.149)	91.341 (11.047)	91.205 (11.37)	**92.764** (9.821)	-34.325 (22.901)
Park2	4	5	99.266 (0.313)	98.514 (0.422)	99.392 (0.235)	99.271 (0.294)	**99.424** (0.194)	99.344 (0.252)	99.357 (0.228)	98.11 (1.637)	65.332 (4.065)
		10	99.25 (0.289)	98.595 (0.381)	99.48 (0.233)	99.286 (0.295)	99.473 (0.223)	**99.494** (0.168)	99.425 (0.26)	99.107 (0.487)	64.986 (4.137)
		15	99.296 (0.245)	98.593 (0.355)	99.492 (0.194)	99.331 (0.268)	99.489 (0.199)	99.536 (0.157)	**99.553** (0.15)	99.391 (0.285)	65.778 (4.167)
		20	99.24 (0.319)	98.643 (0.344)	99.532 (0.15)	99.304 (0.222)	**99.554** (0.159)	99.527 (0.165)	99.527 (0.164)	99.424 (0.211)	65.338 (4.042)
Hartman	6	5	51.189 (17.29)	12.93 (22.988)	45.232 (15.409)	52.767 (15.74)	49.471 (13.637)	55.245 (12.188)	**56.167** (12.423)	-4.033 (35.19)	-27.559 (28.564)
		10	50.214 (12.715)	17.587 (23.413)	48.403 (11.918)	56.064 (11.418)	49.792 (10.46)	**56.303** (10.442)	55.267 (8.619)	20.736 (17.606)	-25.113 (31.015)
		15	52.935 (10.525)	20.691 (18.8)	53.585 (10.627)	55.887 (11.399)	51.921 (10.876)	**57.557** (8.222)	57.144 (9.476)	33.16 (13.855)	-28.756 (30.066)
		20	50.724 (12.496)	20.295 (18.625)	56.275 (9.378)	55.407 (12.112)	55.564 (9.295)	**58.122** (6.926)	57.655 (8.449)	41.248 (9.847)	-28.037 (36.041)
Borehole	8	5	99.453 (0.196)	94.331 (1.007)	99.405 (0.235)	99.414 (0.219)	99.269 (0.278)	**99.48** (0.18)	99.475 (0.166)	99.086 (0.536)	71.667 (2.407)
		10	99.462 (0.196)	94.119 (1.238)	99.55 (0.183)	99.496 (0.182)	99.526 (0.205)	99.563 (0.152)	**99.575** (0.145)	99.495 (0.156)	71.882 (2.397)
		15	99.44 (0.212)	94.483 (1.118)	99.59 (0.172)	99.471 (0.189)	99.594 (0.139)	**99.627** (0.11)	99.624 (0.134)	99.543 (0.19)	71.886 (2.04)
		20	99.441 (0.205)	94.45 (1.07)	99.587 (0.146)	99.487 (0.193)	99.621 (0.132)	99.617 (0.122)	**99.638** (0.114)	99.563 (0.155)	71.657 (2.183)

Table 3.17: R^2 values for each multi/single-fidelity method applied to each benchmark function. Reported are the values at distinct levels of computational cost relative to problem dimensionality. All values are averaged over 100 independent runs, with standard deviation included in brackets.

Test functions	d	cost*	model									
			P	R	A	PR	RA	PA	PRA	HFS	LFS	
Forrester	1	5	0.504 (0.023)	0.867 (0.084)	0.914 (0.057)	0.747 (0.198)	**0.917** (0.059)	0.777 (0.241)	0.784 (0.258)	0.711 (0.301)	0.524 (0.02)	
		10	0.504 (0.023)	0.892 (0.039)	0.988 (0.006)	0.885 (0.051)	0.99 (0.005)	0.989 (0.006)	**0.99** (0.005)	0.988 (0.006)	0.524 (0.02)	
		15	0.502 (0.022)	0.894 (0.024)	0.992 (0.003)	0.885 (0.03)	**0.993** (0.002)	0.992 (0.004)	**0.993** (0.003)	**0.993** (0.005)	0.524 (0.021)	
		20	0.501 (0.026)	0.895 (0.02)	0.993 (0.003)	0.887 (0.022)	0.993 (0.002)	0.993 (0.002)	**0.994** (0.002)	**0.994** (0.004)	0.522 (0.02)	
Sasena	1	5	0.662 (0.018)	0.942 (0.032)	**0.993** (0.008)	0.938 (0.039)	**0.993** (0.006)	0.981 (0.016)	0.978 (0.016)	0.696 (0.127)	0.667 (0.013)	
		10	0.661 (0.018)	0.941 (0.013)	0.997 (0.002)	0.957 (0.014)	**0.998** (0.002)	0.997 (0.003)	0.997 (0.002)	0.991 (0.012)	0.669 (0.015)	
		15	0.66 (0.019)	0.94 (0.009)	**0.998** (0.002)	0.961 (0.008)	**0.998** (0.001)	0.997 (0.002)	0.997 (0.002)	0.996 (0.004)	0.668 (0.013)	
		20	0.661 (0.019)	0.94 (0.008)	0.997 (0.002)	0.961 (0.007)	**0.998** (0.002)	**0.998** (0.002)	**0.998** (0.002)	0.997 (0.003)	0.67 (0.011)	
Currin	2	5	0.964 (0.013)	0.854 (0.056)	0.412 (0.242)	**0.969** (0.014)	0.445 (0.222)	0.963 (0.02)	0.963 (0.02)	0.701 (0.159)	0.966 (0.011)	
		10	0.965 (0.012)	0.866 (0.045)	0.714 (0.228)	0.971 (0.012)	0.759 (0.228)	**0.979** (0.019)	0.978 (0.018)	0.942 (0.093)	0.966 (0.01)	
		15	0.965 (0.01)	0.862 (0.055)	0.842 (0.156)	0.972 (0.008)	0.894 (0.109)	**0.985** (0.016)	**0.985** (0.025)	**0.985** (0.013)	0.966 (0.009)	
		20	0.96 (0.018)	0.87 (0.048)	0.898 (0.127)	0.973 (0.011)	0.918 (0.104)	**0.991** (0.006)	**0.991** (0.006)	0.987 (0.014)	0.964 (0.011)	
Heterogeneous	2	5	0.867 (0.025)	0.829 (0.043)	0.744 (0.155)	**0.902** (0.022)	0.769 (0.115)	0.867 (0.09)	0.883 (0.052)	0.714 (0.191)	0.884 (0.021)	
		10	0.867 (0.024)	0.835 (0.041)	**0.928** (0.065)	0.904 (0.019)	0.916 (0.084)	0.927 (0.053)	0.921 (0.077)	0.911 (0.094)	0.883 (0.021)	
		15	0.868 (0.023)	0.842 (0.033)	0.97 (0.017)	0.905 (0.02)	0.972 (0.017)	0.972 (0.017)	0.968 (0.026)	**0.975** (0.02)	0.883 (0.021)	
		20	0.869 (0.023)	0.845 (0.03)	0.978 (0.011)	0.904 (0.019)	0.98 (0.014)	0.978 (0.019)	0.98 (0.016)	**0.984** (0.008)	0.884 (0.02)	
Rastrigin(shifted)	2	5	0.87 (0.026)	0.874 (0.024)	0.706 (0.139)	0.869 (0.045)	0.747 (0.111)	0.718 (0.116)	0.768 (0.09)	0.254 (0.22)	**0.877** (0.022)	
		10	0.871 (0.025)	0.877 (0.024)	0.773 (0.125)	0.877 (0.024)	0.809 (0.104)	0.843 (0.092)	0.839 (0.102)	0.824 (0.173)	**0.878** (0.021)	
		15	0.875 (0.023)	0.876 (0.025)	0.906 (0.098)	0.878 (0.021)	0.929 (0.051)	0.946 (0.053)	0.948 (0.06)	**0.959** (0.021)	0.879 (0.021)	
		20	0.872 (0.024)	0.876 (0.024)	0.95 (0.058)	0.877 (0.023)	0.948 (0.048)	0.967 (0.038)	0.957 (0.049)	**0.974** (0.017)	0.879 (0.021)	
Rosenbrock	2	5	0.983 (0.008)	0.657 (0.118)	0.646 (0.181)	0.985 (0.009)	0.686 (0.163)	0.984 (0.008)	0.982 (0.009)	0.652 (0.219)	**0.989** (0.006)	
		10	0.98 (0.014)	0.717 (0.08)	0.922 (0.077)	0.984 (0.009)	0.934 (0.058)	0.987 (0.007)	0.986 (0.01)	0.914 (0.107)	**0.989** (0.007)	
		15	0.985 (0.008)	0.735 (0.075)	0.963 (0.036)	0.986 (0.01)	0.972 (0.022)	**0.99** (0.007)	**0.99** (0.006)	0.968 (0.031)	**0.99** (0.005)	
		20	0.98 (0.011)	0.76 (0.055)	0.975 (0.021)	0.986 (0.007)	0.981 (0.013)	**0.991** (0.009)	**0.991** (0.005)	0.984 (0.011)	0.99 (0.004)	
Paciorek	2	5	0.138 (0.057)	0.244 (0.103)	0.541 (0.188)	0.496 (0.171)	0.514 (0.201)	**0.741** (0.173)	0.713 (0.196)	0.725 (0.146)	0.14 (0.053)	
		10	0.138 (0.061)	0.246 (0.09)	0.634 (0.204)	0.59 (0.123)	0.631 (0.224)	0.867 (0.118)	**0.872** (0.13)	0.862 (0.113)	0.139 (0.059)	
		15	0.14 (0.058)	0.249 (0.094)	0.69 (0.217)	0.606 (0.12)	0.723 (0.234)	0.893 (0.107)	0.9 (0.109)	**0.921** (0.082)	0.139 (0.057)	
		20	0.137 (0.056)	0.25 (0.089)	0.776 (0.211)	0.607 (0.098)	0.767 (0.222)	0.919 (0.103)	0.917 (0.106)	**0.932** (0.091)	0.142 (0.058)	
Park2	4	5	0.994 (0.002)	0.994 (0.002)	0.995 (0.002)	0.994 (0.002)	0.995 (0.002)	0.995 (0.002)	0.995 (0.002)	0.986 (0.009)	**0.997** (0.001)	
		10	0.994 (0.002)	0.994 (0.003)	**0.996** (0.002)	0.994 (0.002)	**0.996** (0.001)	**0.996** (0.001)	0.995 (0.002)	0.993 (0.004)	**0.996** (0.001)	
		15	0.994 (0.002)	0.994 (0.002)	0.996 (0.001)	0.995 (0.002)	0.996 (0.001)	0.996 (0.001)	0.996 (0.001)	0.995 (0.002)	**0.997** (0.001)	
		20	0.994 (0.003)	0.994 (0.002)	0.996 (0.001)	0.994 (0.002)	0.996 (0.001)	0.996 (0.001)	0.996 (0.001)	0.995 (0.002)	**0.997** (0.001)	
Hartman	6	5	0.623 (0.055)	**0.655** (0.059)	0.561 (0.103)	0.643 (0.058)	0.591 (0.087)	0.616 (0.074)	0.617 (0.082)	0.235 (0.097)	0.602 (0.062)	
		10	0.612 (0.066)	**0.673** (0.064)	0.539 (0.097)	0.642 (0.067)	0.547 (0.096)	0.589 (0.089)	0.58 (0.081)	0.308 (0.109)	0.614 (0.059)	
		15	0.612 (0.06)	**0.668** (0.059)	0.57 (0.086)	0.642 (0.055)	0.554 (0.085)	0.597 (0.071)	0.594 (0.084)	0.394 (0.1)	0.604 (0.07)	
		20	0.607 (0.059)	**0.678** (0.056)	0.591 (0.082)	0.639 (0.065)	0.588 (0.081)	0.602 (0.067)	0.598 (0.075)	0.447 (0.083)	0.621 (0.063)	
Borehole	8	5	0.996 (0.001)	0.995 (0.002)	0.995 (0.002)	0.996 (0.002)	0.995 (0.002)	0.996 (0.001)	0.996 (0.001)	0.993 (0.004)	**0.997** (0.001)	
		10	0.996 (0.002)	0.995 (0.002)	0.996 (0.001)	0.996 (0.001)	**0.997** (0.001)	0.997 (0.001)	**0.997** (0.001)	0.996 (0.001)	**0.997** (0.001)	
		15	0.996 (0.002)	0.995 (0.002)	**0.997** (0.001)	0.996 (0.001)	**0.997** (0.001)	**0.997** (0.001)	**0.997** (0.001)	0.996 (0.001)	**0.997** (0.001)	
		20	0.996 (0.001)	0.995 (0.001)	**0.997** (0.001)	0.996 (0.001)	**0.997** (0.001)	**0.997** (0.001)	**0.997** (0.001)	**0.997** (0.001)	**0.997** (0.001)	

Table 3.18: Pearson's correlation coefficient values for each multi/single-fidelity method applied to each benchmark function. Reported are the values at distinct levels of computational cost relative to problem dimensionality. All values are averaged over 100 independent runs, with standard deviation included in brackets.

Test functions	d	cost*	model								
			P	R	A	PR	RA	PA	PRA	HFS	LFS
Forrester	1	5	6.33 (2.533)	6.158 (2.473)	7.816 (3.919)	**5.38** (1.682)	7.731 (3.501)	7.015 (3.296)	6.71 (3.198)	5.549 (2.515)	6.671 (1.72)
		10	6.047 (2.55)	**5.324** (1.602)	9.399 (3.03)	5.485 (1.712)	10.221 (3.301)	10.27 (3.172)	10.78 (3.81)	9.294 (4.484)	6.462 (1.751)
		15	4.886 (1.191)	6.09 (2.533)	10.046 (2.871)	**4.803** (1.1)	10.566 (2.936)	10.596 (3.154)	10.764 (3.249)	10.462 (4.832)	6.576 (1.315)
		20	5.012 (1.638)	**4.774** (1.173)	9.974 (3.078)	4.895 (1.404)	11.236 (2.741)	10.709 (3.046)	11.182 (3.158)	12.192 (5.395)	6.525 (1.341)
Sasena	1	5	9.79 (3.224)	7.721 (2.746)	7.775 (3.047)	7.887 (2.948)	7.662 (2.904)	7.586 (2.937)	**7.141** (3.071)	7.205 (3.32)	8.413 (2.87)
		10	9.979 (4.235)	7.676 (2.598)	7.345 (2.944)	7.553 (2.712)	7.234 (2.957)	7.325 (2.894)	**7.184** (2.998)	8.039 (3.475)	8.911 (2.952)
		15	8.893 (2.698)	8.773 (4)	7.38 (3.051)	7.914 (3.119)	7.92 (3.673)	7.209 (2.849)	**7.144** (2.976)	7.763 (3.172)	9.169 (3.152)
		20	8.784 (2.985)	7.878 (2.945)	7.419 (2.936)	7.795 (2.738)	7.448 (3.242)	7.34 (2.833)	**7.091** (2.999)	7.227 (3.258)	8.88 (3.31)
Currin	2	5	11.438 (5.197)	11.774 (4.427)	17.088 (8.887)	8.468 (3.69)	17.164 (9.037)	7.998 (3.381)	8.415 (4.096)	**4.693** (0.605)	10.563 (3.194)
		10	9.844 (3.915)	11.768 (4.289)	16.905 (10.063)	8.254 (3.141)	16.212 (7.815)	7.565 (2.155)	8.285 (2.944)	**5.628** (1.317)	10.891 (4.156)
		15	9.138 (3.942)	11.975 (4.206)	12.627 (6.189)	7.932 (2.727)	12.538 (4.974)	7.314 (2.317)	8.083 (2.714)	**6.152** (1.847)	10.031 (3.704)
		20	9.905 (3.463)	12.928 (5.067)	12.694 (5.922)	7.646 (2.942)	11.672 (4.367)	8.897 (3.476)	7.701 (2.729)	**6.698** (2.251)	10.061 (3.803)
Heterogeneous	2	5	14.6 (8.663)	12.854 (7.611)	13.029 (7.294)	12.341 (7.443)	13.609 (7.4)	14.071 (7.344)	14.08 (7.027)	**12.197** (7.241)	15.553 (6.802)
		10	14.446 (8.845)	**12.388** (7.359)	13.406 (7.395)	12.275 (7.4)	14.157 (7.298)	14.863 (6.94)	15.037 (7.034)	13.542 (7.238)	15.262 (6.513)
		15	12.912 (7.59)	12.39 (7.404)	12.828 (6.999)	**12.309** (7.396)	13.243 (7)	14.17 (7.432)	14.328 (7.251)	13.721 (6.845)	15.315 (6.885)
		20	13.09 (7.476)	**12.161** (7.315)	12.9 (7.344)	12.414 (7.574)	13.297 (6.948)	14.442 (7.185)	14.373 (7.377)	14.772 (7.061)	14.926 (7.143)
Rastrigin(shifted)	2	5	36.441 (23.647)	33.96 (21.674)	38.041 (23.096)	34.918 (21.228)	35.06 (21.599)	34.098 (20.643)	35.33 (20.781)	**28.852** (20.111)	32.138 (21.185)
		10	34.566 (21.392)	32.983 (21.468)	35.277 (20.746)	33.256 (19.994)	34.017 (20.206)	32.208 (17.838)	33.866 (18.396)	**31.974** (19.948)	32.717 (20.188)
		15	32.569 (19.198)	33.012 (22.068)	32.615 (19.048)	34.084 (20.826)	32.056 (19.124)	**30.213** (18.484)	31.561 (19.636)	32.388 (19.331)	31.589 (19.391)
		20	33.388 (20.902)	33.734 (20.781)	31.269 (19.527)	33.199 (21.399)	31.181 (18.49)	30.741 (19.188)	**29.656** (19.814)	31.364 (20.208)	30.46 (20.374)
Rosenbrock	2	5	29.749 (4.002)	26.88 (2.884)	25.261 (2.748)	26.041 (2.376)	24.825 (2.994)	25.332 (2.664)	23.4 (2.832)	**21.047** (2.469)	23.868 (2.607)
		10	28.386 (4.222)	25.185 (2.52)	25.372 (2.262)	25.666 (2.278)	24.928 (2.687)	23.717 (2.556)	22.276 (3.199)	**20.841**(3.254)	22.23 (3.13)
		15	26.324 (2.924)	24.41 (1.824)	25.165 (2.338)	25.807 (2.122)	24.672 (2.564)	21.996 (3.238)	20.774 (3.654)	**19.757** (3.77)	20.341 (3.485)
		20	27.768 (2.945)	25.091 (2.047)	25.497 (2.105)	25.951 (2.356)	24.787 (1.961)	22.262 (2.989)	21.921 (4.276)	**20.76** (3.712)	21.299 (3.888)
Paciorek	2	5	24.448 (19.642)	22.45 (18.523)	26.062 (19.128)	21.537 (18.894)	25.13 (19.445)	22.248 (20.102)	22.22 (19.025)	**18.586** (18.758)	32.154 (18.013)
		10	23.667 (19.106)	21.98 (18.252)	25.138 (18.644)	21.189 (18.479)	24.515 (18.364)	20.152 (18.678)	19.832 (18.588)	**19.527** (18.967)	32.289 (18.15)
		15	22.131 (18.68)	21.558 (18.228)	24.947 (18.873)	21.007 (18.576)	23.998 (18.086)	20.001 (18.653)	**19.675** (19.044)	19.752 (18.903)	32.279 (18.729)
		20	22.639 (18.559)	21.738 (18.435)	25.103 (18.358)	21.131 (18.791)	24.038 (17.569)	19.806 (18.528)	**18.764** (18.452)	19.231 (18.413)	32.002 (18.414)
Park2	4	5	6.017 (3.494)	6.012 (3.575)	5.9 (3.308)	5.58 (3.176)	5.789 (2.943)	5.399 (2.898)	5.316 (3.081)	**5.205** (3.091)	5.271 (3.063)
		10	6.885 (4.369)	6.044 (3.703)	5.985 (3.412)	5.679 (3.18)	5.729 (2.94)	5.327 (3.142)	5.207 (2.88)	**5.139** (2.885)	5.148 (2.953)
		15	6.191 (3.645)	5.955 (3.44)	5.784 (3.258)	5.61 (3.061)	5.518 (2.818)	5.195 (2.96)	**5.038** (2.78)	5.052 (2.818)	5.087 (2.961)
		20	6.172 (3.639)	5.89 (3.412)	5.901 (3.327)	5.668 (3.178)	5.567 (3)	5.304 (3.169)	5.159 (2.809)	5.192 (2.88)	**5.108** (2.896)
Hartman	6	5	52.189 (10.46)	48.834 (9.411)	44.575 (6.577)	46.526 (8.987)	44.375 (5.782)	46.577 (6.609)	45.365 (7.337)	**7.763** (1.543)	47.714 (6.296)
		10	50.28 (9.15)	49.858 (8.494)	45.019 (7.685)	45.062 (6.847)	42.613 (5.023)	45.066 (5.422)	45.289 (4.838)	**12.454** (5.301)	47.587 (6.254)
		15	51.625 (9.805)	48.636 (9.283)	44.164 (6.122)	45.614 (6.011)	44.327 (5.426)	44.693 (6.192)	45.755 (6.725)	**15.407** (1.965)	46.66 (5.65)
		20	50.528 (7.072)	48.541 (7.872)	44.914 (6.457)	45.142 (7.44)	45.81 (5.862)	44.942 (5.411)	46.321 (6.2)	**20.014** (3.271)	47.666 (6.594)
Borehole	8	5	15.529 (7.568)	14.253 (6.813)	13.63 (6.441)	12.681 (7.707)	11.757 (6.896)	9.957 (5.33)	8.725 (4.691)	**8.579** (4.759)	8.623 (4.699)
		10	14.701 (7.328)	14.935 (7.439)	13.676 (6.358)	12.68 (7.658)	11.288 (6.96)	6.046 (1.83)	5.877 (1.67)	**5.56** (1.852)	5.446 (1.397)
		15	14.236 (6.765)	14.083 (6.663)	12.491 (6.732)	12.573 (7.511)	10.91 (6.96)	5.178 (0.998)	5.16 (0.713)	**5.045** (0.531)	5.1 (0.563)
		20	14.45 (7.242)	14.44 (7.051)	12.571 (6.744)	12.312 (7.532)	11.024 (7.168)	**5.806** (2.413)	6.132 (3.21)	5.935 (2.869)	5.955 (2.85)

Table 3.19: Surrogate generation times for each multi/single-fidelity method applied to each benchmark function. Reported are the values at distinct levels of computational cost relative to problem dimensionality. All values are averaged over 100 independent runs, with standard deviation included in brackets.

Test functions	d	cost*	model						
			P	R	A	PR	RA	PA	PRA
Forrester	1	5	0.049 (0.044)	**0.001** (0)	9.236 (2.873)	0.018 (0.021)	11.414 (5.127)	10.042 (3.929)	11.222 (3.785)
		10	0.065 (0.092)	**0.001** (0)	10.069 (2.27)	0.058 (0.059)	12.77 (4.626)	11.427 (3.759)	11.509 (3.128)
		15	0.019 (0.014)	**0.001** (0)	10.612 (2.633)	0.017 (0.007)	12.692 (3.968)	10.897 (2.138)	11.162 (2.758)
		20	0.015 (0.007)	**0.002** (0)	10.671 (2.299)	0.018 (0.006)	12.184 (2.785)	11.58 (2.588)	11.257 (3.058)
Sasena	1	5	0.036 (0.029)	**0.001** (0)	9.885 (4.228)	0.025 (0.027)	10.308 (4.106)	10.848 (4.487)	12.281 (5.559)
		10	0.03 (0.017)	**0.001** (0)	9.65 (4.208)	0.055 (0.054)	10.26 (4.281)	10.112 (4.203)	12.067 (4.868)
		15	0.02 (0.005)	**0.001** (0)	9.783 (4.357)	0.022 (0.006)	11.221 (5.508)	10.333 (4.268)	11.827 (4.462)
		20	0.021 (0.007)	**0.002** (0)	10.071 (4.344)	0.024 (0.007)	10.889 (5.295)	10.28 (4.172)	11.882 (5.122)
Currin	2	5	0.108 (0.122)	**0.002** (0)	9.855 (2.076)	0.099 (0.122)	10.744 (3.189)	11.004 (3.352)	12.559 (3.937)
		10	0.019 (0.009)	**0.003** (0)	11.613 (3.659)	0.021 (0.005)	17.338 (10.597)	18.513 (8.934)	22.65 (12.835)
		15	0.023 (0.016)	**0.004** (0)	12.585 (4.795)	0.021 (0.004)	18.002 (9.74)	24.943 (13.691)	30.98 (16.243)
		20	0.027 (0.005)	**0.006** (0.001)	14.051 (6.15)	0.022 (0.004)	19.31 (10.25)	26.254 (11.816)	28.773 (15.842)
Heterogeneous	2	5	0.078 (0.108)	**0.002** (0)	20.763 (10.482)	0.017 (0.005)	22.2 (10.362)	23.214 (10.968)	24.842 (12.638)
		10	0.017 (0.007)	**0.003** (0)	22.813 (10.532)	0.063 (0.093)	26.651 (11.117)	28.083 (11.396)	28.636 (12.793)
		15	0.027 (0.038)	**0.005** (0)	23.091 (10.127)	0.034 (0.045)	27.156 (10.288)	29.483 (10.25)	31.416 (13.863)
		20	0.028 (0.006)	**0.006** (0.001)	23.84 (11.151)	0.022 (0.005)	28.422 (9.686)	31.308 (9.516)	31.973 (11.84)
Rastrigin(shifted)	2	5	0.076 (0.104)	**0.002** (0)	50.122 (30.634)	0.016 (0.005)	48.609 (30.53)	49.079 (31.521)	49.72 (32.837)
		10	0.016 (0.006)	**0.003** (0)	61.411 (34.908)	0.018 (0.005)	53.456 (32.361)	54.429 (31.301)	49.172 (29.296)
		15	0.016 (0.005)	**0.004** (0)	61.871 (34.428)	0.092 (0.114)	53.429 (32.015)	57.105 (32.743)	52.093 (32.921)
		20	0.039 (0.03)	**0.005** (0.001)	60.286 (33.203)	0.029 (0.032)	51.036 (29.062)	60.563 (37.955)	47.124 (26.651)
Rosenbrock	2	5	0.055 (0.068)	**0.002** (0)	37.379 (2.023)	0.015 (0.004)	38.919 (5.536)	38.655 (12.702)	40.799 (12.923)
		10	0.031 (0.048)	**0.003** (0)	38.844 (3.982)	0.019 (0.006)	41.027 (7.011)	47.615 (12.689)	52.061 (23.101)
		15	0.017 (0.007)	**0.004** (0)	38.179 (4.429)	0.019 (0.005)	40.934 (7.612)	52.857 (16.318)	64.773 (27.06)
		20	0.03 (0.015)	**0.006** (0.001)	39.246 (5.174)	0.022 (0.007)	42.29 (6.788)	59.871 (20.389)	67.668 (24.843)
Paciorek	2	5	0.058 (0.093)	**0.002** (0)	30.41 (26.541)	0.016 (0.006)	31.799 (27.777)	29.495 (27.468)	30.521 (25.971)
		10	0.015 (0.008)	**0.003** (0)	31.004 (26.086)	0.037 (0.053)	35.458 (26.25)	29.465 (26.277)	31.916 (28.003)
		15	0.018 (0.02)	**0.005** (0)	31.754 (26.028)	0.041 (0.059)	35.701 (27.58)	29.888 (26.231)	31.53 (26.35)
		20	0.024 (0.005)	**0.006** (0.001)	33.796 (25.371)	0.02 (0.005)	38.427 (26.113)	29.863 (26.075)	31.407 (25.553)
Park2	4	5	0.015 (0.002)	**0.006** (0)	11.666 (5.042)	0.021 (0.003)	12.341 (4.277)	21.38 (13.893)	28.265 (16.97)
		10	0.024 (0.003)	**0.01** (0.001)	12.894 (4.897)	0.026 (0.003)	13.929 (4.499)	35.896 (24.885)	45.069 (23.042)
		15	0.077 (0.092)	**0.015** (0.001)	13.337 (4.946)	0.153 (0.011)	14.314 (4.369)	45.309 (23.627)	55.032 (28.971)
		20	0.017 (0.003)	**0.019** (0.001)	13.72 (5.154)	0.036 (0.003)	14.608 (4.116)	55.012 (27.615)	48.611 (32.943)
Hartman	6	5	0.02 (0.005)	**0.011** (0)	15.909 (3.799)	0.028 (0.005)	15.274 (2.765)	16.333 (4.936)	18.857 (11.069)
		10	**0.021** (0.004)	0.022 (0.001)	24.731 (5.062)	0.039 (0.004)	23.876 (3.581)	24.965 (7.263)	29.141 (12.509)
		15	0.048 (0.077)	**0.032** (0.001)	34.793 (11.658)	0.052 (0.005)	35.429 (5.423)	33.492 (7.143)	40.767 (15.014)
		20	**0.023** (0.005)	0.042 (0.001)	42.686 (10.427)	0.064 (0.005)	42.94 (6.879)	41.659 (7.446)	51.837 (18.734)
Borehole	8	5	0.03 (0.005)	**0.02** (0.001)	26.086 (9.482)	0.039 (0.004)	23.786 (10.244)	33.69 (21.99)	45.854 (33.723)
		10	**0.023** (0.005)	0.039 (0.001)	27.269 (8.933)	0.06 (0.005)	23.782 (10.975)	68.978 (51.85)	87.323 (55.281)
		15	**0.027** (0.006)	0.058 (0.002)	25.71 (10.203)	0.082 (0.007)	23.603 (11.122)	99.435 (53.384)	107.779 (54.065)
		20	**0.028** (0.005)	0.078 (0.002)	26.076 (10.202)	0.103 (0.008)	25.159 (11.779)	112.286 (50.566)	115.403 (59.272)

Table 3.20: Execution times for each multi-fidelity correction method applied to each benchmark function. Reported are the values at distinct levels of computational cost relative to problem dimensionality. All values are averaged over 100 independent runs, with standard deviation included in brackets.

3.4 Bibliography

A. Aguilera and R. Pérez-Aguila. General n-dimensional rotations. *Proceedings of the WSCG SHORT Communications Papers*, pages 1–8, 2004. URL papers3://publication/uuid/6ADECEC1-F089-42FC-AA61-81C20D0E8B26.

N. M. Alexandrov, R. M. Lewis, C. R. Gumbert, L. L. Green, and P. A. Newman. Approximation and model management in aerodynamic optimization with variable-fidelity models. *Journal of Aircraft*, 38(6):1093–1101, 2001. ISSN 00218669. doi: 10.2514/2.2877.

D. Allaire and K. Willcox. A mathematical and computational framework for multifidelity design and analysis with computer models. *International Journal for Uncertainty Quantification*, 4(1):1–20, 2014. ISSN 21525099. doi: 10.1615/Int.J.UncertaintyQuantification.2013004121.

M. Bakr, J. W. Bandler, K. Madsen, and J. Sondergaard. An Introduction to the Space Mapping Technique. 103(3):239–248, 2002. ISSN 00138703. doi: 10.1023/A.

M. H. Bakr, J. W. Bandler, K. Madsen, J. E. Rayas-Sanchez, and J. Sondergaard. Space mapping optimization of microwave circuits exploiting surrogate models. *IEEE MTT-S International Microwave Symposium Digest*, 3(June):1785–1788, 2000. ISSN 0149645X. doi: 10.1109/mwsym.2000.862325.

M. J. Bayarri, R. Paulo, J. O. Berger, J. Sacks, J. A. Cafeo, J. Cavendish, C. H. Lin, and J. Tu. A framework for validation of computer models. *Technometrics*, 49(2):138–154, 2007. ISSN 00401706. doi: 10.1198/004017007000000092.

B. Beachkofski and R. Grandhi. Improved distributed hypercube sampling. In *43rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, page 1274, 2002.

M. H. Beale, M. T. Hagan, and H. B. Demuth. *Neural Network Toolbox ™ User's Guide R2017b*. 2017. ISBN 0971732108.

P. S. Beran, D. Bryson, A. S. Thelen, M. Diez, and A. Serani. Comparison of multi-fidelity approaches for military vehicle design. In *AIAA AVIATION 2020 FORUM*, page 3158, 2020.

J. Biehler, M. W. Gee, and W. A. Wall. Towards efficient uncertainty quantification in complex and large-scale biomechanical problems based on a Bayesian multi-fidelity scheme. *Biomechanics and Modeling in Mechanobiology*, 14(3):489–513, 2015. ISSN 16177940. doi: 10.1007/s10237-014-0618-0. URL http://dx.doi.org/10.1007/s10237-014-0618-0.

D. E. Bryson and M. P. Rumpfkeil. Variable-fidelity surrogate modeling of lambda wing transonic aerodynamic performance. *54th AIAA Aerospace Sciences Meeting*, (January):1–19, 2016a. doi: 10.2514/6.2016-0294.

D. E. Bryson and M. P. Rumpfkeil. Variable-fidelity surrogate modeling of lambda wing transonic aerodynamic performance. In *54th AIAA Aerospace Sciences Meeting*, page 294, 2016b.

D. L. Clark, H. R. Bae, K. Gobal, and R. Penmetsa. Engineering design exploration using locally optimized covariance kriging. *AIAA Journal*, 54(10):3160–3175, 2016. ISSN 00011452. doi: 10.2514/1.J054860.

D. D. Cox, J. S. Park, and C. E. Singer. A statistical method for tuning a computer code to a data base. *Computational Statistics and Data Analysis*, 37(1):77–92, 2001. ISSN 01679473. doi: 10.1016/S0167-9473(00)00057-8.

C. Currin, T. Mitchell, M. Morris, and D. Ylvisaker. Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association*, 86(416):953–963, 1991. ISSN 1537274X. doi: 10.1080/01621459.1991.10475138.

S. De, J. Britton, M. Reynolds, R. Skinner, K. Jansen, and A. Doostan. On transfer learning of neural networks using bi-fidelity data for uncertainty propagation. *International Journal for Uncertainty Quantification*, 10(6), 2020.

J. Duchon. Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In *Constructive Theory of Functions of Several Variables: Proceedings of a Conference Held at Oberwolfach April 25–May 1, 1976*, pages 85–100. Springer, 1977. ISBN 3540080694.

J. Eason and S. Cremaschi. Adaptive sequential sampling for surrogate model generation with artificial neural networks. *Computers and Chemical Engineering*, 68:220–232, 2014. ISSN 00981354. doi: 10.1016/j.compchemeng.2014.05.021. URL http://dx.doi.org/10.1016/j.compchemeng.2014.05.021.

M. Fernández-Godino, C. Park, N. H. Kim, and R. T. Haftka. Issues in deciding whether to use multifidelity surrogates. *AIAA Journal*, 57(5):2039–2054, 2019. ISSN 00011452. doi: 10.2514/1.J057750.

M. G. Fernández-Godino, C. Park, N.-H. Kim, and R. T. Haftka. Review of multi-fidelity models. 2016. URL http://arxiv.org/abs/1609.07196.

S. Ficini, U. Iemma, R. Pellegrini, A. Serani, and M. Diez. Assessing the Performance of an Adaptive Multi-Fidelity Gaussian Process with Noisy Training Data: A Statistical Analysis.

AIAA Aviation and Aeronautics Forum and Exposition, AIAA AVIATION Forum 2021, pages 1–12, 2021. doi: 10.2514/6.2021-3098.

C. C. Fischer, R. V. Grandhi, and P. S. Beran. Bayesian-enhanced low-fidelity correction approach to multifidelity aerospace design. *AIAA Journal*, 56(8):3295–3306, 2018. ISSN 00011452. doi: 10.2514/1.J056529.

R. Fletcher. *Practical methods of optimization.* John Wiley & Sons, 2013.

A. I. Forrester and A. J. Keane. Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences*, 45(1-3):50–79, 2009. ISSN 03760421. doi: 10.1016/j.paerosci.2008.11.001.

A. I. Forrester, A. Sóbester, and A. J. Keane. Multi-fidelity optimization via surrogate modelling. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 463 (2088):3251–3269, 2007. ISSN 14712946. doi: 10.1098/rspa.2007.1900.

S. E. Gano, J. E. Renaud, and B. Sanders. Hybrid variable fidelity optimization by using a kriging-based scaling function. *AIAA Journal*, 43(11):2422–2430, 2005. ISSN 00011452. doi: 10.2514/1.12466.

R. B. Gramacy and H. K. Lee. Adaptive design and analysis of supercomputer experiments. *Technometrics*, 51(2):130–145, 2009. ISSN 00401706. doi: 10.1198/TECH.2009.0015.

J. Guenther, H. K. Lee, and G. A. Gray. Sequential design for achieving estimated accuracy of global sensitivities. *Applied Stochastic Models in Business and Industry*, 31(6):782–800, 2015. ISSN 15264025. doi: 10.1002/asmb.2091.

H. M. Gutmann. A Radial Basis Function Method for Global Optimization. *Journal of Global Optimization*, 19(3):201–227, 2001. ISSN 09255001. doi: 10.1023/A:1011255519438.

Z. H. Han, S. Görtz, and R. Hain. A Variable-Fidelity Modeling Method for Aero-Loads Prediction. *New results in numerical and experimental fluid mechanics*, (7):17–25, 2010.

J. K. Hartman. Some experiments in global optimization. *Naval Research Logistics Quarterly*, 20(3):569–576, 1973. ISSN 00281441. doi: 10.1002/nav.3800200316.

G. Heath. Mathworks Community Profile - Greg Heath, 2020. URL https://uk.mathworks.com/matlabcentral/profile/authors/2929937-greg-heath.

A. A. Howard, M. Perego, G. E. Karniadakis, and P. Stinis. Multifidelity deep operator networks for data-driven and physics-informed problems. *Journal of Computational Physics*, 493: 112462, 2023.

D. Huang, T. T. Allen, W. I. Notz, and R. A. Miller. Sequential kriging optimization using multiple-fidelity evaluations. *Structural and Multidisciplinary Optimization*, 32(5):369–382, 2006. ISSN 1615147X. doi: 10.1007/s00158-005-0587-0.

M. J. Jansen. Analysis of variance designs for model output. *Computer Physics Communications*, 117(1):35–43, 1999. ISSN 00104655. doi: 10.1016/S0010-4655(98)00154-4.

A. Keane and P. Nair. *Computational approaches for aerospace design: the pursuit of excellence*. John Wiley & Sons, 2005. ISBN 0470855479.

A. J. Keane. Cokriging for robust design optimization. *AIAA Journal*, 50(11):2351–2364, 2012. ISSN 00011452. doi: 10.2514/1.J051391.

M. C. Kennedy and A. O'Hagan. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87(1):1–13, 2000. ISSN 00063444. doi: 10.1093/biomet/87.1.1.

M. C. Kennedy and A. O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001a. ISSN 1369-7412. doi: 10.1111/1467-9868.00294.

M. C. Kennedy and A. O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001b. ISSN 1369-7412. doi: 10.1111/1467-9868.00294.

B. Kim, Y. Lee, and D. H. Choi. Construction of the radial basis function based on a sequential sampling approach using cross-validation. *Journal of Mechanical Science and Technology*, 23 (12):3357–3365, 2009. ISSN 1738494X. doi: 10.1007/s12206-009-1014-z.

S. Koziel and J. W. Bandler. Space-mapping-based modeling utilizing parameter extraction with variable weight coefficients and a data base. *IEEE MTT-S International Microwave Symposium Digest*, 1(2):1763–1766, 2006. ISSN 0149645X. doi: 10.1109/MWSYM.2006.249380.

S. Koziel and J. W. Bandler. Microwave device modeling using space-mapping and radial basis functions. *IEEE MTT-S International Microwave Symposium Digest*, 3:799–802, 2007a. ISSN 0149645X. doi: 10.1109/MWSYM.2007.380079.

S. Koziel and J. W. Bandler. A Space-Mapping Approach to Microwave Device Modeling Exploiting Fuzzy Systems. *IEEE Transactions on Microwave Theory and Techniques*, 55(12): 2539–2547, 2007b. ISSN 15579670. doi: 10.1109/TMTT.2007.909605.

S. Koziel and L. Leifsson. Knowledge-based airfoil shape optimization using space mapping. In *30th AIAA applied aerodynamics conference*, page 3016, 2012.

S. Koziel and L. Leifsson. Shape-preserving response prediction for engineering design optimization. *Procedia Computer Science*, 18(December):879–888, 2013. ISSN 18770509. doi: 10.1016/j.procs.2013.05.253. URL http://dx.doi.org/10.1016/j.procs.2013.05.253.

S. Koziel and L. Leifsson. Shape-Preserving Response Prediction for Surrogate Modeling and Engineering Design Optimization. *Springer Proceedings in Mathematics and Statistics*, 97: 241–253, 2014.

S. Koziel and L. Leifsson. *Simulation-driven design by knowledge-based response correction techniques*. 2016. ISBN 9783319301150. doi: 10.1007/978-3-319-30115-0.

S. Koziel and X.-S. Yang. *Computational Optimization, Methods and Algorithms*. 2011. ISBN 9783642175565.

S. Koziel, J. W. Bandler, and K. Madsen. A space-mapping framework for engineering optimization - Theory and implementation. *IEEE Transactions on Microwave Theory and Techniques*, 54(10):3721–3730, 2006a. ISSN 00189480. doi: 10.1109/TMTT.2006.882894.

S. Koziel, J. W. Bandler, and K. Madsen. Theoretical justification of space-mapping-based modeling utilizing a database and on-demand parameter extraction. *IEEE Transactions on Microwave Theory and Techniques*, 54(12):4316–4321, 2006b. ISSN 00189480. doi: 10.1109/TMTT.2006.884648.

S. Koziel, Q. Cheng, and J. W. Bandler. Space mapping. *IEEE Microwave Magazine*, (December): 105–122, 2008.

S. Koziel, S. Ogurtsov, and S. Szczepanski. Rapid antenna design optimization using shape-preserving response prediction. *Bulletin of the Polish Academy of Sciences: Technical Sciences*, 60(1):143–149, 2012. ISSN 02397528. doi: 10.2478/v10175-012-0020-y.

S. J. Leary, A. Bhaskar, and A. J. Keane. A knowledge-based approach to response surface modelling in multifidelity optimization. *Journal of Global Optimization*, 26(3):297–319, 2003. ISSN 09255001. doi: 10.1023/A:1023283917997.

D. Liu and Y. Wang. Multi-fidelity physics-constrained neural network and its application in materials modeling. *Journal of Mechanical Design*, 141(12):121403, 2019.

H. Liu, Y. S. Ong, and J. Cai. A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design, 2018a. ISSN 16151488.

H. Liu, Y. S. Ong, and J. Cai. A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design. *Structural and Multidisciplinary Optimization*, 57(1):393–416, 2018b. ISSN 16151488. doi: 10.1007/s00158-017-1739-8.

K. Madsen, J. Bandler, S. Dakroury, M. Bakr, J. Sondergaard, Q. Cheng, and A. Mohamed. Space Mapping: The State of the Art. *IEEE Transactions on Microwave Theory and Techniques*, 52 (1):337–361, 2004. ISSN 0018-9480. doi: 10.1109/tmtt.2003.820904.

L. Mainini, A. Serani, M. P. Rumpfkeil, E. Minisci, D. Quagliarella, H. Pehlivan, S. Yildiz, S. Ficini, R. Pellegrini, F. Di Fiore, D. Bryson, M. Nikbay, M. Diez, and P. Beran. Analytical Benchmark Problems for Multifidelity Optimization Methods. pages 1–14, 2022. URL http://arxiv.org/abs/2204.07867.

X. Meng and G. E. Karniadakis. A composite neural network that learns from multi-fidelity data: Application to function approximation and inverse pde problems. *Journal of Computational Physics*, 401:109020, 2020.

T. Mengistu and W. Ghaly. Aerodynamic optimization of turbomachinery blades using evolutionary methods and ANN-based surrogate models. *Optimization and Engineering*, 9(3): 239–255, 2008. ISSN 13894420. doi: 10.1007/s11081-007-9031-1.

E. Minisci and M. Vasile. Robust Design of a Re-entry Unmanned Space Vehicle by Multi-Fidelity Evolution Control. 51:4–6, 2013.

M. D. Morris, T. J. Mitchell, and D. Ylvisaker. Bayesian design and analysis of computer experiments: Use of derivatives in surface prediction. *Technometrics*, 35(3):243–255, 1993. ISSN 15372723. doi: 10.1080/00401706.1993.10485320.

R. H. Myers and D. C. Montgomery. Response Surface Methodology: Process and Product in Optimization Using Designed Experiments. 1995. ISSN 0471581003.

L. W. Ng and K. E. Willcox. Multifidelity approaches for optimization under uncertainty. *International Journal for Numerical Methods in Engineering*, (February):1102–1119, 2014. doi: 10.1002/nme. URL http://onlinelibrary.wiley.com/doi/10.1002/nme.3279/full.

R. M. Paiva, A. R. Carvalho, C. Crawford, and A. Suleman. Comparison of surrogate models in a multidisciplinary optimization framework for wing design. *AIAA Journal*, 48(5):995–1006, 2010. ISSN 00011452. doi: 10.2514/1.45790.

K. Palmer and K. Tsui. A minimum bias Latin hypercube design. *IIE Transactions (Institute of Industrial Engineers)*, 33(9):793–808, 2001. ISSN 0740817X. doi: 10.1080/07408170108936873.

C. Park, R. T. Haftka, and N. H. Kim. Remarks on multi-fidelity surrogates. *Structural and Multidisciplinary Optimization*, 55(3):1029–1050, 2017. ISSN 16151488. doi: 10.1007/s00158-016-1550-y. URL http://dx.doi.org/10.1007/s00158-016-1550-y.

C. Park, R. T. Haftka, and N. H. Kim. Low-fidelity scale factor improves Bayesian multi-fidelity prediction by reducing bumpiness of discrepancy function. *Structural and Multidisciplinary Optimization*, 58(2):399–414, 2018. ISSN 16151488. doi: 10.1007/s00158-018-2031-2.

J. S. Park. Tuning complex computer codes to data and optimal designs. 1991.

B. Parsonage and C. Maddock. A multi-fidelity model management framework for multi-objective aerospace design optimisation. *Frontiers in Aerospace Engineering*, 2(February): 1–21, 2023. doi: 10.3389/fpace.2023.1046177.

B. Parsonage and C. A. Maddock. Multi-stage Multi-fidelity Information Correction for Artificial Neural Network Based Meta-modelling. *2020 IEEE Symposium Series on Computational Intelligence, SSCI 2020*, pages 950–957, 2020. doi: 10.1109/SSCI47803.2020.9308255.

B. Peherstorfer, T. Cui, Y. Marzouk, and K. Willcox. Multi-fidelity Monte Carlo: importance sampling method using high-fidelity and surrogate model. 7831:1–23, 2015.

B. Peherstorfer, K. Willcox, and M. Gunzburger. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *SIAM Review*, 60(3):550–591, 2018. ISSN 00361445. doi: 10.1137/16M1082469.

M. Penwarden, S. Zhe, A. Narayan, and R. M. Kirby. Multifidelity modeling for physics-informed neural networks (pinns). *Journal of Computational Physics*, 451:110844, 2022.

P. Perdikaris, D. Venturi, J. O. Royset, and G. E. Karniadakis. Multi-fidelity modelling via recursive co-kriging and Gaussian-Markov random fields. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2179), 2015. ISSN 14712946. doi: 10.1098/rspa.2015.0018.

P. Z. Qian and C. F. Wu. Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments. *Technometrics*, 50(2):192–204, 2008. ISSN 00401706. doi: 10.1198/004017008000000082.

F. Regazzoni, S. Pagani, A. Cosenza, A. Lombardi, and A. Quarteroni. A physics-informed multi-fidelity approach for the estimation of differential equations parameters in low-data or large-noise regimes. *Rendiconti Lincei*, 32(3):437–470, 2021.

J. Roshanian, A. A. Bataleblu, and M. Ebrahimi. A novel metamodel management strategy for robust trajectory design of an expendable launch vehicle. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, 0(0):1–18, 2019. ISSN 20413025. doi: 10.1177/0954410019860638.

D. E. Rumelhart, B. Widrow, and M. A. Lehr. The Basic Ideas in Neural Networks. *Communications of the ACM*, 37(3):87–92, 1994. ISSN 15577317. doi: 10.1145/175247.175256.

J. Sacks, W. J. Welch, T. J. M. And, and H. P. Wynn. Design and analysis of computer experiments. *Statistical SCience*, 4(4):409–423, 1989. ISSN 0002-6018. doi: 10.1002/9781118842003.ch9.

A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259–270, 2010. ISSN 00104655. doi: 10.1016/j.cpc.2009.09.018. URL http://dx.doi.org/10.1016/j.cpc.2009.09.018.

M. Santos, B. Mattos, and R. Girardi. Aerodynamic coefficient prediction of airfoils using neural networks. In *46th AIAA aerospace sciences meeting and exhibit*, page 887, 2008.

M. J. Sasena. Flexibility and Efficiency Enhancements for Constrained Global Design Optimization with Kriging Approximations. *University of Michigan*, pages 1–237, 2002. doi: 10.1.1.2.4697. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.2.4697.

H. S. Seung, M. Opper, and H. Sompolinsky. Query by committee. In *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory*, pages 287–294, 1992. ISBN 089791497X. doi: 10.1145/130385.130417.

M. Shi, L. Lv, W. Sun, and X. Song. A multi-fidelity surrogate model based on support vector regression. *Structural and Multidisciplinary Optimization*, 2020. ISSN 16151488. doi: 10.1007/s00158-020-02522-6.

T. W. Simpson. Design and Analysis of Computer Experiments in Multidisciplinary Design Optimization: A Review of How Far We Have Come - or Not. *AIAA*, (September):1–21, 2008. doi: doi:10.2514/6.2008-5802.

D. H. Song and D. M. Tartakovsky. Transfer learning on multifidelity data. *Journal of Machine Learning for Modeling and Computing*, 3(1), 2022.

D. J. Toal. Some considerations regarding the use of multi-fidelity Kriging in the construction of surrogate models. *Structural and Multidisciplinary Optimization*, 51(6):1223–1245, 2015. ISSN 16151488. doi: 10.1007/s00158-014-1209-5.

J. D. Toscano, V. Oommen, A. J. Varghese, Z. Zou, N. Ahmadi Daryakenari, C. Wu, and G. E. Karniadakis. From pinns to pikans: Recent advances in physics-informed machine learning. *Machine Learning for Computational Science and Engineering*, 1(1):1–43, 2025.

M. Tyan and J. W. Lee. Efficient multi-response adaptive sampling algorithm for construction of variable-fidelity aerodynamic tables. *Chinese Journal of Aeronautics*, 32(3):547–558, 2019. ISSN 10009361. doi: 10.1016/j.cja.2018.12.012. URL https://doi.org/10.1016/j.cja.2018.12.012.

F. A. Viana and R. T. Haftka. Using multiple surrogates for minimization of the RMS error in meta-modeling. In *Proceedings of the ASME Design Engineering Technical Conference*, volume 1, pages 851–860, 2008. ISBN 9780791843253. doi: 10.1115/DETC2008-49240.

H. Wang, Y. Jin, and J. Doherty. A Generic Test Suite for Evolutionary Multifidelity Optimization. *IEEE Transactions on Evolutionary Computation*, 22(6):836–850, 2018. ISSN 1089778X. doi: 10.1109/TEVC.2017.2758360.

Y. Wang and C.-Y. Lai. Multi-stage neural networks: Function approximator of machine precision. *Journal of Computational Physics*, 504:112865, 2024.

S. Xiong, P. Z. Qian, and C. F. Wu. Sequential design and analysis of high-accuracy and low-accuracy computer codes. *Technometrics*, 55(1):37–46, 2013. ISSN 00401706. doi: 10.1080/00401706.2012.723572.

K. Q. Ye. Orthogonal Column Latin Hypercubes and Their Application in Computer Experiments. *Journal of the American Statistical Association*, 93(444):1430–1439, 1998. ISSN 1537274X. doi: 10.1080/01621459.1998.10473803.

R. Yondo, E. Andrés, and E. Valero. A review on design of experiments and surrogate models in aircraft real-time and many-query aerodynamic analyses. *Progress in Aerospace Sciences*, 96(November 2017):23–61, 2018. ISSN 03760421. doi: 10.1016/j.paerosci.2017.11.003.

L. Zhang, Y. Wu, P. Jiang, S. K. Choi, and Q. Zhou. A multi-fidelity surrogate modeling approach for incorporating multiple non-hierarchical low-fidelity data. *Advanced Engineering Informatics*, 51(September 2021):101430, 2022. ISSN 14740346. doi: 10.1016/j.aei.2021.101430. URL https://doi.org/10.1016/j.aei.2021.101430.

J. Zheng, X. Shao, L. Gao, P. Jiang, and Z. Li. A hybrid variable-fidelity global approximation modelling method combining tuned radial basis function base and kriging correction. *Journal of Engineering Design*, 24(8):604–622, 2013. ISSN 09544828. doi: 10.1080/09544828.2013.788135.

Q. Zhou, X. Shao, P. Jiang, Z. Gao, H. Zhou, and L. Shu. An active learning variable-fidelity metamodelling approach based on ensemble of metamodels and objective-oriented sequential sampling. *Journal of Engineering Design*, 27(4-6):205–231, 2016. ISSN 14661837. doi: 10.1080/09544828.2015.1135236.

Q. Zhou, Y. Wang, S. K. Choi, P. Jiang, X. Shao, and J. Hu. A sequential multi-fidelity metamodeling approach for data regression. *Knowledge-Based Systems*, 134:199–212, 2017. ISSN 09507051. doi: 10.1016/j.knosys.2017.07.033. URL http://dx.doi.org/10.1016/j.knosys.2017.07.033.

Q. Zhou, Y. Wang, S. K. Choi, P. Jiang, X. Shao, J. Hu, and L. Shu. A robust optimization approach based on multi-fidelity metamodel. *Structural and Multidisciplinary Optimization*, 57(2):775–797, 2018. ISSN 16151488. doi: 10.1007/s00158-017-1783-4.

Q. Zhou, Y. Wu, Z. Guo, J. Hu, and P. Jin. A generalized hierarchical co-Kriging model for multi-fidelity data fusion. *Structural and Multidisciplinary Optimization*, 62(4):1885–1904, 2020. ISSN 16151488. doi: 10.1007/s00158-020-02583-7.

Engineering Test Case: Transatmospheric Vehicle and Mission Design 4

4.1 Multi-Fidelity Modelling . . . 195
4.1.1 Geometry Parametrisation . . . 195
4.1.2 Aerodynamics . . . 196
4.1.3 2D/3D Geometry Meshing . . . 198
4.2 Multi-Fidelity Data Correction . . . 199
4.3 Optimisation . . . 202
4.3.1 System Models . . . 203
4.3.2 Problem Formulation . . . 205
4.3.3 Numerical Settings . . . 207
4.4 Results . . . 208
4.4.1 High/Low-Fidelity Solutions . . . 208
4.4.2 Multi-Fidelity Solutions . . . 210
4.4.3 K-fold Cross-Validation Error . . . 213
4.5 Cost . . . 214
4.6 Conclusion . . . 218
4.7 Bibliography . . . 220

In this chapter, the complete multi-fidelity data correction, surrogate model generation and *Multi-Fidelity Model Management* (MFMM) developed in Chapter 3 is applied to an engineering test case involving a parameterised *Unmanned Space Vehicle* (USV). The contents

of this chapter were largely reported in (Parsonage and Maddock, 2023). This example is intended to demonstrate the efficacy of the proposed framework for efficiently incorporating increased levels of simulation fidelity from multiple, competing sources into an early-stage *Multidisciplinary Design Optimisation* (MDO) scenario.

USVs, such as the Boeing X-37B *Orbital Test Vehicle* (OTV), have been used to return experiments from *Low Earth Orbit* (LEO) and are often considered as a test bed for the development of future re-usable launchers, with particular emphasis on the key technologies enabling atmospheric re-entry. The USV examined in this study is based on an idealised unpowered waverider (Starkey and Lewis, 2000; Minisci and Vasile, 2013), with the goal of defining the technical specifications for a small, relatively inexpensive technological demonstrator vehicle. Of particular interest is the range of geometric configurations (and corresponding re-entry trajectories) enabling the return of certain types of payload (e.g., delicate or time sensitive) owing to the management of aerodynamic and/or thermal loads. The practicality of such a vehicle is thus inherently dependant on the interaction between geometric configuration and aerodynamic/aerothermal subsystems. Simulating the latter to facilitate numerical design optimisation is traditionally more heavily limited by available computational resource. This problem, therefore, is well suited for examining via a multi-fidelity approach.

In Minisci and Vasile (2013), the authors construct and adaptively update an ANN surrogate of the vehicle's aerodynamic response within a model management framework. The surrogate is initially trained using only low-fidelity data obtained via a physics-based aerodynamic simulation tool. The outer-loop *Multi-Objective Optimisation* (MOO) algorithm then cycles through successive generations utilising the current surrogates to evaluate optimal trajectories. Simultaneously, an external model management process actively monitors the status of the current surrogates and schedules new evaluations from high-fidelity aerodynamic solver. New datapoints are selected based on the density of data within the vicinity of the current set of optimal trajectories. When new data is available, it is added to the training database with all lower fidelity information within that vicinity removed from subsequent training. All surrogates are then updated with the new database.

The approach presented here differs in that low-fidelity data, rather than being removed as new high-fidelity data becomes available, is instead gradually corrected using the new information, allowing a smoother progression towards global accuracy in the vicinity of the current solutions. The correction process is decomposed into several distinct parametric/non-parametric stages (see Section 3.1.2), each leveraging alternate aspects of the available model responses. Globally approximating surrogates are constructed at each degree of fidelity (low, mid, and high) via an automated hyper-parameter selection and training procedure (Section 3.1.3). The resulting hierarchy drives the optimisation process, with local refinement managed according to a confidence-based multi-response adaptive sampling procedure, with bias given

to global parameter sensitivities (see Section 3.2). The test vehicle under consideration is subjected to a static/dynamic parameter optimisation for three separate single-objective problems (Sections 4.3 and 4.4), representing relevant design conditions/criteria for high-level, early stage multidisciplinary design. It is found that the proposed data correction process facilitates increased efficiency in attaining a desired approximation accuracy relative to a single-fidelity equivalent model. When applied within the proposed multi-fidelity management framework, clear convergence to the objective optimum is observed for each examined design scenario, outperforming an equivalent single-fidelity approach in terms of computational efficiency and solution variability. Section 4.5 presents the associated cost savings, and finally, Section 4.6 presents the conclusions and recommendations. The work presented in the following chapter can be found in the following peer-reviewed publications:

- B. Parsonage and C. A. Maddock. Multi-stage Multi-fidelity Information Correction for Artificial Neural Network Based Meta-modelling. *2020 IEEE Symposium Series on Computational Intelligence, SSCI 2020*, pages 950–957, 2020. doi: 10.1109/SSCI47803.2020.9308255
- B. Parsonage and C. Maddock. A multi-fidelity model management framework for multi-objective aerospace design optimisation. *Frontiers in Aerospace Engineering*, 2 (February):1–21, 2023. doi: 10.3389/fpace.2023.1046177

4.1 Multi-Fidelity Modelling

This section details the modelling strategies employed to represent the vehicle geometry and aerodynamics, the latter via two alternative, competing levels of simulation fidelity.

4.1.1 Geometry Parametrisation

A waverider USV may be parameterised according to the 2D power law equations presented by (Starkey and Lewis, 2000). The curvatures of the planform p and the upper surface u are given by:

$$y_p = Ax^n \tag{4.1}$$

$$y_u = B(z_u)^n \rightarrow z_u = \left(\frac{y_u}{B}\right)^{\frac{1}{n}} \tag{4.2}$$

where the exponent $n \in [0, 1]$ and the positive scaling coefficients A and B are given by:

$$A = w/2L^n \tag{4.3}$$

$$B = A/\tan^n(\beta) \tag{4.4}$$

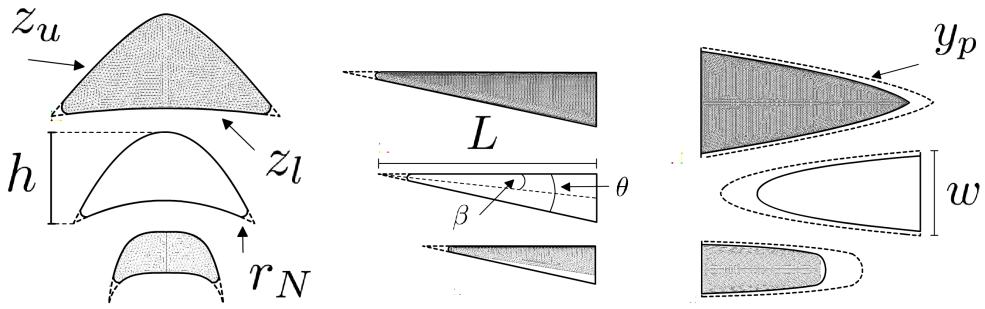


FIGURE 4.1: Example geometric configurations of a waverider-based USV defined by (4.1)-(4.8) (Starkey and Lewis, 2000). Orthogonal cross sections are shown aligning with the vehicle's **(A)** longitudinal x-axis, **(B)** lateral y-axis and **(C)** vertical Z-axis. The three examples included in each subfigure represent the lower, middle and upper values of the vehicle design vector $\mathbf{d}(l, w, n, \theta, r_N)$.

where L and w are the vehicle length and width respectively and β is the shock angle. The curvature of the lower surface with an attached shock is given by:

$$z_l = x \tan\theta + (y/A)^{1/n}(\tan\beta - \tan\theta) \tag{4.5}$$

where θ is the wedge angle. Vehicle height h and the surface areas of the planform and base sections S_p and S_b are given by:

$$h = L \tan\beta \tag{4.6}$$

$$S_p = wL/(n+1) \tag{4.7}$$

$$S_b = S_p \tan\theta \tag{4.8}$$

The geometric design vector $\mathbf{d}$ is defined according to the following variable bounds (Minisci and Vasile, 2013); vehicle length $L \in [2.9, 4.2]$m, width $w \in [1, 2]$m, exponent $n \in [0.2, 0.7]$, wedge angle $\theta_w \in [7, 11]$ deg and leading-edge radius $r_N \in [0.04, 0.06]$m. The shock angle β is held fixed at 12 deg. By parameterising in this manner, small variations in $\mathbf{d}$ can correspond to significant change in the resulting geometry (Starkey and Lewis, 2000). Three examples, representing the maximum, mean and minimum values of $\mathbf{d}$, are shown in Figure 4.1.

4.1.2 Aerodynamics

Two distinct levels of fidelity are considered for the evaluation of the aerodynamic forces acting on the vehicle during re-entry. Each is referred to respectively as 'high' and 'low' fidelity in primarily a relative sense, in that there exists a significant disparity across both numerical complexity and computational cost between the two. Given the purposes of this initial demonstration were more concerned with any inherent *disagreement* between alternative

models rather than their absolute accuracy, extended validation was considered outwith the scope of this study.

4.1.2.1 Low-fidelity aerodynamics

Low-fidelity aerodynamics are computed using the *Free Open Source Tool for Re-Entry of Asteroids and Debris* (FOSTRAD) under development at the University of Strathclyde (Mehta et al., 2015; Falchi et al., 2017b). FOSTRAD discretises solid geometry into surface meshes of triangular elements (panels), to which *Local Surface Inclination* (LSI) based methods are applied. Modified Newtonian Theory and the Schaaf and Chambre analytic model are used to approximate continuum and free-molecular aerodynamic coefficients respectively (Mehta et al., 2015; Falchi et al., 2017a). Transitional regime aerodynamics, defined in (Mehta et al., 2015) as $10^{-4} < Kn < 100$, are approximated using a sigmoid (base 10) bridging function.

4.1.2.2 High-fidelity aerodynamics

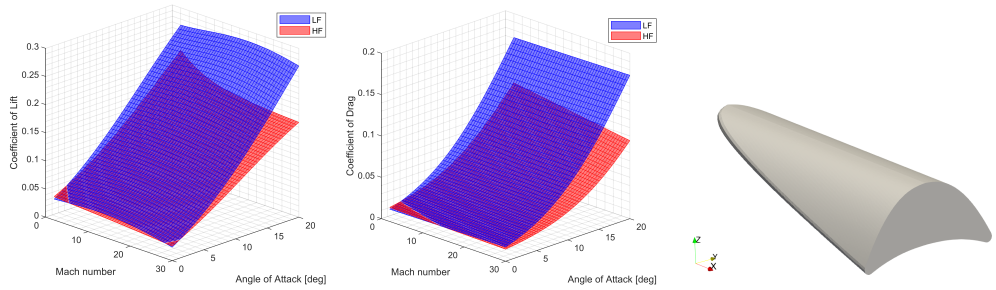
High-fidelity aerodynamics are computed using the *Stanford University Unstructured Code* (SU2) (Palacios et al., 2013). The 3-dimensional RANS equations are employed with a standard SST turbulence model and a *Finite Volume Method* (FVM) discretisation. The JST scheme is used in conjunction with a second-order scalar upwind discretisation and Venkatakrishnan's limiter to model convective fluxes. The gradients of the spatial flow variables, required to evaluate viscous fluxes, are calculated using the Green-Gauss method with a CFL number of 1. The linear system is solved using the GMRES method with an error tolerance of 10^{-10} and the steady state solution is achieved using the Euler implicit method for time integration. A Cauchy convergence criteria is applied to the Drag function after iteration 10. Convergence is defined as the reduction of the residuals by five orders of magnitude with respect to their initial values, within a maximum number of iterations fixed at 5,000.

4.1.2.3 Correlation

Table 4.1 summarises the general numerical correlation between both aerodynamics approximations, where the coefficients of lift and drag, C_L and C_D, of the vehicle defined by design vector $\mathbf{d}(L, w, n, \theta, r_N)$ (Section 4.1.1) are estimated for the operational range defined by Mach number Ma and angle of attack α:

$$\begin{aligned} 2 \leq Ma \leq 30 \\ 0^{\circ} \leq \alpha \leq 20^{\circ} \end{aligned} \tag{4.9}$$

Of particular note, the values of Pearson's correlation coefficient r^2 is noticeably higher relative to R^2. This suggests the two models share similar functional form but are more significantly displaced in output space. This is in line with expectation, as despite differing



(a) (b) (c)

FIGURE 4.2: Correlation between the coefficients of **(A)** lift C_L and **(B)** drag C_D, for **(C)** the vehicle defined by $\mathbf{d}(L, w, n, \theta, r_N) = [3.55, 1.5, 0.45, 9, 0.05]$ relative to Mach number Ma and Angle of Attack α.

levels of simulation complexity, both aim to represent the same physical process and thus an appreciable level of similarity would be expected.

Property	**Value** $[C_L, C_D]$
RRMSE	1.075, 1.331
RMAE	2.872, 3.074
R^2	43.45%, 42.00%
r^2	85.53%, 96.67%

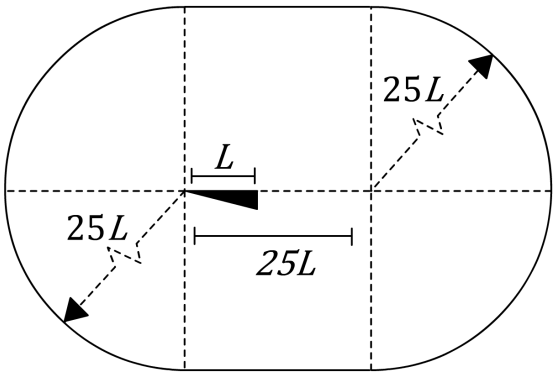
Table 4.1: Numerical accuracy of the low-fidelity model relative to the high-fidelity counterpart.

Figure 4.2 shows the response surface of the coefficients of lift and drag, C_L and C_D, for the vehicle configuration: $L = 3.55$m, $w = 1.5$m, $n = 0.45$, $\theta = 9$deg and $r_N = 0.05$m for the operational range given by (4.9). In this case, the low-fidelity model can be seen to generally over-predict the aerodynamic response, particularly at higher angles of attack. Furthermore, as the low-fidelity model includes no specific flow modelling, the functional form (predominantly of C_L) as Mach number increases is notably simpler than that of the high-fidelity model. Regardless, the general agreement in form is seen to be quite good, thus conforming to the overall metrics presented in Table 4.1.

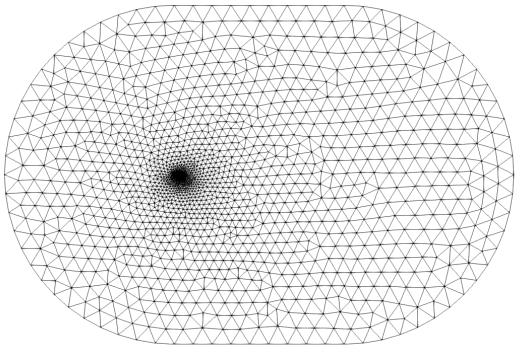
4.1.3 2D/3D Geometry Meshing

The open source meshing tool GMSH (v4.6.0) (Geuzaine and Remacle, 2009), incorporating the OpenCascade geometry kernel, is selected for mesh generation. Equations (4.1), (4.2) and (4.5) are used to define the base section and leading edge profile according to the user input vector **d**. A solid surface `.stl` mesh may then be generated and exported.

To produce a volume mesh suitable for CFD analyses, a half-cylindrical far-field with semi-spherical input/output caps is created around the surface mesh, with the y-symmetry plane



(a)



(b)

FIGURE 4.3: An example of the far-field volumetric mesh used for CFD (high-fidelity) simulation, where **(A)** gives the dimensions relative to vehicle length L (Leifsson et al., 2016) and **(B)** shows the resulting unstructured mesh, concentrated around the vehicle body surface.

bisecting the vehicle geometry along the longitudinal axis, see Figure 4.3. A single volume mesh is thus created via Boolean subtraction. A simplified boundary mesh definition is incorporated by specifying a vehicle-centered bounding box with internal and external mesh element sizes of 0.025m, and 10m respectively, with a transition gradient layer of thickness $t = 20$m. The vehicle surface mesh is comprised of between 3500 and 8000 vertices and between 7500 and 17000 triangular elements. Likewise, the CFD volume mesh of between 60,000 and 200,000 nodes and 400,000 and 1,400,000 elements. Input, output, far-field and vehicle surfaces are labelled accordingly. The complete volume mesh is exported in the native SU2 format.

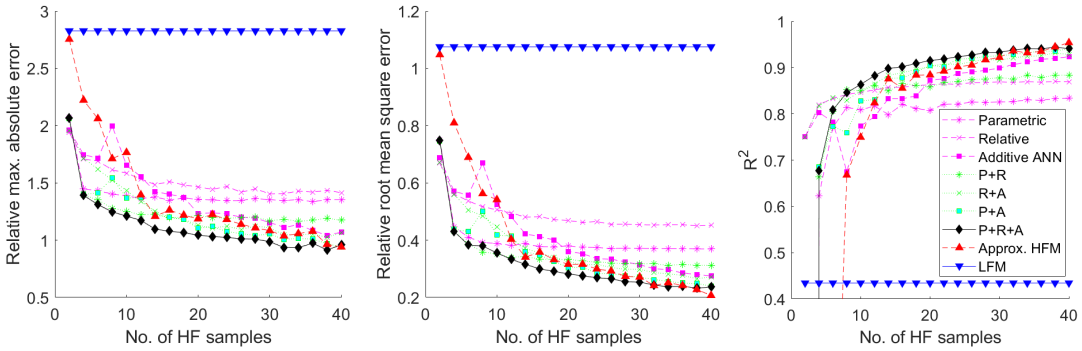
4.2 Multi-Fidelity Data Correction

In an earlier version of this work, Parsonage and Maddock (2020) presented an isolated test of the proposed response correction functions for a series of fixed geometric designs as well as the complete input domain as described in Section 4.1. This section presents a similar demonstration, though extends the range of high-fidelity sampling and focuses on more

comparative continuous metrics, including consideration of the variance exhibited by each approach. This section is included to further demonstrate the efficacy of the proposed multi-fidelity correction methods prior to the commencement of the engineering test case comprising the main subject of this chapter.

The coefficients of lift and drag, C_L and C_D, for the vehicle defined by design vector $\mathbf{d}(L, w, n, \theta, r_N)$ (Section 4.1.1) are predicted for the operational range defined by Mach number $2 \leq Ma \leq 30$, angle of attack $0^o \leq \alpha \leq 20^o$ and altitude $10^3 m \leq h \leq 10^5 m$. Sample sets of M low-fidelity points and N high-fidelity points are selected according to Section 3.1.1, where $M = 30 d_{in}$ (given that the cost of the low-fidelity model is relatively low) and N is increased in regular increments from 2 to 40. All samples sets are generated via LHS. For each value of N, the corrected predictions of C_L and C_D are evaluated against an independent testing set of M high-fidelity responses. As for the numerical tests presented in Section 3.1.4, several corrective process variants are considered, where the stages - Parametric (P), Relative (R) and Additive (A) - are applied individually and in sequential combinations. Figures 4.4 and 4.5 show the values of RMAE, RRMSE and R^2 across all values of N for both responses, averaged over 100 separate analyses. Included for reference are the corresponding values of the low-fidelity model, and of a single-fidelity surrogate model trained only on the set of N high-fidelity responses available at the current iteration. This model thus represents the baseline above which the multi-fidelity surrogate must perform.

Considering the responses of both C_D and C_L, a clear hierarchy can be observed between method variants. Within this hierarchy, the performance of each process largely conforms to the behaviour observed for the numerical tests presented previously in Section 3.1.4. For instance, the single stage methods exhibit significant limitations in attainable correction quality. The P-stage is limited by the simplistic nature of the defining parametric operation, and thus while initially offering a strong correction (due to the relatively high model correlation), the effectiveness of successive samples exponentially decreases. In a similar manner, the R-stage is limited by the inherent complexity of the low-fidelity model, and the A-stage demonstrates an increase in correction quality proportional to N, albeit remaining poor at lower values of N. Again conforming to expectation, applying stages in selected sequential combinations (PR, RA, and PA) is seen to alleviate some of these limitations. For example, preconditioning with the P-stage both improves the converged value of the R-stage and the correction quality/stability for the A-stage at lower values of N. In a similar manner, the RA method both reduces correction instability of the A-stage and the premature convergence of the R-stage. Of note here is that unlike the others, the R-stage does not explicitly aim to reduce sample point discrepancy, and thus is effective as a pre-conditioner to the A-stage. Furthermore, the sequential combination of the P-stage and A-stage is perhaps inefficient for this case, as both are driven by sample point discrepancy. Thus in any combination, the elimination of sample point error renders the



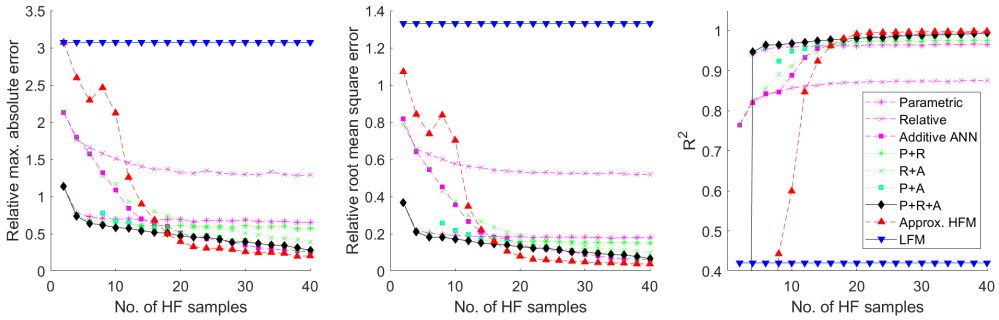
(a) (b) (c)

FIGURE 4.4: Performance metrics of the Parametric (P), Relative (R) and Additive (A) corrective stages applied individually and in sequential combinations for the USV Coefficient of Lift C_L relative to the number of high-fidelity evaluations, where **(A)** is the RMAE **(B)** is the RRMSE and **(C)** is the Coefficient of Determination R^2. Included for reference are the error values of the low-fidelity model, and a single-fidelity surrogate trained only on the available high-fidelity samples at each iteration.

other obsolete, largely undesirable when the true response is not yet fully represented. The most appropriate sequence thus includes all 3 stages, arranged PRA, which can indeed be seen to offer the strongest correction quality for both cases.

As before, the single-fidelity surrogate offers the poorest prediction at lower values of N, but tends to exceed the multi-fidleity approaches above some critical number of high-fidelity samples N_{crit}. The value of N_{crit} can be seen in this case to be around 38 and 18 samples for C_L and C_D respectively. The significant difference between responses is perhaps surprising given the similar level of correlation between models (see Section 4.1.2.3).

The variance in corrective quality was additionally examined as a further extension to the work presented in (Parsonage and Maddock, 2020). Figures 4.6 and 4.7 show the variance in RRMSE observed for each correction stage configuration for C_L and C_D respectively. Increased variance is observed for each method utilising a neural network (i.e. non-deterministic) stage. Whilst the collaborative effects of each method on prediction quality variance seem to follow the same general behaviour as previously discussed, of particular interest is that even after N_{crit} has been reached, there still exists notable variation in the prediction quality of the single fidelity surrogate, likely due to the combination of limited available samples and a lack of any deterministic pre-processing stage(s).



(a) (b) (c)

FIGURE 4.5: Performance metrics of the Parametric (P), Relative (R) and Additive (A) corrective stages applied individually and in sequential combinations for the USV Coefficient of Drag C_D relative to the number of high-fidelity evaluations, where **(A)** is the RMAE **(B)** is the RRMSE and **(C)** is the Coefficient of Determination R^2. Included for reference are the error values of the low-fidelity model, and a single-fidelity surrogate trained only on the available high-fidelity samples at each iteration.

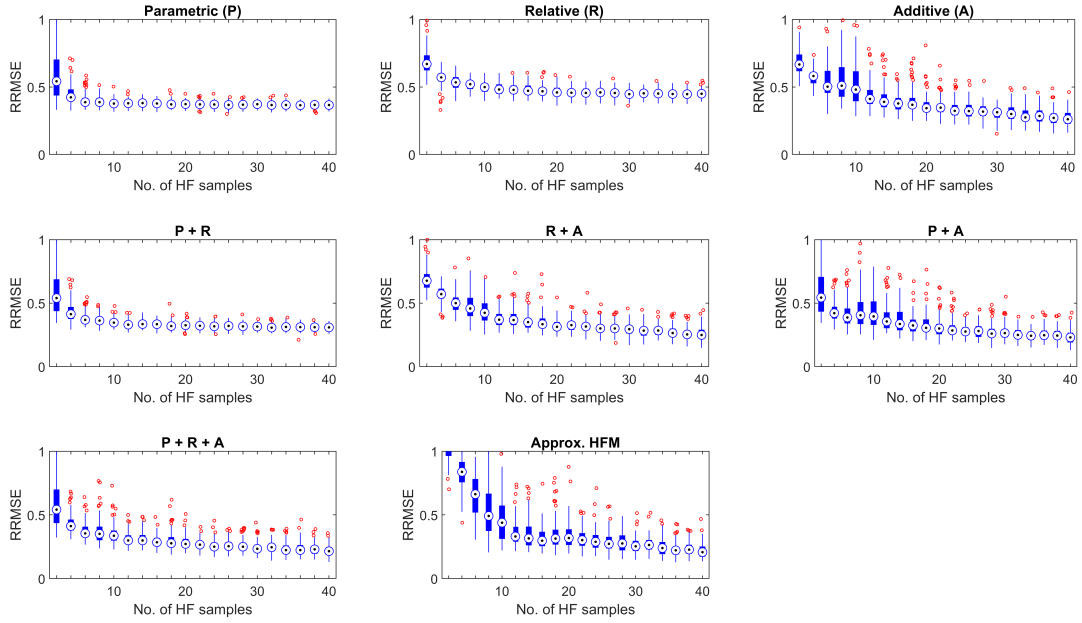


FIGURE 4.6: Variance in RRMSE of the Parametric (P), Relative (R) and Additive (A) corrective stages applied individually and in sequential combinations for the USV Coefficient of Drag C_L relative to the number of high-fidelity evaluations. Included for reference is the corresponding performance of the single-fidelity surrogate trained only on the available high-fidelity samples at each iteration.

4.3 Optimisation

This section briefly describes the system dependant processes, problem formulation and settings for the numerical solver.

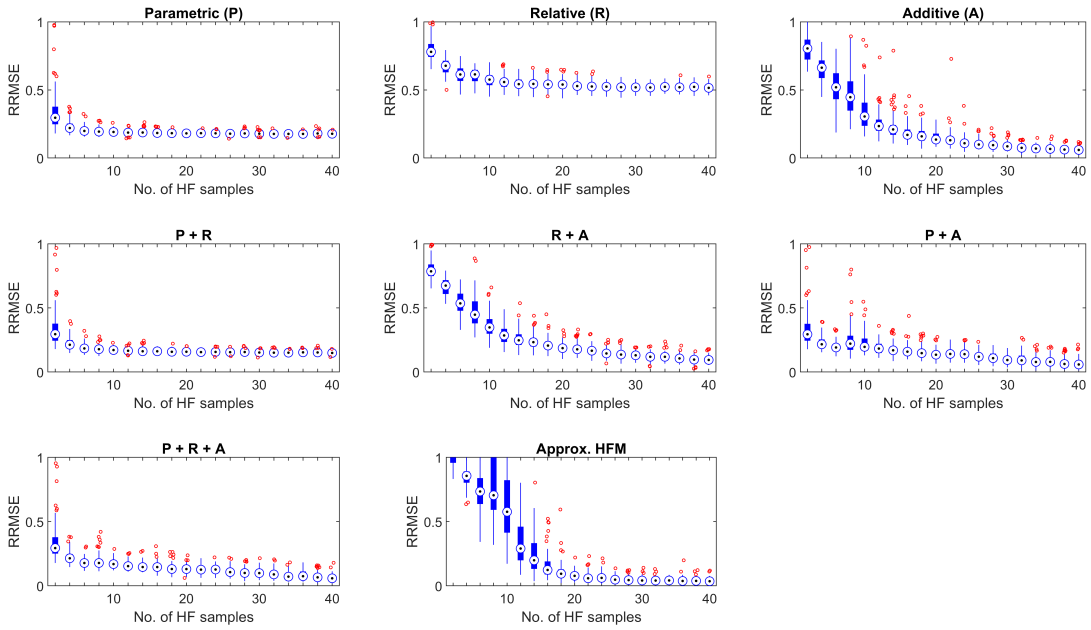


FIGURE 4.7: Variance in RRMSE of the Parametric (P), Relative (R) and Additive (A) corrective stages applied individually and in sequential combinations for the USV Coefficient of Drag C_D relative to the number of high-fidelity evaluations. Included for reference is the corresponding performance of the single-fidelity surrogate trained only on the available high-fidelity samples at each iteration.

4.3.1 System Models

The following system models are incorporated into the optimisation problem formulation, in addition to the geometric and aerodynamic model definitions given in Section 4.1.

4.3.1.1 Environment

The Earth is modelled as an oblate spheroid based on the WSG-84 model. The gravitational field is modelled using 4th order spherical harmonics (accounting for J2, J3 and J4 terms) for accelerations in the radial g_r and transverse g_t directions. Atmospheric conditions (temperature T_{atm}, pressure P_{atm}, density ρ_{atm} and speed of sound c) are modelled using the Standard US76 global static atmospheric model.

4.3.1.2 Configuration

Geometric characteristics of the vehicle are taken as presented in Section 4.1. The total mass of the vehicle is comprised of the structural mass $m_{structure}$, the mass of the thermal protection system m_{TPS} and the mass of the payload (e.g. avionics and power systems) $m_{payload}$:

$$m_{total} = m_{structure} + m_{TPS} + m_{payload} \tag{4.10}$$

The structural mass is given by:

$$m_{structure} = \rho_{body}(2S_p + S_b) * t_{shell} \tag{4.11}$$

where S_p and S_b are the surface areas of the planform and base sections respectively, and assuming a titanium frame of density $\rho_{body} = 4400kg/m^3$ and shell thickness t_{shell} of $5mm$. The mass of the TPS, chosen as $3mm$ thick Zirconium Diboride, $\rho_{TPS} = 6000kg/m^3$, is taken as:

$$m_{TPS} = m_{nose} + m_{skin} \tag{4.12}$$

where the mass of the nose m_{nose} is:

$$m_{nose} = \rho_{TPS} V_{nose} \tag{4.13}$$

and the mass of the TPS skin covering the remainder of the vehicle m_{skin}:

$$m_{skin} = \rho_{TPS} S_{TPS} t_{TPS} \tag{4.14}$$

where S_{TPS} is the surface area of the vehicle excluding the nose section, approximated by:

$$S_{TPS} = 2S_p + S_b - 2S_n \tag{4.15}$$

where S_n is the surface areas of the nose section. The payload mass $m_{payload}$ is assumed 40% of the structural mass:

$$m_{payload} = 0.4 m_{structure} \tag{4.16}$$

The vehicle 'nose' section length is defined as $0.1L$, with a corresponding volume V_n approximated as:

$$V_n = \frac{w_n L_n^2 \tan\theta}{(n+1)(n+2)} \tag{4.17}$$

4.3.1.3 Aerodynamics

Aerodynamic evaluation may be performed using either the low- or high-fidelity solvers introduced in Section 4.1.

4.3.1.4 Aerothermodynamics

For this simplified analysis, it is assumed that the peak heat flux exists at the vehicle nose section (Minisci et al., 2011). The convective heat flux is evaluated using the analytical formula as given by (Anderson Jr, 2006):

$$\dot{q}_{conv} = K_e \sqrt{\frac{\rho_\infty}{r_N}} v^3 \tag{4.18}$$

where $Ke = 1.742e^{-4}$ for $\dot{q}_{conv}[W/m^2]$, r_N is the radius of curvature of the nose leading edge, v is the vehicle velocity and ρ_{inf} is the free-stream (atmospheric) density.

4.3.1.5 Dynamics

A 3-DOF dynamic model expressed in an *Earth Centred Earth Fixed* (ECEF) reference frame is used. The state vector is comprised of the spherical coordinates for position $\mathbf{r} = [r, \lambda, \theta]$ and velocity $\mathbf{v} = [\nu, \gamma, \chi]$, where r is the radial distance, (λ, θ) are the latitude and longitude, ν is the magnitude of the relative velocity vector directed by the flight path angle γ and the flight heading angle χ.

$$\dot{r} = \nu \sin\gamma \tag{4.19}$$

$$\dot{\lambda} = \frac{\nu \cos\gamma \cos\chi}{r} \tag{4.20}$$

$$\dot{\theta} = \frac{\nu \cos\gamma \sin\chi}{r \cos\lambda} \tag{4.21}$$

$$\begin{aligned}\dot{\nu} = \frac{F_T(\tau)\cos\alpha - D(\alpha)}{m} - g_r \sin\gamma + g_t \cos\gamma \cos\chi \\ +\omega_E^2 r \cos\lambda(\sin\gamma\cos\lambda - \cos\gamma\cos\chi\sin\lambda)\end{aligned} \tag{4.22}$$

$$\begin{aligned}\dot{\gamma} = \frac{1}{\nu}\left(\frac{F_T(\tau)\sin\alpha + L(\alpha)}{m}\cos\beta - g_r\cos\gamma - g_t\sin\gamma\cos\chi\right) + \frac{\nu}{r}\cos\gamma \\ +\frac{\omega_E^2 r}{\nu}\cos\lambda(\sin\gamma\cos\chi\sin\lambda + \cos\gamma\cos\lambda) + 2\omega_E\sin\chi\cos\lambda\end{aligned} \tag{4.23}$$

$$\begin{aligned}\dot{\chi} = \frac{1}{\nu\cos\gamma}\left(\frac{F_T(\tau)\sin\alpha + L(\alpha)}{m}\sin\beta - g_t\sin\chi\right) + \frac{v}{r}\cos\gamma\sin\chi\tan\lambda \\ +\omega_E^2 r\frac{\sin\chi\sin\lambda\cos\lambda}{\nu\cos\gamma} + 2\omega_E(\sin\lambda - \tan\gamma\cos\chi\cos\lambda)\end{aligned} \tag{4.24}$$

$$\dot{\alpha} = \frac{\partial\alpha}{\partial t} \tag{4.25}$$

In this case, the thrust force $F_T = 0$ and angle $\tau = 0\,\text{deg}$. No out-of-plane movement is considered, thus bank angle $\beta = 0\,\text{deg}$.

4.3.2 Problem Formulation

A generic optimal control problem may be formulated as follows (Ricciardi et al., 2019):

$$\begin{aligned}&\min_{u\in U,\, b\in B}\mathbf{J} \\ s.t.\quad &\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x},\mathbf{u},\mathbf{b},t) \\ &\mathbf{g}(\mathbf{x},\mathbf{u},\mathbf{b},t) \le 0 \\ &\boldsymbol{\psi}(\mathbf{x}_0,\mathbf{x}_f,t_0,t_f,\mathbf{b}) \le 0 \\ &t \in [t_0,t_f]\end{aligned} \tag{4.26}$$

where $\mathbf{J} = [J_1,\ldots,J_m]^T$ is a vector of objectives J_i each a function of the state vector $\mathbf{x} : [t_0,t_f] \to \mathbb{R}^n$, control variables $\mathbf{u} \in L^\infty(U \subseteq \mathbb{R}^{n_u}$, static parameters $\mathbf{b} \in \mathbf{B} \subseteq \mathbb{R}^{n_b}$ and time t. The functions $\mathbf{x}(t)$ belong to the Sobolev space $W^{1,\infty}$ while the other objective

functions are $J_i : \mathbb{R}^{3n+2} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_b} \rightarrow \mathbb{R}$. The objective vector is subject to a set of dynamic constraints with $\mathbf{F} : \mathbb{R}^n \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_b} \times [t_0, t_f] \rightarrow \mathbb{R}^n$, algebraic constraints $\mathbf{g} : \mathbb{R}^n \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_b} \times [t_0, t_f] \rightarrow \mathbb{R}^{n_g}$ and boundary conditions $\boldsymbol{\psi} : \mathbb{R}^{2n+2} \times \mathbb{R}^{n_b} \rightarrow \mathbb{R}^{\psi}$.

The problem considered in this chapter has two major constituents. The first is a time-dependant problem considering the optimal trajectory of the vehicle given a finite set of controllable parameters. This constitutes a a standard OC problem as described later in Section 5.1. The second considers the optimisation of the set of time-independent (static) parameters defining the vehicle shape/configuration. This constitutes a more traditional engineering design problem. Table 4.2 lists the bounds, initial and final conditions of the primary problem variables. The time-dependant vehicle state is defined using a fixed initial condition and bounded final condition, all within global bounds. Similarly, the time-independent design variables and the single control variable are defined using global bounds and an initial guess at the midpoint for each element. The list of static variables includes the set of design variables $\mathbf{d}(l, w, n, \theta, r_N)$ as well as three slack variables $(\dot{q}_{conv})_{max}$, a_{max} and q_{max} defining the maximum acceptable values of the heat flux, acceleration and dynamic pressure respectively, such that $\mathbf{b} = [\mathbf{d}, (\dot{q}_{conv})_{max}, a_{max}, q_{max}]$. These variables are used to define additional path constraints such that the obtained trajectories are confined within these operational bounds:

$$\dot{q}_{conv}(\mathbf{x}, \mathbf{u}, \mathbf{b}, t) \leq (\dot{q}_{conv})_{max} \tag{4.27}$$

$$q(\mathbf{x}, \mathbf{u}, \mathbf{b}, t) \leq q_{max} \tag{4.28}$$

$$\mathbf{a}(\mathbf{x}, \mathbf{u}, \mathbf{b}, t)^{(1,3)} \leq a_{max} \tag{4.29}$$

Three problem objectives are considered, each requiring a distinct degree of aerodynamic performance; the minimisation of peak convective heat flux experienced during re-entry $\dot{q}_{conv}$, the minimisation of total time required to reach terminal conditions t_f, and the minimisation of the peak dynamic pressure q:

$$J_1(\mathbf{b}) = (\dot{q}_{conv})_{max} \tag{4.30}$$

$$J_2(t_f) = t_f \tag{4.31}$$

$$J_3(\mathbf{b}) = q_{max} \tag{4.32}$$

Each problem is subject to path constraints (Eqs. (4.27)-(4.29)) on convective heat flux $\dot{q}_{conv}$, axial and transverse acceleration $\mathbf{a}^{(1,3)} = [a_x, a_z]$ and dynamic pressure q. The time domain of the problem is defined within a global upper lower bound $t \in [0, 6500]$, containing a bounded final condition $t_f \in [500, 6500]$ and a fixed initial condition of $t_0 = 0$. The computational

resource required to converge in each case is examined using both the MFMM framework and an equivalent single-fidelity (SF) approach.

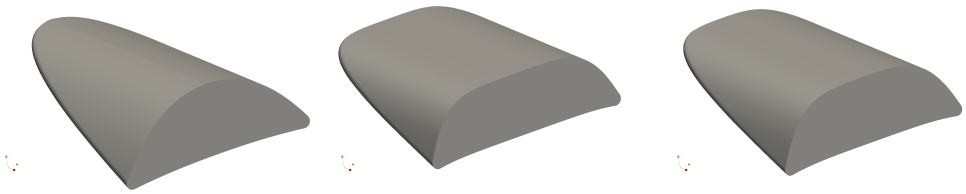
	Parameter	Bounds	Initial	Final
State variables	Altitude [m]	[15000, 100000]	100000	[15000, 25000]
	Velocity [m/s]	[100, 10000]	7700	[900, 1100]
	Flight path angle [deg]	[-80, 10]	-0.3	[-80, 10]
	Heading angle [deg]	[0, 225]	145	[0, 225]
	Latitude [deg]	[-90, 90]	68	[-90, 90]
	Longitude [deg]	[-180, 180]	21	[-180, 180]
	Angle of attack [deg]	[0, 20]	0	[0, 20]
Static variables	Length [m]	[2.9, 4.2]	3.5	–
	Width [m]	[1, 2]	1.5	–
	Exponent	[0.2, 0.7]	0.45	–
	Wedge angle [deg]	[7, 11]	9	–
	Nose radius [m]	[0.04, 0.06]	0.05	–
	Peak heat flux [W/cm^2]	[0, 500]	200	–
	Peak acceleration [m/s^2]	[0, 19.6]	19.6	–
	Peak dynamic pressure [kg/m^2]	[0, 20000]	20000	–
Control variables	Control rate $\partial\alpha/\partial t$ [deg/s]	[-0.1, 0.1]	0	–
Time	Total time [s]	[0, 6500]	0	[500,6500]

Table 4.2: Optimisation parameters.

The problem is solved using the in-house, open source *Multi-Objective Direct Hybrid Optimal Control* (MODHOC) (Ricciardi and Vasile, 2018) algorithm, incorporating the presented multi-fidelity management framework. MODHOC transcribes the optimisation problem into a NLP problem via the DFET method (Vasile, 2010b).

4.3.3 Numerical Settings

Each problem is considered as a single time phase, with boundary conditions given in Table 4.2 and discretised into 6 finite elements. States and controls are represented by element specific 9th order Bernstein polynomials, giving 496 total optimisation parameters (60 control, 427 state and 9 static). Single objective optimisation was completed using the gradient based solver fmincon with a *Sequential Quadratic Programming* (SQP) algorithm. Limits of 1000 objective function evaluations and 100 outer level iterations were imposed for each run. The maximum function, constraint violation and step tolerances are set at 1e-9, 1e-12 and 1e-12 respectively.



(a) (b) (c)

FIGURE 4.8: Geometric design solutions obtained using only high-fidelity aerodynamic model. The single objective in each case is the minimisation of **(A)** peak convective heat flux at the vehicle nose $(\dot{q}_{conv})_{max}$ **(B)** the total time taken to reach terminal conditions t_f and **(C)** the peak dynamic pressure experience during re-entry q_{max}.

4.4 Results

4.4.1 High/Low-Fidelity Solutions

As a reference, the problem was first solved for each objective using pre-constructed well-fit[1] surrogate approximations of the high-fidelity and low-fidelity aerodynamic models. The optimal designs for each single-objective problem, relative to fidelity level, are given in Table 4.3. Figures 4.8 and 4.9 show the geometry and state trajectories for each problem solved using the high-fidelity aerodynamic model.

Level	Objective	L [m]	w [m]	n	θ_n [deg]	r_N [m]	m [kg]	J
high-fidelity	min $(\dot{q}_{conv})_{max}$	3.3517	2.0	0.4506	11.0	0.06	515.10	1.79292e6 W/m^2
	min t_f	2.9	2.0	0.2	11.0	0.06	553.49	1215.4 s
	min q_{max}	3.0785	1.885	0.2507	10.9802	0.0564	528.84	16006.2 kg/m^2
low-fidelity	min $(\dot{q}_{conv})_{max}$	2.9	2.0	0.2	10.5720	0.06	549.79	1.36021e6 W/m^2
	min t_f	2.9	2.0	0.2	7.0	0.06	519.33	1724.65 s
	min q_{max}	2.9311	1.968	0.2144	7.4468	0.0598	513.56	16006.2 kg/m^2

Table 4.3: High/Low fidelity solution summary.

As a general theme, the optimiser favours shorter, wider vehicles with larger values of leading edge radius (the latter likely due to the significant effect on peak heating). The primary difference between solutions is then the resulting mass of each vehicle, driven by the shape exponent value n. This is possibly a result of an attempt to maximise effective lifting surface whilst incurring minimal mass penalty. The effects may be seen in the high-fidelity trajectory solutions, where the lighter vehicles utilise greater proportions of high altitude gliding to dissipate excess energy, ultimately allowing a gentler descent through the denser regions of the atmosphere. In contrast, the heavier solution (minimum t_f) exhibits a rapid descent through the upper regions of the atmosphere before utilising larger drag forces to decelerate as air density increases.

Similar contrasts can be seen in the angle of attack profiles. The minimum t_f solution maintains

[1] i.e. using LHS selected sampling sets of size $n_s = 30n_{input}$, trained using the procedure described in Section 3.1.3 to $R^2 > 99.5$.

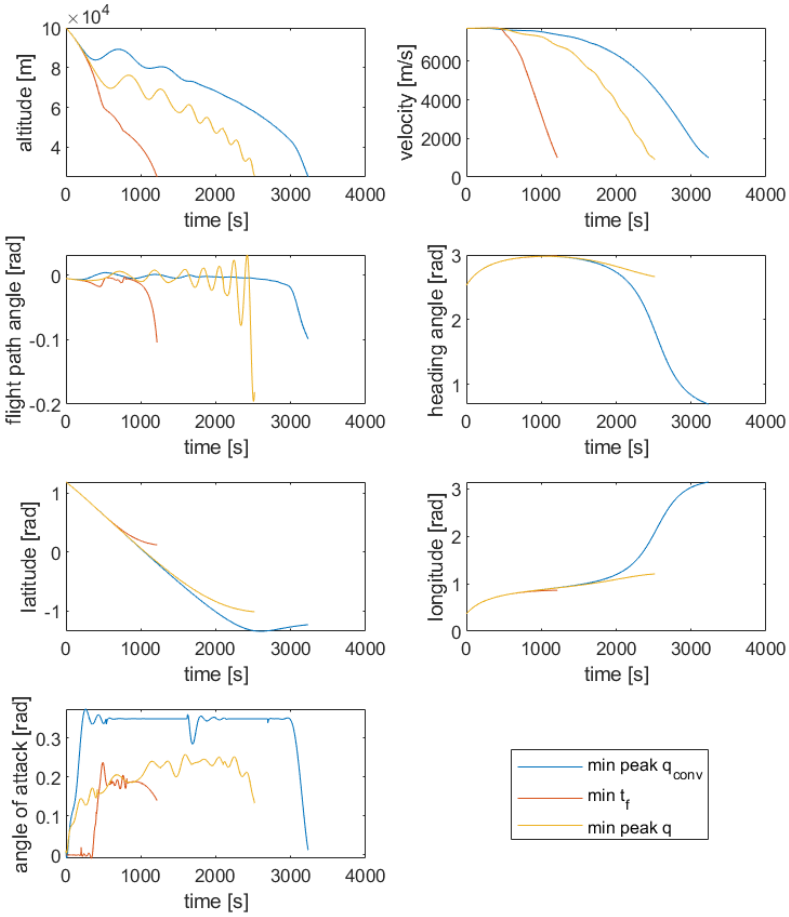


FIGURE 4.9: Dynamic state solutions obtained using only the high-fidelity aerodynamic model. The objective functions considered are the minimisation of peak convective heat flux at the vehicle nose $(\dot{q}_{conv})_{max}$, the total time taken to reach terminal conditions t_f and the peak dynamic pressure experience during re-entry q_{max}.

close to 0deg during the initial descent, before flaring to 10deg at around 60km in order to decelerate within the aerodynamic loading constraints. The minimum $(\dot{q}_{conv})_{max}$ solution displays an almost constant saturation at the maximum allowable value of 20deg, suggesting that this bound is imposing a significant constraint on the available solution. While at this incidence, the stagnation point for convective heat transfer is likely located at the vehicle nose (Minisci et al., 2011), allowing the vehicle to reach more realistic values (40-45deg) would possibly shift this point further down the lower surface, and thus require more sophisticated aero-thermodynamic modelling. In the case where aerodynamic heating is less of a constraint, the optimiser appears to correlate a smaller leading edge radius with improved hypersonic

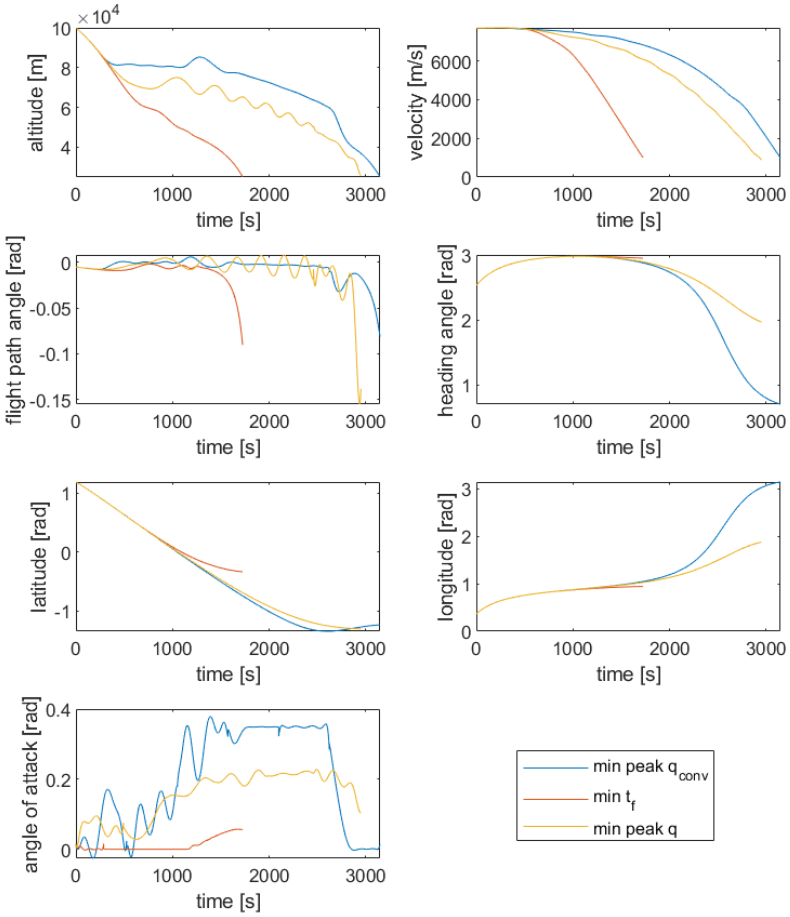
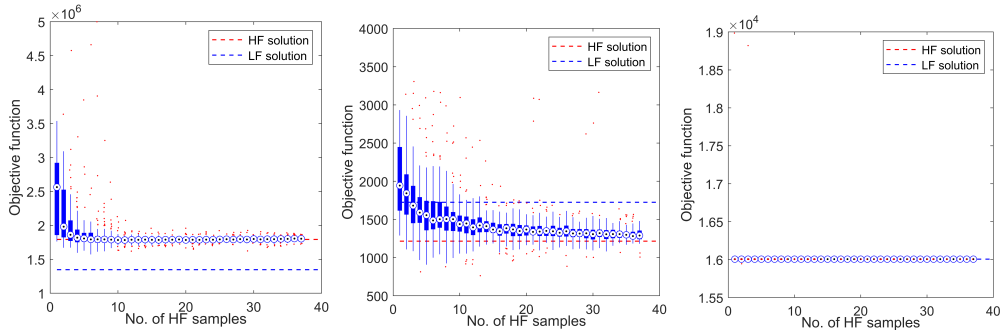


FIGURE 4.10: Dynamic state solutions obtained using only the low-fidelity aerodynamic model. The objective functions considered are the minimisation of peak convective heat flux at the vehicle nose $(\dot{q}_{conv})_{max}$, the total time taken to reach terminal conditions t_f and the peak dynamic pressure experience during re-entry q_{max}.

aerodynamic performance.

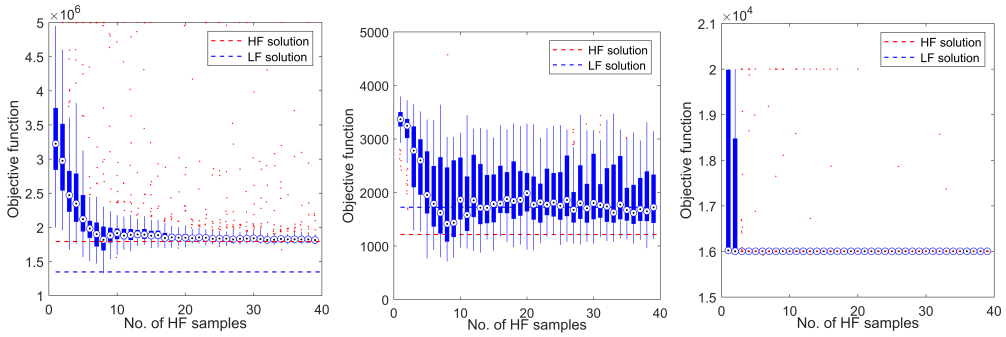
4.4.2 Multi-Fidelity Solutions

The proposed framework is used to solve each single-objective problem. Each analysis is terminated upon reaching a designated maximum number of allowable high-fidelity evaluations, $N_{max} = 40$. Figures 4.11 and 4.12 show the convergence of each problem solution relative to the number of samples, where the latter utilises only a single-fidelity surrogate. In this case, the model is generated using direct non-adaptive sample sets of size N. The solutions obtained when using only the low-fidelity and high-fidelity models respectively are included



(a) (b) (c)

FIGURE 4.11: Objective function convergence when using the multi-fidelity surrogate management framework, where the objective functions considered are the minimisation of **(A)** peak convective heat flux $(\dot{q}_{conv})_{max}$, **(B)** total time taken to reach terminal conditions t_f and **(C)** the peak dynamic pressure experienced during re-entry q_{max}.



(a) (b) (c)

FIGURE 4.12: Objective function convergence when using a single-fidelity aerodynamic surrogate trained only on the available high-fidelity samples. The objective functions considered are the minimisation of **(A)** peak convective heat flux $(\dot{q}_{conv})_{max}$, **(B)** total time taken to reach terminal conditions t_f and **(C)** the peak dynamic pressure experienced during re-entry q_{max}.

for reference.

In each case, the multi-fidelity optimisation demonstrates convergence to the high-fidelity optimum. The single-fidelity method achieves reasonable convergence for the min peak $\dot{q}_{conv}$ and min q_{max} cases, though fails to show similar performance for the min t_f case. While it is possible that outwith the limited number of high-fidelity model evaluations considered the single-fidelity surrogate would indeed allow convergence to the 'true' solution, it can nonetheless be seen that the model management framework does so within a reduced computational cost. Furthermore, the multi-fidelity solution is seen to offer a reduced measure of uncertainty compared with the single-fidelity alternative. Disproportionate variation relative to the expected surrogate accuracy (see Section 4.2) is likely the result of a cascading effect driven by the inherent randomisation within the model training procedure i.e. slight

differences in approximation encourage the optimiser to converge on a significantly altered solution. Intuitively, this issue would appear to be more prevalent for the surrogates built on limited datasets (and thus altogether less reliably representative of global trends), as is the case for the single-fidelity surrogate. The multi-fidelity surrogate however, whilst incorporating the same amount of high-fidelity data, is trained using a comparatively larger dataset of unique points given the inclusion of supplemental low-fidelity data during the information correction process. Thus it would be expected that the multi-fidelity surrogate would exhibit proportionally less variation between successive updates. A consequential implication is that although a given iteration above such a 'critical value' of N (see Section 4.2) may generate a single-fidelity surrogate of higher measurable accuracy than the multi-fidelity counterpart, it's use within an iterative refinement scheme may still yet be misleading given the remaining probabilistic uncertainty in current and future iterations.

Less definite convergence is seen when examining the progression of the vehicle design vector $\mathbf{d}(L, w, n, \theta, r_N)$. For example, Figure 4.13 shows the average component values of $\mathbf{d}(L, w, n, \theta, r_N)$ relative to the number of high-fidelity samples N for both the model management approach and the single-fidelity approach. It can be seen that the vehicle width w, wedge angle θ and leading edge radius r_N design parameters tend to reasonably adherence to the high-fidelity solution within the prescribed budget, with length L and exponent n values showing only weak convergence trends at best. Nonetheless, it is observed that in all cases, the model management framework can be said to facilitate convergence upon the 'true' solution in less high-fidelity evaluations than the equivalent single-fidelity approach, if indeed the latter converges at all within the allotted budget.

In all cases however, the probability of variation in each design variable is much higher than observed for the corresponding objective function. Particularly in cases that exhibit a strong objective convergence with respect to N, the disparity across the design vector $\mathbf{d}$ suggests that the sensitivity (or lack thereof) of the problem objective function to a particular design variable has a significant impact on the correction procedure. For example, Figure 4.14 shows the distribution of vehicle length L relative to N for both multi-fidelity and single-fidelity approaches. Though the average value tends to converge on the 'true' value in less high-fidelity samples than the equivalent single-fidelity method, at the final iteration there still exists a significant interquartile range, and may indeed still saturate almost to either minimum or maximum bound. This is supported via a closer examination of the sensitivities of the low-fidelity and high-fidelity models. Table 4.4 lists the total contribution of each input variable, including all interactions with other variables, to the variance of each output for both the low-fidelity and high-fidelity models. These values were calculated using the formula for total sensitivity given by Eq. (3.56) in Section 3.2.1 (Jansen, 1999; Saltelli et al., 2010). A clear dominance of the angle of attack α on the function responses compared to the remaining

input variables is seen, supporting the conjecture that design vector sampling has been largely overshadowed by the predominance of the Angle of Attack on aerodynamic response. This then raises the question that, despite the cost savings achievable via a sensitivity-biased sampling approach, if the 'less important' variables are more relevant to design/mission performance via a different subsystem, there then may exist a functional disconnect within the optimisation process, potentially compromising the obtained results.

However, the obvious solution here is to then incorporate this different subsystem into the optimisation problem, such as to introduce a driver for the variables not as important to the first subsystem. This could be achieved through the use of one of several *Multidisciplinary Design Optimisation* (MDO) architectures, or perhaps more simply, the reformulation of the problem as a *Multi-Objective Optimisation* (MOO) problem. For example, considering the current problem, an additional subsystem could be one broadly representing the production/operational costs of the subject vehicle, including (for instance) material costs, launch costs (dependant on the vehicle mass) or manufacturing costs (related to the time required to produce the vehicle). These costs are then more inherently dependant on the vehicle design parameters. Thus if any of the previously presented optimisation problems were converted into a MOO problem with an additional objective to minimise the production costs of the vehicle, it is likely that the resulting surrogate would demonstrate a higher global accuracy (a result of increased spreading of additional samples) with respect to its input parameters.

Input Variable	**LF**		**HF**	
	C_L	C_D	C_L	C_D
M	5.5159e-3	2.4265e-4	3.4315e-2	1.7561e-2
α	9.0887e-1	9.1943e-1	7.4819e-1	9.1166e-1
L	7.4432e-4	1.3480e-3	5.6945e-2	8.3851e-6
w	5.8889e-4	2.4266e-3	1.6004e-2	1.0273e-2
n	6.0094e-4	1.2420e-2	5.4573e-2	3.6097e-3
θ	9.3243e-2	8.2324e-2	1.2155e-1	8.4129e-2
r_N	2.3512e-3	1.1366e-3	5.2708e-3	7.5035e-4

Table 4.4: The total contribution of each input variable, including all interactions with other variables, to the variance of each output for both the low-fidelity and high-fidelity models (where $N \approx 34e3$) (Jansen, 1999; Saltelli et al., 2010).

4.4.3 K-fold Cross-Validation Error

To briefly demonstrate the surrogate quality achieved via the proposed methodology during the optimisation process, Figure 4.15 shows the kCVE recorded for each multi-fidelity aerodynamic surrogate, relative to each test case. Of note is that the relative improvement in kCVE is largely independent of the particular objective function considered.

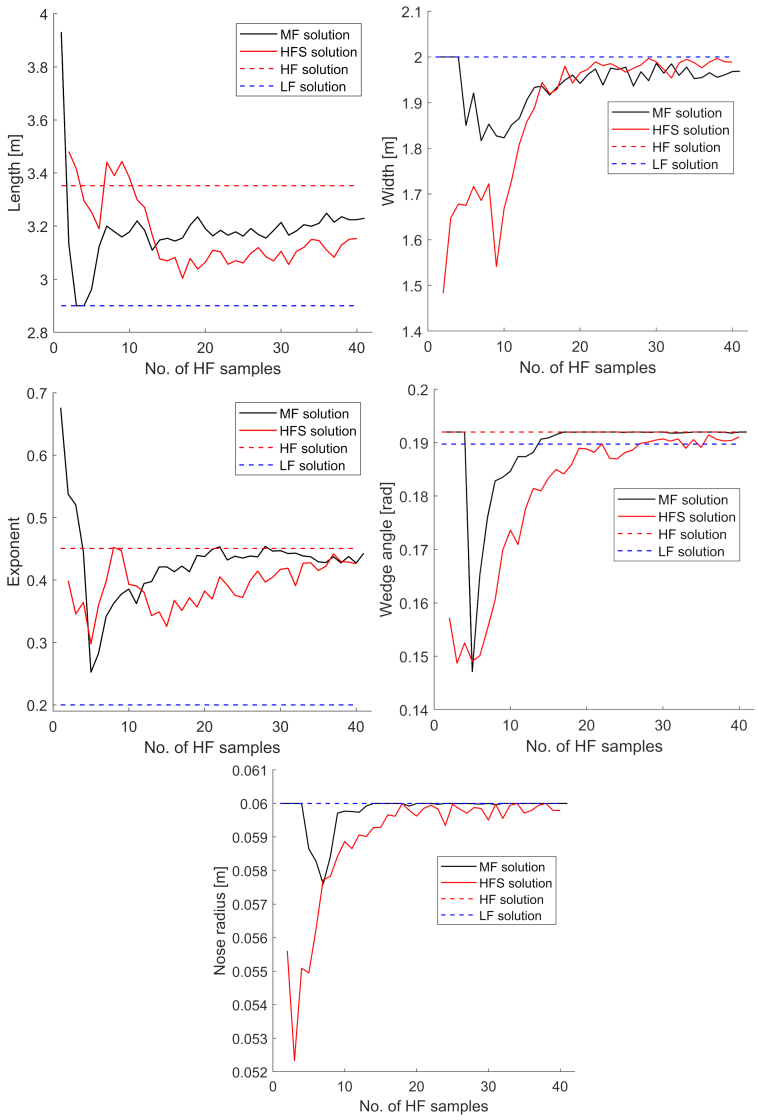
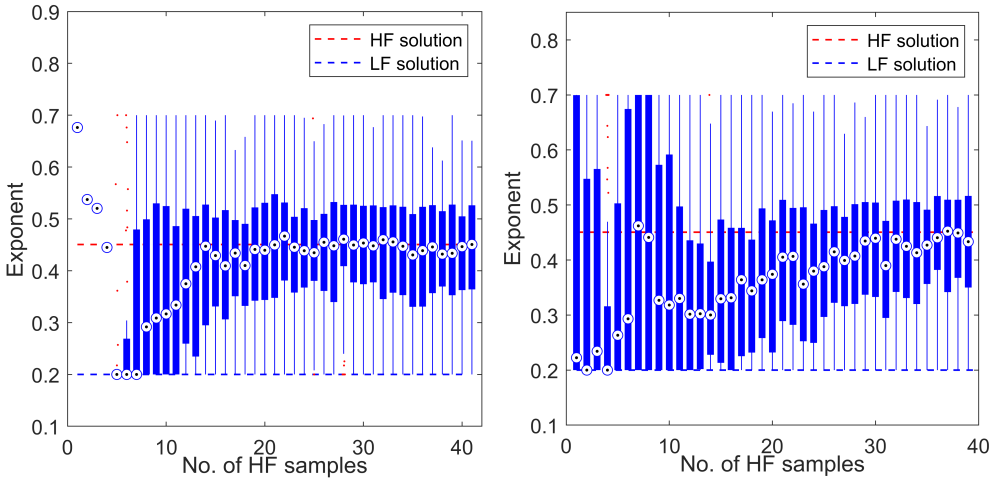


FIGURE 4.13: Averaged values of vehicle design vector $\mathbf{d}(l, w, n, \theta, r_N)$ relative to the number of available high-fidelity samples during the optimisation process, where the objective function considered is the minimisation of peak convective heat flux $(\dot{q}_{conv})_{max}$. The black line represents the results obtained using the multi-fidelity management framework. The red line represents those obtained using only a single-fidelity aerodynamic surrogate trained on the available high-fidelity samples. Included for reference are the values obtained when solving the prescribed problem using only the low-fidelity and high-fidelity aerodynamic models respectively.

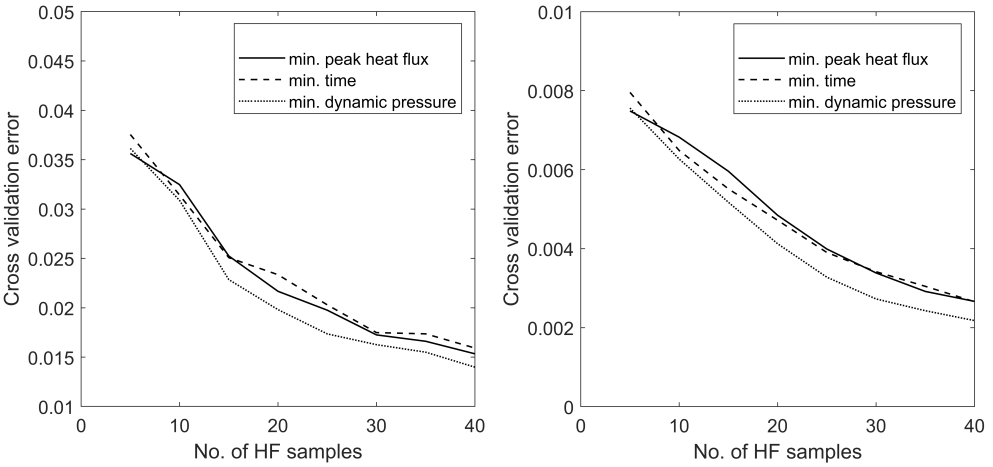
4.5 Cost

With regards to the presented application, the goal was to demonstrate the cost savings offered by the proposed multi-fidelity approach. This section presents a brief cost report considering the framework as a whole. The presented results were obtained over 100 independent runs



(a) (b)

FIGURE 4.14: Distribution values of the vehicle exponent n relative to the number of high-fidelity evaluations when solving for the minimum peak convective heat flux $(\dot{q}_{conv})_{max}$ using **(A)** the multi-fidelity management framework and **(B)** the equivalent single-fidelity surrogate.



(a) (b)

FIGURE 4.15: kCVEs recorded for the multi-fidelity surrogates of **(A)** Coefficient of Lift C_L and **(B)** Coefficient of Drag C_D when solving for each prescribed objective function using the multi-fidelity model management framework.

Mode	Objective	init. SM gen.* [s]	init. guess gen. [s]	optimisation [s]	SM update* [s]
multi-fidelity	min $(\dot{q}_{conv})_{max}$	355.5 (111.6)	245.1 (997.3)	78.8 (217.4)	54.5 (8.9)
	min t_f	383.5 (125.0)	218.7 (833.1)	41.7 (158.5)	56.1 (11.1)
	min q_{max}	438.3 (139.8)	234.4 (938.5)	15.1 (135.7)	56.1 (10.8)
single-fidelity	min $(\dot{q}_{conv})_{max}$	18.5 (4.4)	2396 (1954)	334.3 (636.6)	-
	min t_f	18.8 (5.6)	2383 (2036)	694.0 (1154.8)	-
	min q_{max}	19.2 (4.7)	2410 (2019)	201.6 (584.8)	-

*not including high/low-fidelity model execution time

Table 4.5: Average execution times for the primary stages of the multi-fidelity optimisation process. Included for reference are the corresponding values for the single-fidelity process. Standard deviations for each are included in brackets.

using a 2.5GHz Intel(R) Core(TM) i5-7200 CPU with 8GB RAM.

The average execution times (with standard deviations included in brackets) recorded for a single run of the aerodynamic solvers FOSTRAD and SU2 were 17.21s (4.96s) and 6165s (4077s) respectively. Geometry generation using GMSH accounted for an additional 12.24s (4.1s) for a 2D surface mesh and 34.43s (11.19s) for a 3D volume mesh. The average run-time of a trained ANN surrogate was found to be 1.0388e-4s (4.7078e-04s).

Table 4.5 presents the execution times required by each stage of the optimisation process. Included where appropriate are the corresponding values for the single-fidelity process using only high-fidelity function evaluations. A number of significant disparities can be seen. Firstly, the generation of initial surrogates takes much longer for the multi-fidelity managed approach due to the iterative initial sampling procedure, terminating only when the surrogates have reached acceptable quality. The single-fidelity approach in contrast simply generates surrogates following simple one-shot sampling. Furthermore, this procedure is performed only once during the multi-fidelity process, wheres the single-fidelity results are representative of the time taken each iteration. Secondly, the multi-fidelity managed procedure is an order of magnitude faster when generating the initial guess to be supplied to the optimiser and subsequently converging on an optimal solution. This is in part due to the higher quality of the multi-fidelity surrogate models given the extended sampling and training procedures in operation, but also likely given the stability in variation as a result of the proposed correction process, discussed previously in Section 4.2. Finally, the reported times for surrogate update are representative of the time taken to adaptively select new sample points for multi-fidelity correction, and the subsequent update of the current set of surrogates.

Considering the complete process, Figure 4.16 shows the total execution time of the multi-fidelity management framework and the single-fidelity optimisation process for each of the examined test cases. In each case, the multi-fidelity management framework is seen to facilitate convergence to the high-fidelity solution at a increased rate relative to the single-fidelity approach.

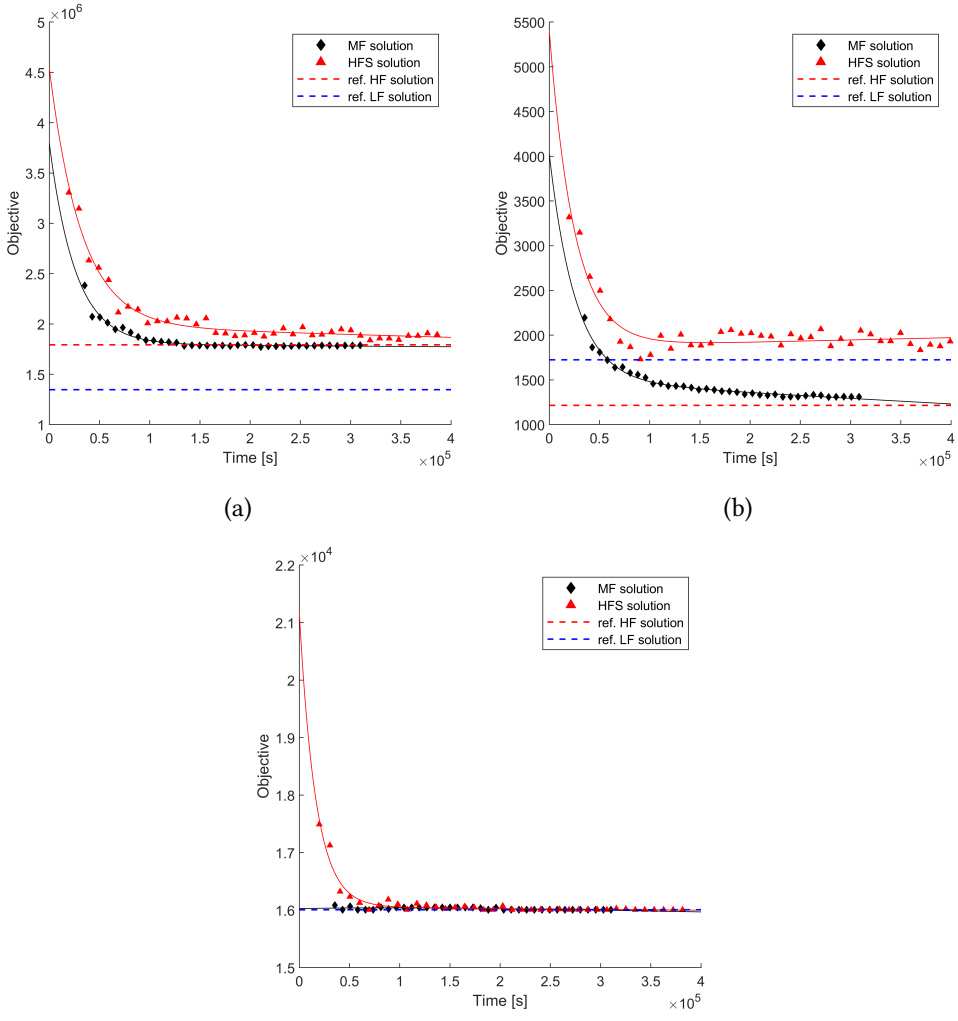


FIGURE 4.16: Total execution times of the multi-fidelity management framework and the single-fidelity optimisation process for the **(A)** minimum peak heat flux $(\dot{q}_{conv})_{max}$, **(B)** minimum time t_f and **(C)** minimum dynamic pressure q_{max} test cases.

4.6 Conclusion

This chapter has demonstrated an approach to incorporate increased levels of simulation fidelity into a vehicle design optimisation problem within a minimised cost budget. Alternative aspects of low-fidelity data were used to facilitate an information correction process, with configurable alternatives compared in terms of available accuracy and computational cost. It was found that the sequential application of corrective stages allows a greater attainable approximation accuracy relative to a single-fidelity equivalent model for a set number of high-fidelity samples up to some critical value. The effectiveness of such correction is generally found to be proportional to the ratio N_{crit}/N_{local}.

This approach was then applied to predict the aerodynamic response of a parametrised waverider subjected to a static/dynamic parameter optimisation. Correction quality of the resulting surrogate was found to be consistent across the input domain, though naturally decreasing with dimensionality. Clear convergence to the optimum was observed via the proposed multi-fidelity management framework for each of three examined single-objective problem variants, in two cases significantly outperforming the equivalent single-fidelity solution in terms of number of required high-fidelity samples. In these cases, objective function convergence occurred both faster and with higher stability, as the constituent surrogates were less susceptible to destructive updates from additional samples.

An observed limitation was that the convergence of the static design solution was far less consistent, likely a result of the sensitivity based adaptive sampling approach tending to favour variables exhibiting a larger influence on the surrogate output parameters. Whilst limiting in this case, a logical mitigation strategy would be to incorporate additional objectives into the optimisation problem such as to introduce a design driver for the variables not as important to the original objective. The result would be a more balanced sampling strategy, simultaneously favouring variables important to any (or multiple) objective(s). This could also be achieved through the use of one of several *Multidisciplinary Design Optimisation* (MDO) architectures.

Another slight limitation was that only minor differences in zero- and first-order surrogate approximation (a result of the random initialisations present in the ANN-based modelling) were capable of significantly altering the resulting solutions. This was seen most prevalently in the single fidelity networks (though to some degree affecting all networks trained on very limited datasets), with the multi-fidelity networks largely reducing the effects of this behaviour, albeit not completely eliminating it.

A final disadvantage of this approach arguably stem from the attempted generalisation, in that the assumption of no *a priori* knowledge is perhaps unfounded for most real-world applications. In this case, multi-fidelity approaches more tailored to the concerned physical process(es) may

achieve higher performance. However, a definitive advantage of the proposed framework and methodologies is the ease to which they may be generally incorporated into early-phase optimisation and/or analysis activities. This then aids the rapid validation of design trade-offs and the determination of interdisciplinary coupling/sensitivities.

The results presented in Sections 4.2 and 4.4 were obtained using the ARCHIE-WeSt High Performance Computer (www.archie-west.ac.uk)

4.7 Bibliography

J. D. Anderson Jr. *Hypersonic and high-temperature gas dynamics*. American Institute of Aeronautics and Astronautics, 2006. ISBN 1600861954.

A. Falchi, E. Minisci, M. Vasile, D. Rastelli, and N. Bellini. DSMC-based Correction Factor for Low-fidelity Hypersonic Aerodynamics of Re-entering Objects and Space Debris. *7th European Conference of Aeronautics and Aerospace Sciences*, 10(July), 2017a.

A. Falchi, M. Vasile, A. Falchi, V. Renato, E. Minisci, and M. Vasile. FOSTRAD : An Advanced Open Source Tool for Re-entry Analysis. *Reinventing Space Conference*, (October):1–15, 2017b.

C. Geuzaine and J.-F. Remacle. Gmsh: A 3-D finite element mesh generator with built-in pre- and post-processing facilities. *International Journal for Numerical Methods in Engineering*, 79: 1309–1331, 2009. doi: 10.1002/nme. URL http://onlinelibrary.wiley.com/doi/10.1002/nme.3279/full.

M. J. Jansen. Analysis of variance designs for model output. *Computer Physics Communications*, 117(1):35–43, 1999. ISSN 00104655. doi: 10.1016/S0010-4655(98)00154-4.

L. Leifsson, S. Koziel, and Y. Tesfahunegn. Fast Multi-Objective Aerodynamic Optimization Using Space-Mapping-Corrected Multi-Fidelity Models and Kriging Interpolation. 153, 2016. doi: 10.1007/978-3-319-27517-8. URL http://link.springer.com/10.1007/978-3-319-27517-8.

P. M. Mehta, E. Minisci, M. Vasile, A. C. Walker, and M. Brown. An Open-source Hypersonic Aerodynamic And Aerothermodynamic Modeling Tool. *8th European Symposium on Aerothermodynamics for Space Vehicles*, (1):9, 2015.

E. Minisci and M. Vasile. Robust Design of a Re-entry Unmanned Space Vehicle by Multi-Fidelity Evolution Control. 51:4–6, 2013.

E. Minisci, M. Vasile, and H. Liqiang. Robust multi-fidelity design of a micro re-entry unmanned space vehicle. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, 225(11):1195–1209, 2011. ISSN 20413025. doi: 10.1177/0954410011410124.

F. Palacios, M. R. Colonno, A. C. Aranake, A. Campos, S. R. Copeland, T. D. Economon, A. K. Lonkar, T. W. Lukaczyk, T. W. Taylor, and J. J. Alonso. Stanford University Unstructured (SU2): An open-source integrated computational environment for multi-physics simulation and design. *51st AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition 2013*, (January):1–60, 2013. doi: 10.2514/6.2013-287.

B. Parsonage and C. Maddock. A multi-fidelity model management framework for multi-objective aerospace design optimisation. *Frontiers in Aerospace Engineering*, 2(February): 1–21, 2023. doi: 10.3389/fpace.2023.1046177.

B. Parsonage and C. A. Maddock. Multi-stage Multi-fidelity Information Correction for Artificial Neural Network Based Meta-modelling. *2020 IEEE Symposium Series on Computational Intelligence, SSCI 2020*, pages 950–957, 2020. doi: 10.1109/SSCI47803.2020.9308255.

L. A. Ricciardi and M. Vasile. MODHOC - Multi Objective Direct Hybrid Optimal Control. 2018. URL https://strathprints.strath.ac.uk/67085/.

L. A. Ricciardi, C. A. Maddock, and M. Vasile. Direct Solution of Multi-Objective Optimal Control Problems Applied to Spaceplane Mission Design. *Journal of Guidance, Control, and Dynamics*, 42(1):30–46, 2019. ISSN 0731-5090. doi: 10.2514/1.g003839.

A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259–270, 2010. ISSN 00104655. doi: 10.1016/j.cpc.2009.09.018. URL http://dx.doi.org/10.1016/j.cpc.2009.09.018.

R. P. Starkey and M. J. Lewis. Analytical off-design lift-to-drag-ratio analysis for hypersonic waveriders. *Journal of Spacecraft and Rockets*, 37(5):684–691, 2000. ISSN 00224650. doi: 10.2514/2.3618.

M. Vasile. Finite elements in time: A direct transcription method for optimal control problems. *AIAA/AAS Astrodynamics Specialist Conference 2010*, pages 1–25, 2010. doi: 10.2514/6.2010-8275.

Part II

Resolution-based Fidelity Management

Background 5

5.1 Optimal Control . . . 224
5.1.1 Indirect Methods . . . 225
5.1.2 Direct Methods . . . 227
5.2 Adaptive Mesh Refinement for Optimal Control . . . 233
5.3 Multi-Objective Optimisation . . . 236
5.3.1 Scalarisation-based . . . 237
5.3.2 Dominance-based . . . 238
5.3.3 Others . . . 240
5.4 Multi-Objective Optimal Control . . . 240
5.5 Adaptive Mesh Refinement for Multi-Objective Optimal Control . . . 242
5.6 Conclusion . . . 244
5.7 Bibliography . . . 246

This chapter compiles selected background information from the associated literature relevant to adaptive mesh refinement of *Multi-Objective Optimal Control Problems* (MOCPs). Section 5.1 briefly reviews the two primary classes of method used to solve Optimal Control problems, namely; *direct* and *indirect* methods. Section 5.2 introduces the issue of mesh refinement for direct Optimal Control problems and discusses several prevalent methods with which this can be achieved in an adaptive format. Section 5.3 introduces and discusses *Multi-Objective Optimisation* (MOO), including a brief description of scalarisation vs. dominance-based techniques, with several examples of prominent methods for each. Section 5.4 discusses the application of *Multi-Objective Optimisation* (MOO) methodology to *Multi-Objective Optimal Control Problems* (MOCPs), again with several examples of existing approaches. Finally, Section 5.5 introduces the premise behind applying adaptive mesh refinement methods to aid solving *Multi-Objective Optimal Control Problems* (MOCPs), and the implications thereof.

5.1 Optimal Control

Modelling the dynamic behaviour of a controllable system requires the inclusion of algebraic variables in the mathematical description. The associated design problem can thus become (in part) the selection/determination of these control variables such that some objective function is optimised according to the set of *Ordinary Differential Equations* (ODEs) describing the differentiated system state (Betts and Huffman, 1998). Some problem statements may also include algebraic constraints on the system state and/or controls as a further complicating factor.

The problem of how to control a general non-linear dynamical system from one initial state to a desired final state in some optimal sense, is approached mathematically by formulating and solving an *Optimal Control* (OC) problem. This class of problem can be considered an extension of the NLP problem to include an infinite number of variables (Betts, 1998). However, unless the problem is so simple as to exhibit only academic relevance (e.g., the infinite-horizon linear quadratic problem, see Kwakernaak et al. (1974)), analytical solutions are rarely attainable. Therefore, numerical techniques must be employed.

The historical development of the numerical methods with which to solve OC problems has closely paralleled the progress made in the underlying NLP methods (itself closely linked to the emergence of the digital computer, as discussed in Section 1). For example, until the early 1960s, most implementations consisted of a parametric optimisation approach (i.e. by hand) using a simple Newton's method (Betts, 1998). Problem size was therefore practically limited to around $n = m \approx 10$. Quasi-Newton methods became popular in the 1970s (Fletcher, 2013; Dennis Jr and Schnabel, 1996). Prevalent approaches included applying an unconstrained minimisation algorithm to a modified form of the objective function as a way of dealing with constraints. Alternatively, a method for constrained problems (known as the reduced-gradient approach), identified a set of basic variables that could be used to eliminate active constraints, thus facilitating the selection of the non-basic variables using an unconstrained optimisation method. Most applications in this decade and into the early 1980s were of moderate size, i.e., $n = m < 10^2$ (Betts, 1998). As computing power continued to increase exponentially, methods exploiting matrix sparsity began to emerge, allowing problems of up to $n = m \approx 10^5$ to be numerically solved (Gill et al., 2005).

Numerous interesting works that survey Optimal Control methods and their development are available in the open literature (Bryson and Earl, 2018). Considering modern historical development, the survey by Polak (1973) provides an overview of computational methods prior to the widespread emergence of the digital computer, including the pioneering investigations by Bellman (1957); Bellman and Dreyfus (1962). Advancements in gradient-based methods for Optimal Control made during the 1960s and early 1970s are reviewed in (Miele, 1975). Bryson

(1996) extends this scope of review to primarily cover from 1950 to 1985, though including an interesting historical perspective on the roots of the associated *calculus of variations* developed during the 1600s[1]. Hull (1997) presents a brief review of methods for direct transcription of a continuous-time Optimal Control problem into a finite-dimensional parameter optimisation problem. The work by Betts (1998) includes an excellent review of modern numerical methods for Optimal Control and trajectory optimisation, including a summary of the significant amount of research completed prior to 1990. Finally, Rao (2013) presents a review of modern numerical methods for solving Optimal Control problems. Included is a discussion of pertinent computational issues, available software and an approach to select the most appropriate method(s) for a given problem.

The numerical approaches described in the above surveys can be separated into two broad categories; *indirect* and *direct* methods. To help further classify the distinction between the two, Rao (2013) first describes the practice of solving Optimal Control problems as reliant on three fundamental areas:

- Solving systems of differential equations and integrating functions.
- Solving systems of non-linear algebraic equations.
- Solving non-linear optimisation problems.

All numerical methods for Optimal Control require solving differential equations and integrating functions. Further discussion of these aspects is reserved for the appendices. An indirect method combines these processes with numerically solving systems of non-linear algebraic equations, whilst a direct method combines instead with non-linear optimisation. This relationship is summarised graphically in Figure 5.1 (Rao, 2013). Whilst precluding an attempt to simply reproduce existing derivations, the following section briefly introduces the most widely used approaches within each of these classifications. Of note, von Stryk and Bulirsch (1992) provides a brief list of commonly used methods for Optimal Control developed prior to the 1990s and emphasises the usefulness of combining indirect methods and direct methods (i.e. *hybrid* approaches). These methods, however, are considered outwith the scope of this work.

5.1.1 Indirect Methods

Indirect methods use the calculus of variations to derive and explicitly solve the necessary conditions for local first-order optimality in terms of the adjoint differential equations, Pontryagin's

[1] See also the works by Bliss (1963); Goldstine (1980); Sussmann and Willems (1997) for further fundamental background on this topic.

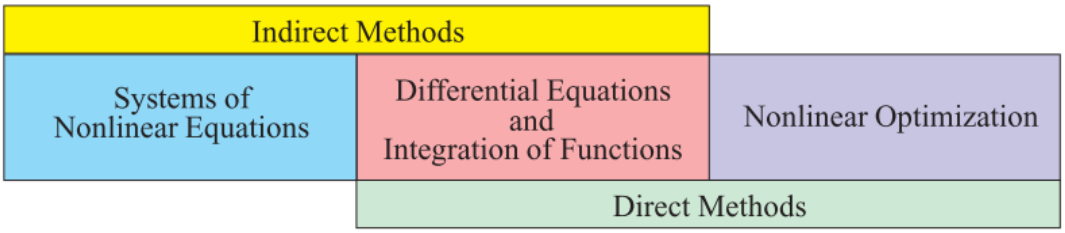


FIGURE 5.1: Solving Optimal Control problems: *Indirect* vs. *direct* methods.

maximum principle and the associated boundary (transversality) conditions. The maximum principle behaves as a constrained optimisation problem, where the control variables $\mathbf{u}(t)$ are chosen such as to optimise the system Hamiltonian at every instance of time t subject to the limitations imposed by the state and control path constraints (Betts, 1998). The complete set of necessary conditions includes a *Differential Algebraic system of Equations* (DAE) with boundary conditions imposed at the initial and final instances of time, commonly t_0 and t_f respectively. This is referred to as a two-point boundary value problem. which may be repeatedly solved to give a number of candidate solutions (extremals). Each extremal is analysed to determine local optimality, and the particular solution with the lowest cost function value is selected (Rao, 2013).

The solution of multiple point boundary value problems can be determined by solving the system of non-linear algebraic equations of the form:

$$\begin{aligned} f(\mathbf{z}) &= 0 \\ g_{min} \leq g(\mathbf{z}) &\leq g_{max} \end{aligned} \tag{5.1}$$

It can therefore be said that indirect methods attempt to find a root of the necessary condition for optimality $F'(\mathbf{x}) = 0$ (that is, a stationary point of the objective function F). By analogy, an indirect method for optimising a function of n variables requires analytically computing the function gradient, then locating a set of variables using a root-finding algorithm such that the gradient is equal to zero. An indirect method must therefore be able to compute the system derivative and determine whether it is sufficiently close to 0.

This introduces a number of difficulties. Computing system derivatives requires explicitly deriving the coupled set of adjoint DAE describing the states and co-states of the system, the control equations and all of the transversality conditions (Betts, 1998). For complicated non-linear dynamics, this may not be a simple task. Secondly, the region of convergence for a root-finding algorithm can be small, especially when it is necessary to guess the initial values of the adjoint variables (co-states), that often do not have an obvious physical interpretation. Thirdly, for problems with path inequalities, it is necessary to explicitly state the sequence of constrained and unconstrained sub-arcs, which may be difficult to estimate *a priori*, before

iteration can begin. Finally, the optimality conditions are only first order local conditions.

Indirect methods are highly accurate, particularly for simple problems where closed form analytical solutions may be derived. In these cases however, should the objective function, boundary conditions or dynamics of the system change, the necessary conditions for optimality must be completely re-derived (through analytic manipulation) from scratch. However, most practical applications do not fit this this requirement for simplicity, thus numerical techniques must be employed. For the reasons stated above then, indirect methods were not considered further as part of this work.

5.1.2 Direct Methods

Direct methods, in a fundamental departure from indirect methods, seek to approximate the state and/or control variables of the original Optimal Control problem across a finite set of grid points (nodes) that subdivide the current time phase into a series of intervals. In this manner, a continuous-time infinite-dimensional Optimal Control problem is discretised into a finite-dimensional NLP problem, where states and/or controls are represented by an appropriate set of polynomials. The complete transcribed problem may then be solved (usually) via any standard gradient-based NLP solver. In essence then, the dynamic variables are adjusted to directly optimise the problem objective function.

Direct methods offer a number of numerical advantages over indirect methods. An analytic expression for the necessary conditions is not required and therefore neither is an initial guess for the adjoint variables (co-states). As the transcription is automatic, the process only needs re-run should alternative/additional objectives, boundary conditions or system dynamics be required. The imposition of path constraints is similarly straightforward. Finally, a large number of efficient and high-quality gradient based NLP solvers are available both open source and commercially. Solving direct Optimal Control problems can be summarised as following the general approach:

1. Transcribe the continuous dynamic system into a problem with a finite set of variables.
2. Solve the finite dimensional (i.e. NLP) problem using a suitable parameter optimisation method.
3. Assess the accuracy of the finite dimensional approximation. If necessary, repeat (1) and (2) at a higher level of fidelity (for example, via some appropriate grid refinement strategy).

With regards to the first point, direct methods are most commonly grouped according to the extent of parameterisation employed during the discretisation step. For example, whether

only the control variable(s), or both the control and the state variables are parameterised into piecewise polynomials. The following briefly overviews the most commonly applied direct methods, including single and multiple shooting methods and orthogonal/pseudospectral collocation methods.

5.1.2.1 Shooting

The Shooting (or 'Single Shooting') method is the simplest direct method for solving Optimal Control problems. Particularly useful for launch trajectory and orbital transfer optimisation problems, which lend themselves to parameterisation with relatively few NLP variables (Betts, 1998). The control vector $\mathbf{u}(t)$ is parametrised across m phases according to the functional form:

$$\mathbf{u}(t) \approx \sum_{i=1}^{m} \mathbf{a}_i \phi_i(t) \tag{5.2}$$

where $\phi(t)_i, (i = 1, \ldots, m)$ are the basis functions (e.g. B-spline) and $a_i, (i = 1, \ldots, m)$ are the parameters to be determined by the optimisation algorithm. To satisfy the system dynamics, the differential equations are integrated across all phases using some appropriate explicit time-marching algorithm (for example, Runge-Kutta). The constraint and objective functions are evaluated at one or more of the phase boundaries.

The basic form of the single-shooting method involves an initial guess of the control vector $\mathbf{u(t)}$ at $t = t_0$, subsequently used to propagate the system state vector according to the defined system dynamics until $t = t_f$ (see Figure 5.2). The final state is evaluated in terms of a pre-defined objective, and the initial guess is updated accordingly. This process is repeated until some selected measure of optimality indicates convergence within a specified tolerance, at which point the process is then terminated.

Direct single-shooting applications are successful due to their ability to represent a problem using a relatively small number of variables. The accuracy of these approaches thus relies upon both the order of the integrator used and the size of the corresponding time steps. Considering the latter, adaptive step size methods may be employed to locally refine the solution in areas of higher non-linearity.

Single shooting methods do however suffer from large resultant variations in the projected final conditions propagated from relatively small changes early in the trajectory (i.e. near t_0). In other words, the constraints can exhibit significant non-linear behaviour with respect to the control variables, consequentially complicating the optimisation problem (Betts, 1998). Another issue is the resultant computational cost of calculating the necessary derivative information. Most commonly this is achieved using finite difference approximations. However,

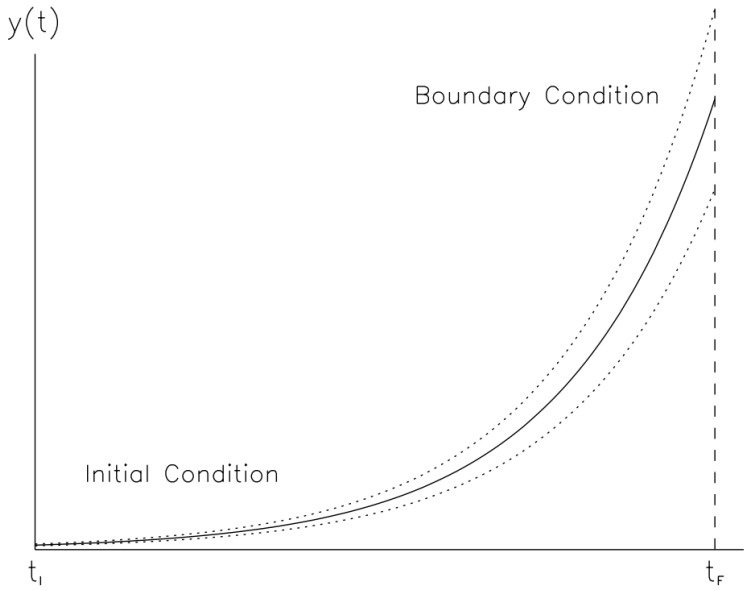


FIGURE 5.2: Single-Shooting (Betts, 1998).

with this approach it becomes necessary to integrate the trajectory for each perturbation. Accordingly, at least n solutions are required to evaluate a finite difference gradient, which itself may be required at each NLP iteration. Using finite difference approximations also incurs an additional risk of noise in the gradient calculations, especially if using variable step-size methods Betts (1998). Forward differences offer an accuracy of order $\mathcal{O}(\delta)$, compared to the accuracy of $\mathcal{O}(\delta^2)$ using central difference methods (though at twice the computational cost). An alternative to finite difference approximations for this application, albeit uncommon, is to integrate the *variational equations*. This involves an additional differential equation introduced for each NLP variable, to be solved alongside the state equations as part of an augmented system of differential equations. However, these equations must be derived separately for each application, the inconvenience of which resulting in the apparent rarity of this approach Betts (1998). Lastly, difficulties can be encountered if equality path constraints are included in the problem definition, or in the equivalent case where the dynamics are described as a DAE. This is because the values of the intermediate stages between the boundaries of each shooting segment are not explicitly represented, thus are invisible to the NLP solver (Betts, 1998).

5.1.2.2 Multi-shooting

The multi-shooting method (Keller, 1968) instead divides the complete time interval $[t_0, t_f]$ into a series of $M + 1$ segments. Each individual segment is solved using the single shooting approach. In this case, the employed numerical optimisation method selects the values of the state vector at the beginning of each interval in addition to the control parametrisation unknowns. This method requires additional defect constraints to enforce continuity between

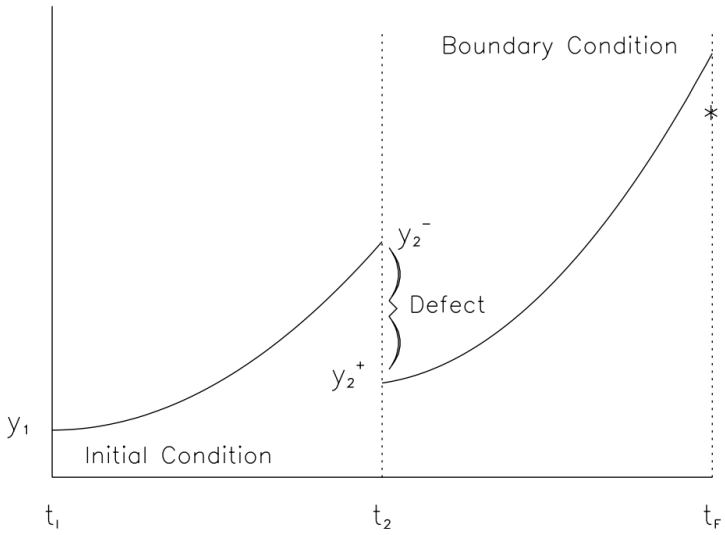


FIGURE 5.3: Multi-Shooting (Betts, 1998).

consecutive intervals, thus ensuring that the state vector exhibits continuous behaviour across each interface.

$$\mathbf{x}(t_i^-) = \mathbf{x}(t_i^+) \iff \mathbf{x}(t_i^-) - \mathbf{x}(t_i^+) = 0 \tag{5.3}$$

Equation (5.3) can be formed into a vector root-finding problem, where the objective is thus to reduce the error between $\mathbf{x}(t_i^-)$ and $\mathbf{x}(t_i^+)$ to zero. Naturally the size of the associated NLP required for the multi-shooting method is significantly larger than for single-shooting, owing to the additional variables and constraints. Mitigation strategies thus include the exploitation of matrix sparsity to more efficiently solve the NLP problem, or utilising parallel processing capabilities. Regardless, the primary advantage remains the greatly reduced instability effects experienced (as compared to the single-shooting approach) due to the shorter time intervals the integration is applied over.

5.1.2.3 Collocation

The most widely used direct methods implementing state and control parametrisation are known as Collocation methods (Ross and Karpenko, 2012). Collocation methods can be further sub-categorised into local and global methods. Local collocation methods vary the number of segments the problem time interval is divided into whilst approximating the state through each using a fixed degree polynomial, whereas global collocation methods instead fix the number of segments and vary the degree of the approximating polynomial. Both local and global collocation methods tend to employ Chebyshev or Lagrange polynomials as basis functions.

Local collocation methods approximate the state and control profiles using a successive series of piecewise continuous polynomials that satisfy (collocate) the system ODEs at the series

of grid points (also referred to as collocation points, mesh points, or nodes) t_j subdividing the current time domain[2] i.e., $t_j \leq t \leq t_{j+1}$ (Betts, 1998). Like all direct methods, the collocation approach does not require the explicit specification of the necessary conditions for optimality, i.e., the adjoint equations, transversality conditions and the maximum principle. This suggests versatility and robustness, making this approach attractive for complicated applications. Unlike other direct methods however, collocation does not require the prior specification of the particular arc sequence for problems with path inequalities. This is because the underlying NLP effectively determines the subarcs using an active set strategy.

Local collocation for Optimal Control is typically employed using one of two classes of discretisation: Runge Kutta (with implicit and explicit variants, although with the former being much more practical due to the improved stability characteristics, and therefore more widely used) and Orthogonal Collocation methods (Rao, 2013). Orthogonal Collocation for Optimal Control was first investigated by Reddien (1979) where cubic splines approximating states and controls were defined over a set of Legendre-Gauss points. The subsequent work by Cuthrell and Biegler (1987, 1989) also used Legendre-Gauss points, though instead with Lagrange polynomials. Cuthrell and Biegler (1989) then demonstrated that an indirect transcription using *Legendre-Gauss* (LG) points was in fact equivalent to the *Karush-Kuhn-Tucker* (KKT) conditions obtained from the NLP of the direct formulation. Throughout the 1990s, the use of higher-order Gauss-Lobatto collocation methods was developed (Herman and Conway, 1996; Enright and Conway, 1992). More recent work involved the study of orthogonal collocation convergence rates using *Legendre-Gauss-Radau* (LGR) points (Kameswaran and Biegler, 2008).

The key notion of the collocation methods is to replace the original set of ODEs describing the differentiated system state with the set of defect constraints $\zeta_i = 0$, which are imposed on each interval in the discretisation and subsequently combined with the original path constraints. Naturally this formulation results in a large NLP problem, with the numbers of NLP variables $n \approx (n_y + n_u)MN$ (where n_y is the number of state variables, n_u the number of control variables, M is the number of grid points and N is the number of phases) and a similar number of NLP constraints. Thus a typical trajectory with 7 states, 2 controls, 100 grid points per phase and 5 phases produces an NLP with $n = 4500$ (Betts, 1998). Fortunately, the problem Jacobian and Hessian are sparse (typically with 1% of matrix elements non-zero). Thus methods exploiting matrix sparsity, such as SNOPT (Gill et al., 2005), SPRNLP (Betts and Frank, 1994) and KNITRO (Byrd et al., 2006), are particularly successful in terms of reducing computational storage and execution time.

A common form of global collocation is the *pseudospectral* approach. Pseudospectral methods approximate the state vector using a global polynomial (such as Chebyshev or Legendre), as opposed to the piecewise polynomial approach of local collocation methods. Collocation is

[2] In essence, similar to the approach of multi-shooting.

then performed at specified points. By virtue of the global state approximating polynomial and orthogonally collocated points, pseudospectral methods converge upon the solution *spectrally* (exponentially) relative to the number of collocation points. The term pseudospectral could almost be considered as a synonym for collocation (Jain and Tsiotras, 2008). However, in practice, pseudospectral is usually applied only when collocation is performed in conjunction with a basis of global functions. Another major distinction is that pseudospectral methods employ differentiation, relying on the discretisation of the *tangent bundle* (the left-hand side of the differential equations). Most other collocation methods, in comparison, are based on integration, relying on the approximation of the vector field (i.e., the right-hand side of the differential equations) (Jain and Tsiotras, 2008). For example, orthogonal collocation methods typically approximate the state variables over time interval $\tau \in [-1, 1]$, such that for a N phase problem:

$$\mathbf{x}(\tau) \approx \mathbf{X}(\tau) = \sum_{k=1}^{N} \mathbf{c}_k L_K(\tau) \tag{5.4}$$

where the functions $L_k(\tau), (k = 1, \ldots, N)$ are Lagrange polynomials. Lagrange polynomials satisfy the isolation property (Rao, 2013):

$$L_k(\tau_i) = \begin{cases} 1, & k = j \\ 0, & k \neq j \end{cases} \tag{5.5}$$

such that together, Eqs. (5.4) and (5.5) lead to the calculation of parameters $\mathbf{c}$ as:

$$\mathbf{c}_k = \mathbf{x}(\tau_k) \tag{5.6}$$

Earlier examples of pseudospectral methods applied to Optimal Control problems predominantly utilised Chebyshev points (with Chebyshev basis functions). It was found that, in general, this approach did not satisfy the isolation property, which leads to much more complicated collocation conditions (Rao, 2013). As a result, the majority of modern applications of pseudospectral methods use Lagrange polynomials as the basis functions.

Pseudospectral methods that utilise Lagrange polynomials can be further categorised into three subsets, defined by their corresponding type of nodes; Gauss-Lobatto points (either Legendre-Gauss-Lobatto or Chebyshev-Gauss-Lobatto points, or Legendre-Gauss-Radau points. These various permutations can be summarised by analysing their relative similarities and differences. Both the LGL (Legendre pseudospectral method with Gauss-Lobatto points) and the infinite-horizon LGR (Legendre-Gauss-Radau points) methods approximate the state using a global polynomial of degree M with the dynamics collocated at $M + 1$ points. Collocating the system derivative at each approximation point gives rise to a differentiation operator that is both square and singular. Conversely, the GPM (Gauss pseudospectral method, usually employing Legendre-Gauss points) and RPM (Radau pseudospectral method, a variation on

LGR) approximate the state using a polynomial of degree M and collocate the dynamics at M points. As in this case the derivative is not evaluated at each approximation point, the differentiation matrix is left full rank and non-square.

One pitfall of all pseudospectral methods is the dependence on the smoothness of the target functions to allow spectral convergence/accuracy (Fahroo and Ross, 2008). For example, if the solution contains highly non-linear or discontinuous regions, for instance due to *bang-bang* control, pseudospectral methods may exhibit spurious oscillations (see Gibbs phenomenon), a known risk when applying polynomial interpolation. In these cases then, spectral convergence cannot be achieved. Several methods based on basic pseudospectral methods (orthogonal collocation) have been developed to mitigate this dependency. For example, dividing a non-smooth problem into finite elements to capture discontinuities (Cuthrell and Biegler, 1987) (or equivalently, reformulating into a multi-phase problem and optimising the temporal boundaries of each phase, using the previous solution as a first guess), spectral patching (Fahroo and Ross, 2000), or allowing non-linear transformations at the interface between trajectory segments (Ross and Fahroo, 2004).

5.2 Adaptive Mesh Refinement for Optimal Control

A fundamental issue concerning numerical methods to solve Optimal Control problems is how well the chosen scheme may approximate the 'true' solution (Betts and Huffman, 1998). Most practical Optimal Control problems include some number of state constraints, control constraints and equality/inequality boundary and/or path constraints. As a consequence, particularly when the path constraints do not involve the control variable explicitly, the optimal solution(s) may exhibit rapid changes or discontinuities in state and control profiles and/or their respective higher order derivatives (Zhao and Li, 2017). Similarly, when the control variables are represented as linear functions within the system differential equations, *bang-bang* solutions can often be expected (Betts and Huffman, 1998).

Considering direct transcription methods (Section 5.1.2), the accuracy of obtained solutions may therefore be highly dependant on the grid resolution of the discretisation nodes at which states and controls are evaluated. Specifically considering discontinuities, some level of inaccuracy must be expected unless the location of the discontinuities are introduced explicitly as phase boundaries. If the constraint functions are to be satisfied to within a strict tolerance (as is often required in the case of multidisciplinary design, for example), a high-resolution (dense) discretisation grid is typically required. However, the use of dense grids (as with high-accuracy numerical schemes) during the initial iterations, or indeed at any iteration if the grid is to remain uniform, may require an excessive computational budget (in terms of both CPU time and memory) given the resulting size of the NLP problem and the associated augmentation

of search space dimensionality. Such a strategy may therefore remain impractical for most applications, especially if the resulting NLP problem is not sparse (Jain and Tsiotras, 2008; Betts and Huffman, 1998). Furthermore, as the number of grid points increases, the chosen algorithm may become ill-conditioned and exhibit poorer convergence characteristics, ultimately failing to acceptably converge (Zhao and Li, 2017). For similar comments corroborating this point of view, see also Jain and Tsiotras (2008); Schwartz (1996).

Significant research effort has therefore been focused on defining suitable sparse representations or on the reduction of the high computational load associated with uniform grid discretisations whilst balancing the fundamental desire for accuracy (Jain and Tsiotras, 2008). This latter aspect is most commonly approached via a progressive *non-uniform* refinement of the grid resolution (Jain and Tsiotras, 2008; Betts and Huffman, 1998; Betts, 1998; Ross and Fahroo, 2004; Gong and Ross, 2006; Darby et al., 2011). Non-uniform refinement entails the adaptive addition, subtraction or redistribution of discretisation nodes to concentrate the grid around non-smooth regions (or, particular areas of the solution that dominate the overall accuracy), while smooth regions remain relatively sparse[3]. In this manner, a *multi-resolution* grid may capture highly non-linear or discontinuous behaviour within a significantly reduced computational budget.

Such approaches have been well documented in the literature (Betts and Huffman, 1998; Betts, 1998; Jain and Tsiotras, 2008; Zhou et al., 2017; Schwartz, 1996; Binder et al., 2000; Schlegel et al., 2005; Assassa and Marquardt, 2014; Zhao and Tsiotras, 2011). The works by Betts and Huffman (1998); Betts (1998) select new grid points by solving an integer programming problem that minimises the maximum discretisation error by subdividing the current grid. Schwartz (1996) proposed two static mesh refinement strategies minimising the maximum local discretisation error in solving optimal control problems based on Runge-Kutta (RK) integration. Binder et al. (2000) uses a direct shooting approach in which the OC problem is converted into a NLP problem by parameterising the control profile, combined with a bi-orthogonal wavelet analysis of the gradients of the Lagrangian function with respect to the parameterising functions at the optimal points to determine the regions that require refinement. Schlegel et al. (2005) also employ a wavelet analysis, though in this case of the control profile, in order to select regions for refinement in problems with state and/or control path constraints. Both Binder et al. (2000) and Schlegel et al. (2005) use wavelet-based ideas to locate possible singularities in the solution and to refine the grid locally. However, because both methods operate solely in the wavelet domain, they may induce increased computational overhead owing to the need to transform back and forth from the physical domain. In the spectral patching method by Fahroo and Ross (2000), the time interval is divided into several non-overlapping continuous

[3]This can be considered a form of *h*-refinement, as opposed to adapting the order of the approximating polynomial (*p*-refinement) Babuška and Suri (1990).

sub-intervals, where the continuity for both states and controls is enforced at each interface. Ross and Fahroo (2004) generalise this approach into the so-called pseudospectral knotting method. Information is exchanged across patches in the form of event conditions associated with the OC problem, hence removing the restriction of continuity in the solution across the end points of the phases. The phase boundaries, termed as knots, can be fixed or free, with the free knots being part of the optimisation process. On each phase, the problem is solved using the Legendre or Chebyshev pseudospectral method. Gong and Ross (2006) further build upon this work with an algorithm where the user specifies the number of nodes to be increased within a particular phase, in case the error of the computed optimal control between two successive iterations is greater than a prescribed threshold. The gradient of the control is used to approximately determine the location of the knots. However, for these methods (Ross and Fahroo, 2004; Gong and Ross, 2006), the approximate number and location of singularities in the solution must be specified *a priori*, which may be difficult for most problems. Accordingly, the number of nodes to be added to a particular phase must be defined by the user before starting the algorithm. Jain and Tsiotras (2008) present an automated adaptive multi-resolution technique, where the OC problem is discretised onto a series of complimentary, refinable dyadic grids. The criterion for deciding the region in which to add/remove grid points (each may be completed in a single step) is based on simple local interpolation error analysis. This avoids the need for any secondary optimisation as in the works by Betts et al. (2002); Betts and Huffman (1998); Betts (2010). Furthermore, this algorithm does not require any integrations, as opposed to the highly accurate integrations (Romberg quadratures) employed by Betts et al. (2002), which can be computationally expensive for non-linear dynamics. This algorithm will automatically detect the regions in the solution that are non-smooth, thus does not require the *a priori* specification of the locations of any irregularities. Finally, the technique is shown to be generally robust. Because of the multi-level nature of dyadic grid points, most do not shift in successive iterations, which benefits the overall convergence characteristics.

More recently, Zhao and Li (2017); Zhao and Shang (2018); Zhao and Li (2020) present a multi-resolution approach for solving Optimal Control problems using direct collocation. Enabled by the work of Jain and Tsiotras (2008), the algorithm may start from any uniform initial mesh with an odd number of grid points and refine the mesh at locations where the interpolation error remains outside a specific threshold. The resulting method is extended to generalised dyadic grids. In a related work, Zhao and Li (2019) adapt and apply this method to optimise the descent trajectory of a Mars reentry vehicle with respect to various boundary and path constraints. A similar approach is proposed by Li et al. (2020), where a mesh refinement procedure and trajectory optimisation scheme are implemented in a nested, bi-level structure. In the inner-level, a convex optimisation algorithm locally solves the trajectory problem, with the solution subsequently passed to the outer-level for grid refinement. This approach is applied to solve a trajectory planning problem for spacecraft relative motion.

5.3 Multi-Objective Optimisation

Generally speaking, optimisation problems (including the majority of Optimal Control problems) tend to be formulated according to only one performance criterion that needs to be optimised (Bryson, 1996; Betts, 2010). However, most examples of engineering design are fundamentally multi-objective in nature, due to the inherent tensions between four main areas: performance, cost, schedule and risk (Maier, 2009). Schedule and risk levels (e.g. probability of component failure) are often fixed (or uncontrollable), therefore improving the overall system performance is most easily achieved by increasing economic cost. In essence, one can say that 'improving' any single one of these aspects causes a corresponding deterioration in the others. Thus multi-objective design is defined largely by the need to *compromise* (de Weck, 2004). Quantifying, visualising and resolving such compromises can be thought of as one of the key duties of system designers, not least to avoid convergence to some sub-optimal configuration.

Optimising the performance of an engineering system may itself involve the consideration of any number of constituent sub-objectives. When such objectives are found to represent competing design interests[4] the design problem becomes the representation of the corresponding multi-dimensional trade-off surface, often expressed in the form of a continuous linear/non-linear relationship. This relationship is most commonly referred to as the Pareto front, after Vilfredo Pareto and his pioneering work analysing economic problems with mathematical methods (Pareto, 1919), building upon the initial work by Edgeworth (1881) exploring multi-criteria decision making and non-inferiority in economics. Should the design objective functions themselves be continuous, the Pareto front may potentially be comprised of an infinite number of possible design configurations, each representing a certain level of compromise between the included objectives.

A generic MOO problem may be written as (de Weck, 2004):

$$\begin{aligned} &\min_{\mathbf{x}\in S} \mathbf{J}(\mathbf{x},\mathbf{p}) \\ s.t. \quad &\mathbf{g}(\mathbf{x},\mathbf{p}) \leq 0 \\ &\mathbf{h}(\mathbf{x},\mathbf{p}) = 0 \end{aligned} \tag{5.7}$$

where $\mathbf{J} = [J_1(\mathbf{x}), \ldots, J_z(\mathbf{x})]^T$ is a column vector of z objectives conforming to $J_i \in \mathbb{R}$. Individual objectives are dependant on the vector of n design variables $\mathbf{x} = [x_1, \ldots, x_n]^T$, where $x_{i,LB} \leq x \leq x_{i,UB}\,(i = 1, \ldots, n)$, as well as a vector of fixed parameters $\mathbf{p}$. Design variables are assumed continuous and can be independently varied by the optimiser between lower and upper bounds $\mathbf{x}_{LB}$ and $\mathbf{x}_{UB}$ respectively. For a particular design to lie within the feasible domain S, the vectors of m_1 inequality constraints $\mathbf{g} = [g_1(\mathbf{x}), \ldots, g_{m_1}(\mathbf{x})]^T$ and

[4] Thus there will exist no unique solution optimal with respect to all included objectives.

m_2 equality constraints $\mathbf{h} = [h_1(\mathbf{x}), \ldots, h_{m_2}(\mathbf{x})]^T$ must both be satisfied. The problem is therefore to simultaneously minimise all elements of the objective vector $\mathbf{J}$.

Fundamentally, solving a Multi-Objective Optimisation problem enables a greater understanding of the numerical influence of the included objectives, thus informing future decision making. This recognises the fact that alternative design configurations, in practice, may be subject to selection based on objectives which may not be explicitly represented in the mathematical formulation of the problem, or perhaps even easily quantifiable at all (de Weck, 2004). Another useful application of this approach is when considering heavily constrained problems. Should a particular constraint be difficult to satisfy, it may be removed and re-introduced instead as an additional objective. The resulting multi-objective trade-off surface will then visualise the extent to which this constraint may be relaxed within some acceptable level of overall performance (Mezura-Montes and Coello, 2008).

The solution to a Multi-Objective Optimisation problem is often given as a finite set of solutions approximating the Pareto front. As individual solutions are still subject to the same complexities as in single-objective optimisation (e.g. the presence of non-linearity in objective functions and/or constraints, points of non-differentiability or discontinuity and the existence of local optima or multi-modality), numerical methods with which to solve multi-objective problems are most easily categorised according to the manner in which alternative objectives are handled/compared. More specifically, this choice often condenses to simply *when* the decision-maker articulates preference towards different objectives; never, before (*a priori*), during (*progressive*) or after (*a posteriori*) the optimisation procedure (Andersson, 2000). The two most common, fundamental high-level definitions are: scalarisation-based and dominance-based approaches. While the reader is referred to Andersson (2000) for a comprehensive overview and comparison of these methods, the following sections present a brief summary of each.

5.3.1 Scalarisation-based

Scalarisation reduces the multi-objective problem into a single (or series of) parametric single-objective sub-problem(s) (de Weck, 2004; Eichfelder, 2009), such that optimal solutions to the single-objective problem(s) are Pareto-optimal solutions to the multi-objective problem. Assuming that the problem objectives can in fact be meaningfully combined, this approach requires the specification of an overarching aggregate objective function (or dimensionless utility function) that contains the weighted contributions of each sub-objective, normally by using coefficients, exponents, constraint limits, etc (Gambier and Badreddin, 2007). Such weights thus represent the relative importance attributed to each objective. Accordingly, a single set of weights produces a single solution representing a particular compromise between the considered sub-objectives. To represent the associated trade-off surface through

multiple points then, a series of weights must be provided. With reference to the previously introduced classification (Andersson, 2000) then, this defines scalarisation as primarily an *a priori* preference expression method. However, it may be difficult to assign appropriate values of weights *a priori* in order to give a well spread representation of the Pareto front. Put simply, a uniformly spread set of weights may not result in a similarly uniform spread of solutions in objective space. Furthermore, a particular set of weights may entirely miss many interesting or relevant solutions, or indeed give no feasible solutions at all. However, specifying weights in this manner essentially defines a set of descent directions that each solution must follow, implying that any standard gradient-based NLP solver may be used (which also then guarantees local optimality). For these reasons, scalarisation approaches are typically more computationally efficient than dominance-based approaches for multi-objective problems.

Some well established scalarisation-based methods include (de Weck, 2004): the weighted sum approach (Cohon, 2004; Steuer, 1986), the Normalized Normal Constraints Method (Messac and Mattson, 2004), Compromise Programming (Non-linear combinations) (Ringuest, 1992), Multiattribute Utility Analysis (MAUA) (Keeney and Raiffa, 1993), Goal Programming (Steuer, 1986), Lexicographic Approaches, Acceptability Functions (Wallace et al., 1996; Kim and Wallace, 1997), and Fuzzy Logic (Lee and Li, 1993). Attempting to address the various practical issues with *a priori* preference articulation, adaptive methods for the selection of appropriate weights is an active area of research. For example, see the works (Cheng et al., 2016; Deb and Jain, 2014; Qi et al., 2014).

5.3.2 Dominance-based

Dominance-based methods[5], e.g. NSGA-II (Deb et al., 2002), MOPSO (Coello Coello and Lechuga, 2002) and multiMADS (Audet et al., 2010), are named after the dominance criterion. The dominance criterion is practically applied in order to distinguish between inferior and non-inferior solutions of a solution set represented in z-dimensional objective space (i.e. each element of the objective vector is separately considered). The dominance criterion may be mathematically defined as follows: Let $\mathbf{J}^1, \mathbf{J}^2 \in \mathbb{R}^z$ represent two feasible objective vectors. It is said that $\mathbf{J}^1$ dominates $\mathbf{J}^2$ (weakly) if:

$$\mathbf{J}^1 \leq \mathbf{J}^2 \text{ and } \mathbf{J}^1 \neq \mathbf{J}^2 \tag{5.8}$$

or, more precisely:

$$J_i^1 \leq J_i^2 \quad \forall i \tag{5.9}$$

$$\text{and } J_i^1 < J_i^2 \text{ for at least one } i \tag{5.10}$$

[5] also referred to as *Pareto*-methods.

For strong dominance, *all* elements of $\mathbf{J}^1$ would have to be better (i.e. $\mathbf{J}^1 < \mathbf{J}^2$) than the corresponding elements of $\mathbf{J}^2$. A feasible solution vector is said to Pareto optimal if there is no other feasible solution representing an improvement in one objective function without simultaneously exhibiting a deterioration in at least one other (Rao, 2009). A set of Pareto optimal solutions thus represents a finite approximation of the Pareto front.

Dominance methods do not require any specification of weights vectors. However, as the criterion itself in non-differentiable, knowledge of all other points composing the current approximation of the Pareto front is required. The simplest approaches are therefore a combination of design space exploration and dominance filtering (also referred to as Pareto filtering). However, these methods operate around evaluating a full-factorial exploration of the design space, which may be wholly infeasible (in terms of computational expense) for a problem of n-dimensionality. Secondly, solutions obtained in this manner remain only approximations of the Pareto front, in that they only satisfy non-dominance (de Weck, 2004).

Prevalent dominance-based methods include the *Normal Boundary Intersection* (NBI) method (Das and Dennis, 1998) and the wider class of *Evolutionary Algorithms* (EAs) (or, population-based *heuristic* algorithms), including methods such as *Multi-Objective Genetic Algorithms* (MOGA), *Particle Swarm Optimisation* (PSO) and *Multi-Objective Simulated Annealing* (MOSA). The NBI method can produce a near uniform representation of the Pareto front. However, due to reliance on equality constraints normal to a line connecting the anchor points (e.g. $\mathbf{x}^{1*}$ and $\mathbf{x}^{2*}$) in objective space, NBI will converge to local optima for complex, non-linear problems resulting in dominated or non-Pareto points. Evolutionary (or heuristic) algorithms typically evolve populations of designs gradually such that the Pareto frontier may be approximated as closely as possible. However, the large number of repeated function evaluations required for these approaches induces a large, often prohibitive, computational expense as well as a tendency for *niching* (clumping of solutions in objective space). This can ultimately result in underrepresented regions of the Pareto front.

Another class of *a posteriori* preference expression methods involves adaptively selecting and updating search preferences during the optimisation process itself. In this type of approach, an initial, lower quality approximation of the Pareto front is constructed using a small number of points. This initial approximation is then used to identify and exploit unexplored or poorly represented regions of objective space. An example is the *Adaptive Weighted Sum* (AWS) method proposed by de Weck (2004). By concentrating computational effort within promising regions (where information may be most easily obtained), the algorithm iteratively builds and advances the current approximation. Furthermore, the adaptive selection of points ensures that the final approximation is composed of a well-spread set of solutions.

5.3.3 Others

Indicator-based approaches aim to maximise some relevant scalar metric of the approximated Pareto front, thus effectively acting as an intermediate between scalarisation and dominance based approaches. A complicating factor is that most useful metrics require the specification of a reference Pareto set of solutions, which may not be known *a priori*, particularly for non-academic problems. The only metric that does not include this caveat is the hypervolume indicator. Methods utilising this latter metric include SMS-EMOEA (Beume et al., 2007) and HypE (Bader and Zitzler, 2011), though it has since been established that the hypervolume indicator in fact influences the selection of points comprising the current approximation of the Pareto front, favouring the central or convex regions (Auger et al., 2009). A further difficulty is the computational expense of computing the hypervolume, which itself increases exponentially relative to the number of objectives. Methods to approximate the hypervolume metric exist, and are indeed represent an active area of research into improving the computational efficiency of the indicator-based approach (Fonseca et al., 2006).

5.4 Multi-Objective Optimal Control

In reference to the discussion in Section 5.1, it has been established that the performance of a dynamic, controllable system may be equally dependant on the (time dependant) control law by which it must operate as it is on static (time independent) system design parameters. It follows that the same system may be controlled in alternative ways to optimise performance according to alternative objectives. Where these objectives are contradictory and/or incommensurable, yet all deemed integral to the design definition (and therefore must be simultaneously considered), the problem may be formulated as a *Multi-Objective Optimal Control Problem* (MOCP). A common example is minimising the total time of operation whilst simultaneously minimising the energy consumption (control effort). This class of problem exists as an extension to the previously discussed MOO problem, as static design parameters may be incorporated in addition to the associated time-dependant control law. Problems of this type arise in many applications where the complexity of the system under consideration precludes performance evaluation relative to only one single metric. A common example in aerospace engineering is spacecraft trajectory design, where objectives may include (but are not limited to) minimising consumed propellant as well as the transition time between two given orbits. Examples within other fields include the design of optimal reactor feeding rates in (bio)chemical engineering, optimal power management of fuel cells in electrical engineering and optimal robot paths in mechanical engineering.

Considering MOCPs, for much the same reasons as for general MOO problems, it is often far more desirable to the designer to achieve a set of compromise solutions, as opposed to one single solution. Classical methods of generating such a Pareto-optimal set primarily centre

around scalarisation-based methods (Section 5.3.1), such as the weighted sum approach. The Bolza problem (see Section 6.1), which defines the overarching objective as a sum of two functions, one dependant on the system boundary states and the other on an integral function computed along the complete trajectory, is another example of such. In Section 5.3, it was stated that weighted sum approaches exhibit a number of practical limitations. The first, and often the most significant, is that they are unable to converge to non-convex areas of the Pareto front (Kaya and Maurer, 2014). Secondly, as with general scalarisation methods, it is difficult to estimate *a priori* how the positioning of solutions composing the Pareto front is influenced by changes in weights, even if only minor. This problem is especially concerning should the objectives be defined over largely different scales.

Recent work has studied the multi-objective adaptation of the KKT optimality conditions and related theoretical aspects of multi-objective optimal control (Kaya and Maurer, 2014; Zhu, 2000; Kien et al., 2012; de Oliveira and Silva, 2016; Ngo and Hayek, 2017). Methods with which to solve MOCPs have similarly been the subject of much study. Ober-Blobaum et al. (2012) solve an interplanetary trajectory optimisation problem by coupling a direct transcription method with a scalarisation-based approach. Predefined 'unreachable' points in criteria space are used to determine the search direction vectors that may be assigned to individual multi-objective solutions. Each solution is then solved using a standard gradient-based solver. A similar approach is employed by Kaya and Maurer (2014), where Pascoletti-Serafini scalarisation (Pascoletti and Serafini, 1984) is used to transform the multi-objective problem into a set of single-objective problems. The complete approach is applied to solve chemical reaction engineering and drug dosage problems. Logist et al. (2009) combine direct transcription methods with various scalarisation techniques such as the Normalised Normal Constraint method (Messac and Mattson, 2004) and the NBI method (Das and Dennis, 1998) in order to approximate the Pareto-front of non-convex problems, specifically that of a bio-chemical process. Examples in the field of aerospace engineering include the works by Bairstow et al. (2006) and Roshanian et al. (2018), which both optimise the ascent phase of a two-stage to orbit launcher. The former solves to represent the trade off between payload-to-orbit and the associated cost, whilst the latter minimises both the mean and variance of the gross take-off mass where several design and operative parameters where subject to uncertainty. In both works, the control-law was represented as a predefined parametric shape, the parameters of which included as optimisation variables.

Stochastic methods have also been employed for solving MOCPs. For example, Coverstone-Carroll et al. (2000) defined a dual-loop algorithm incorporating both Optimal Control theory and GAs. A *Multi-Objective Genetic Algorithms* (MOGA) was employed in the outer-loop to generate vectors of co-states and times of flight. Each time of flight was then optimised within the inner-loop with the goal to minimise the corresponding propellant consumption. The dual-

loop approach was also demonstrated by Englander et al. (2019). In this case, the outer-loop employed a MOGA to handle the multi-objective problem through a set of categorical variables and the inner-loop solved a set of single objective constrained Optimal Control problems using Monotonic Basin Hopping (Leary, 2000).

5.5 Adaptive Mesh Refinement for Multi-Objective Optimal Control

The importance of multi-objective formulations in modern design has been discussed several times in this work already (see Sections 2.2.6 and 5.3). However, MOCPs, especially those involving complex, non-linear or discontinuous solutions, inherently induce a significantly high computational burden. In Section 5.2, mesh refinement methods for direct Optimal Control problems were discussed in the context of improving computational efficiency, i.e. obtaining results of higher numerical accuracy at a reduced cost. Such methods in practice are most often applied to single-objective problems (Zhao and Li, 2019; Ross and Fahroo, 2004; Gong and Ross, 2006; Darby et al., 2011; Patterson et al., 2015). For example, Zhao and Li (2019) employ an adaptive mesh refinement algorithm in combination with a local direct collocation method to optimise the descent trajectory of a Mars re-entry vehicle with respect to various boundary and path constraints. Another example is the work by Ross and Fahroo (2004) that divides an Optimal Control problem into phases by specifying the location of the phase boundaries (knots) at non-smooth locations. Knots may be either fixed or free, where free knots are part of the optimisation process. Knot locations are first estimated by manually checking the solution of a preliminary optimisation, then optimised in a second optimisation. This removes the restriction of continuity in the solution across the end points of neighbouring phases. Each phase is solved using either the Legendre or Chebyshev pseudospectral method, which simultaneously collates the grid points about the knots (the grid points being the roots of orthogonal polynomials, which are by definition denser towards the ends of the specified phase). The proposed methodology is applied to an example problem maximising the final altitude of a two-stage vertically launched rocket. Also considering pseudospectral methods, Gong and Ross (2006) allow the number of nodes to be added to a particular phase (if the error between two successive iterations is greater than a prescribed threshold) to be either specified *a priori* or determined automatically. The location of the knots is (approximately) determined using the gradient of the control as part of a pseudospectral differentiation matrix that locates switches, kinks, corners, and other discontinuities. Another example is the *hp*-adaptive pseudospectral method presented by Darby et al. (2011), which iteratively determines the number of elements and the width/polynomial degree of each to obtain a solution within a user-specified accuracy. The proposed methodology is applied to a series of single-objective problems sourced from the open literature. Patterson et al. (2015) presented a similar mesh

refinement method (the *ph* method), which attempts to reach the prescribed error tolerance by first increasing the degree of the approximating polynomials then, if unsuccessful, by increasing the number of mesh elements.

Despite this depth of literature, mesh refinement methods for MOCPs are far more scarce. This is likely due to one significant complicating factor: the orthogonal objective solutions may exhibit non-linear/non-smooth behaviour within entirely different regions of the time domain. With reference to the prescribed grid resolution problem, this implies that to properly represent a multi-objective trade-off in a numerically efficient manner, individual solutions of a Pareto-optimal set must be separately discretized upon their own unique, non-uniform grids.

Some authors have attempted to navigate this complication through the application of *scalarisation* methods Andersson (2000), reducing the multi-objective problem into a (or series of) single-objective sub-problem(s) Yang and Qi (2016); Tsiantis and Banga (2020); Zou et al. (2018). Scalarisation (see Section 5.3.1) requires the specification of an overarching aggregate objective function (or dimensionless utility function) that contains the weighted contributions of each sub-objective (representing the relative importance attributed to each objective de Weck (2004)). Accordingly, a single weights vector produces a single solution representing a certain level of compromise between sub-objectives. Solutions treated in isolation thus allow the simple incorporation of iterative mesh refinement schemes. To represent a trade-off surface however, a series of weights must be provided (as seen in Santos et al. (2018)), and thus a series of separate refinement procedures must follow. This approach is seen in Yang and Qi (2016), where the authors use a Runge-Kutta discretisation method to transcribe the problem into an NLP problem with solution states and controls defined upon refinable dyadic grids. The scalarised aggregate function is then solved using a local optimisation method.

However, it may be difficult to assign appropriate values of weights *a priori* in order to give a well spread representation of the Pareto front. Put simply, a uniformly spread set of weights may not result in a similarly uniform spread of solutions in objective space. Furthermore, a particular set of weights may entirely miss many interesting or relevant solutions, or indeed give no feasible solutions at all. Ding et al. (2017) attempt to circumvent this particular issue by replacing the weights-based approach with a Physical Programming method (Messac, 1996). In this case, the overall cost function is determined by combining each of the four included objectives using a series of preference functions. Applying this method in conjunction with a *hp*-adaptive pseudospectral method allows the optimiser to produce a number of solutions, each defined on a unique non-uniform grid, equal to the number of defined preference functions. This approach then is still inherently limited in terms of the detail to which the Pareto front may be represented.

In the multi-objective literature, the issues associated with scalarisation can be mitigated or

entirely avoided by employing dominance-based approaches (see Section 5.3.2) as an alternative to scalarisation.

A common approach to dominance-based optimisation is the *Evolutionary Algorithm* (EA). Inspired by the mechanics of natural selection, a typical EA iteratively processes a population of memetic individuals according to a series of evolutionary operators designed to mimic fitness-based selection and reproduction.

An issue throughout the field of EAs is that the restriction of flexibility in genotype length/-structure required by standard combinatorial operators naturally precludes their use in more complex design scenarios (Harvey, 1992). Indeed, a traditional fixed-length representation in computational evolution may altogether sacrifice certain forms of genotypic variation otherwise common in natural genomes, thus potentially compromising the dynamics of the evolutionary process as a whole (Merlevede et al., 2019). Examples of such are the duplication or removal of certain genes/sequences (a crucial process for the emergence of new genes in natural evolution (Long et al., 2003), and intimately linked to genetic properties such as robustness, evolvability, and functional specialisation (Calabretta et al., 2000)), or the structural relocation of genes across the genotype (Merlevede et al., 2019).

From the computational perspective, variable-length representations of the problem genome (Koza, 1992) allow evolutionary modelling to mimic these phenomena to a higher degree. In consequence, this facilitates a more natural exploration of open-ended design problems where solution complexity and/or dimensionality are in themselves adaptive unknowns (Banzhaf et al., 2006).

This may be particularly relevant when considering problems where solution fidelity/feasibility is inherently linked to the size of the associated genotype. For example, path optimisation problems where the number of waypoints is not known *a priori* (Qiongbing and Lixin, 2016; Cruz-Piris et al., 2019; Rodríguez-Molina et al., 2021), systems design problems where the number of discrete components is a optimisable parameter (Hitomi and Selva, 2018; Decraene et al., 2011; Aliefa and Suyanto, 2020), or direct optimal control problems requiring adaptive mesh refinement to capture discontinuous/non-smooth behaviour in the system dynamics (Betts and Huffman, 1998; Jain and Tsiotras, 2008; Zhao and Li, 2017).

5.6 Conclusion

This chapter has offered a brief introduction to several areas of research relevant to adaptive mesh refinement for MOCPs. This is a complex issue with several interconnected components, thus the individual sections of Chapter 5 gradually introduce each component in turn.

Section 5.1 begins with a brief overview of solving *Optimal Control* (OC) problems and an

introduction to the two primary classes of approach: indirect and direct methods. Indirect methods were identified as largely unsuitable for this study, in part due to the added complexity in defining the necessary conditions for optimality. Discussion is therefore focused on direct methods, with prevalent examples (such as single-shooting, multi-shooting and collocation-based approaches) given. Global collocation methods are considered to be particularly attractive for the purposes of this study due to the offered flexibility in representing the problem of interest, and without relying on numerical integration, amongst other reasons.

However, global collocation methods (and indeed direct methods in general) can exhibit difficult in approximating non-smooth or discontinuous profiles, common features of complex engineering design problems, particularly involving multiple disciplines. This difficulty relates to the number and distribution of grid nodes across which the time varying state/control profiles are constructed. Section 5.2 therefore discusses several approaches present in the literature for the adaptive refinement of these grid points. Of these, methods exploiting dyadic grid structures were found to be effective and potentially suitable for incorporation into this work, though outside of this work had only been applied to single-objective problems.

Next, Section 5.3 introduces *Multi-Objective Optimisation* (MOO) and discusses the two most prevalent classes of approach to solving MOO problems: scalarisation-based and dominance-based methods. The extension of MOO methodology to include *Optimal Control* (OC) problems, forming *Multi-Objective Optimal Control Problems* (MOCPs), is then discussed in Section 5.4. This section briefly reviews examples of scalarisation and dominance-based methods in application to MOCPs.

Finally, Section 5.5 discusses the fact that direct methods for OC are just as susceptible to issues related to highly non-linear or non-smooth/discontinuous behaviour whether the problem is single-objective or multi-objective. One may employ mesh refinement strategies to address these issues, but a particular difficulty faced by MOCPs is that the location and scope of the necessary refinement may differ between optimal solutions lying at different locations along the Pareto-front, often dramatically. This issue becomes non-trivial when then considering that to properly represent the trade-off surface of such a problem, a set of solutions is required where each has been independently refined with respect to its own position in objective space, and thus must be defined with a unique grid discretisation. This issue has yet to be adequately addressed in the literature and as such forms the research basis for the remainder of this thesis.

5.7 Bibliography

M. H. Aliefa and S. Suyanto. Variable-Length Chromosome for Optimizing the Structure of Recurrent Neural Network. *2020 International Conference on Data Science and Its Applications, ICoDSA 2020*, 2020. doi: 10.1109/ICoDSA50139.2020.9213012.

J. Andersson. *A survey of multiobjective optimization in engineering design.* PhD thesis, Linktjping University, Sweden, 2000.

F. Assassa and W. Marquardt. Dynamic optimization using adaptive direct multiple shooting. *Computers and Chemical Engineering*, 60:242–259, 2014. ISSN 00981354. doi: 10.1016/j.compchemeng.2013.09.017. URL http://dx.doi.org/10.1016/j.compchemeng.2013.09.017.

C. Audet, G. Savard, and W. Zghal. A mesh adaptive direct search algorithm for multiobjective optimization. *European Journal of Operational Research*, 204(3):545–556, 2010. ISSN 03772217. doi: 10.1016/j.ejor.2009.11.010. URL http://dx.doi.org/10.1016/j.ejor.2009.11.010.

A. Auger, J. Bader, D. Brockhoff, and E. Zitzler. Theory of the hypervolume indicator: Optimal μ-distributions and the choice of the reference point. *Proceedings of the 10th ACM SIGEVO Workshop on Foundations of Genetic Algorithms, FOGA'09*, pages 87–102, 2009. doi: 10.1145/1527125.1527138.

I. Babuška and M. Suri. The p-and hp versions of the finite element method, an overview. *Computer methods in applied mechanics and engineering*, 80(1-3):5–26, 1990. ISSN 0045-7825. doi: https://doi.org/10.1016/0045-7825(90)90011-A.

J. Bader and E. Zitzler. HypE : An Algorithm for Fast Hypervolume-Based. *Evolutionary computation*, 19(1):45–76, 2011.

B. K. Bairstow, O. De Weck, and J. Sobieszczanski-Sobieski. Multiobjective optimization of two-stage rockets for earth-to-orbit launch. *Collection of Technical Papers - AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, 3(May):1585–1600, 2006. ISSN 02734508. doi: 10.2514/6.2006-1720.

W. Banzhaf, G. Beslon, S. Christensen, J. A. Foster, F. Képès, V. Lefort, J. F. Miller, M. Radman, and J. J. Ramsden. Guidelines: From artificial evolution to computational evolution: A research agenda. *Nature Reviews Genetics*, 7(9):729–735, 2006. ISSN 14710056. doi: 10.1038/nrg1921.

R. E. Bellman. Dynamic Programming. 1957.

R. E. Bellman and S. E. Dreyfus. Applied dynamic programming. 1962.

J. T. Betts. Survey of Numerical Methods for Trajectory Optimization. *Journal of Guidance, Control, and Dynamics*, 21(2):193–207, 1998. ISSN 0731-5090. doi: 10.2514/2.4231.

J. T. Betts. *Practical methods for optimal control and estimation using nonlinear programming*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2010. ISBN 978-0-89871-688-7. doi: https://doi.org/10.1137/1.9780898718577.

J. T. Betts and P. D. Frank. A sparse nonlinear optimization algorithm. *Journal of Optimization Theory and Applications*, 82(3):519–541, 1994. ISSN 00223239. doi: 10.1007/BF02192216.

J. T. Betts and W. P. Huffman. Mesh refinement in direct transcription methods for optimal control. *Optimal Control Applications and Methods*, 19(1):1–21, 1998.

J. T. Betts, N. Biehn, S. L. Campbell, and W. P. Huffman. Compensating for order variation in mesh refinement for direct transcription methods II: Computational experience. *Journal of Computational and Applied Mathematics*, 143(2):237–261, 2002. ISSN 03770427. doi: 10.1016/S0377-0427(01)00509-X.

N. Beume, B. Naujoks, and M. Emmerich. SMS-EMOA: Multiobjective selection based on dominated hypervolume. *European Journal of Operational Research*, 181(3):1653–1669, 2007. ISSN 03772217. doi: 10.1016/j.ejor.2006.08.008.

T. Binder, L. Blank, W. Dahmen, and W. Marquardt. Grid refinement in multiscale dynamic optimization. *Computer Aided Chemical Engineering*, 8(C):31–36, 2000. ISSN 15707946. doi: 10.1016/S1570-7946(00)80007-3.

G. A. Bliss. *Lectures on the Calculus of Variations*. Number 504. University of Chicago press, 1963.

A. E. Bryson. Optimal control-1950 to 1985. *IEEE Control Systems Magazine*, 16(3):26–33, 1996.

D. E. Bryson and A. Earl. *Applied optimal control: optimization, estimation and control*. Routledge, 2018.

R. H. Byrd, J. Nocedal, and R. A. Waltz. Knitro: An Integrated Package for Nonlinear Optimization. pages 35–59, 2006. doi: 10.1007/0-387-30065-1_4.

R. Calabretta, S. Nolfi, D. Parisi, and G. P. Wagner. Duplication of modules facilitates the evolution of functional specialization. *Artificial Life*, 6(1):69–84, 2000. ISSN 10645462. doi: 10.1162/106454600568320.

R. Cheng, Y. Jin, M. Olhofer, and B. Sendhoff. A Reference Vector Guided Evolutionary Algorithm for Many-Objective Optimization. *IEEE Transactions on Evolutionary Computation*, 20(5):773–791, 2016. ISSN 1089778X. doi: 10.1109/TEVC.2016.2519378.

C. A. Coello Coello and M. S. Lechuga. MOPSO: A proposal for multiple objective particle swarm optimization. *Proceedings of the 2002 Congress on Evolutionary Computation, CEC 2002*, 2:1051–1056, 2002. doi: 10.1109/CEC.2002.1004388.

J. L. Cohon. *Multiobjective programming and planning*, volume 140. Courier Corporation, 2004.

V. Coverstone-Carroll, J. W. Hartmann, and W. J. Mason. Optimal multi-objective low-thrust spacecraft trajectories. *Computer Methods in Applied Mechanics and Engineering*, 186(2-4): 387–402, 2000. ISSN 00457825. doi: 10.1016/S0045-7825(99)00393-X.

L. Cruz-Piris, I. Marsa-Maestre, and M. A. Lopez-Carmona. A Variable-Length Chromosome Genetic Algorithm to Solve a Road Traffic Coordination Multipath Problem. *IEEE Access*, 7: 111968–111981, 2019. ISSN 21693536. doi: 10.1109/ACCESS.2019.2935041.

J. E. Cuthrell and L. T. Biegler. On the optimization of differential-algebraic process systems. *AIChE Journal*, 33(8):1257–1270, 1987. ISSN 15475905. doi: 10.1002/aic.690330804.

J. E. Cuthrell and L. T. Biegler. Simultaneous optimization and solution methods for batch reactor control profiles. *Computers and Chemical Engineering*, 13(1-2):49–62, 1989. ISSN 00981354. doi: 10.1016/0098-1354(89)89006-4.

C. L. Darby, W. W. Hager, and A. V. Rao. An hp-adaptive pseudospectral method for solving optimal control problems. *Optimal Control Applications and Methods*, 32(4):476–502, 2011. ISSN 01432087. doi: 10.1002/oca.957.

I. Das and J. Dennis. Normal-boundary intersection: A new method for generating the pareto surface in nonlinear multicriteria optimization problems. *SIAM Journal of Optimisation*, 8 (3):631–657, 1998.

V. A. de Oliveira and G. N. Silva. On sufficient optimality conditions for multiobjective control problems. *Journal of Global Optimization*, 64(4):721–744, 2016. ISSN 15732916. doi: 10.1007/s10898-015-0351-y.

O. L. de Weck. Multiobjective Optimization: History and Promise. In *Invited Keynote Paper, GL2-2, The Third China-Japan-Korea Joint Symposium on Optimization of Structural and Mechanical Systems, Kanazawa, Japan*, 2004.

K. Deb and H. Jain. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, Part I: Solving problems with box constraints. *IEEE Transactions on Evolutionary Computation*, 18(4):577–601, 2014. ISSN 1089778X. doi: 10.1109/TEVC.2013.2281535.

K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002. ISSN 1089778X. doi: 10.1109/4235.996017.

J. Decraene, M. Chandramohan, F. Zeng, M. Yoke Hean Low, and W. Cai. Evolving Agent-Based Model Structures using Variable-Length Genomes. In *Proceedings of the Fourth International workshop on Optimisation in Multi-Agent Systems*, 2011. URL http://pdcc.ntu.edu.sg/EVOSIM/.

J. E. Dennis Jr and R. B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*. SIAM, 1996.

D. Ding, H. Guo, L. Huang, and Y. Wang. Multi-objective optimization of reentry trajectory for Hypersonic Gliding Vehicle. *CGNCC 2016 - 2016 IEEE Chinese Guidance, Navigation and Control Conference*, pages 1436–1443, 2017. doi: 10.1109/CGNCC.2016.7829000.

F. Y. Edgeworth. *Mathematical psychics: An essay on the application of mathematics to the moral sciences*, volume 10. CK Paul, 1881.

G. Eichfelder. An Adaptive Scalarization Method in Multiobjective Optimisation. *SIAM Journal of Optimisation*, 19(4):1694–1718, 2009.

J. A. Englander, M. A. Vavrina, and A. R. Ghosh. Aas 15-227 Multi-Objective Hybrid Optimal Control for Multiple-Flyby Low-Thrust Mission Design. pages 1–20, 2019.

P. J. Enright and B. A. Conway. Discrete approximations to optimal trajectories using direct transcription and nonlinear programming. *Journal of Guidance, Control, and Dynamics*, 15 (4):994–1002, 1992. ISSN 07315090. doi: 10.2514/3.20934.

F. Fahroo and I. M. Ross. A spectral patching method for direct trajectory optimization. *The Journal of the Astronautical Sciences*, 48:269–286, 2000.

F. Fahroo and I. M. Ross. Convergence of the costates does not imply convergence of the control. *Journal of Guidance, Control, and Dynamics*, 31(5):1492–1496, 2008. ISSN 15333884. doi: 10.2514/1.37331.

R. Fletcher. *Practical methods of optimization*. John Wiley & Sons, 2013.

C. M. Fonseca, L. Paquete, and M. López-Ibáñez. An improved dimension-sweep algorithm for the hypervolume indicator. *2006 IEEE Congress on Evolutionary Computation, CEC 2006*, pages 1157–1163, 2006. doi: 10.1109/cec.2006.1688440.

A. Gambier and E. Badreddin. Multi-objective Optimal Control: An Overview. (October):1–3, 2007.

P. E. Gill, W. Murray, and M. A. Saunders. SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM Review*, 47(1):99–131, 2005. ISSN 00361445. doi: 10.1137/S0036144504446096.

H. H. Goldstine. A history of the calculus of variations from the 17th through the 19th century. 1980.

Q. Gong and I. M. Ross. Autonomous pseudospectral knotting methods for space mission optimization. In *16th AAS/AIAA Space Flight Mechanics Meeting, American Astronautical Society Paper*, pages 06–151, 2006.

I. Harvey. The SAGA Cross: The Mechanics of Recombination for Species with Variable-Length Genotypes. *Parallel Problem Solving from Nature*, 2:269–278, 1992.

A. L. Herman and B. A. Conway. Direct optimization using collocation based on high-order Gauss-Lobatto quadrature rules. *Journal of Guidance, Control, and Dynamics*, 19(3):592–599, 1996. ISSN 15333884. doi: 10.2514/3.21662.

N. Hitomi and D. Selva. Constellation optimization using an evolutionary algorithm with a variable-length chromosome. *IEEE Aerospace Conference Proceedings*, 2018-March:1–12, 2018. ISSN 1095323X. doi: 10.1109/AERO.2018.8396743.

D. G. Hull. Conversion of optimal control problems into parameter optimization problems. *Journal of guidance, control, and dynamics*, 20(1):57–60, 1997.

S. Jain and P. Tsiotras. Trajectory optimization using multiresolution techniques. *Journal of Guidance, Control, and Dynamics*, 31(5):1424–1436, 2008. ISSN 15333884. doi: 10.2514/1.32220.

S. Kameswaran and L. T. Biegler. Convergence rates for direct transcription of optimal control problems using collocation at Radau points. *Computational Optimization and Applications*, 41(1):81–126, 2008. ISSN 09266003. doi: 10.1007/s10589-007-9098-9.

C. Y. Kaya and H. Maurer. A numerical method for nonconvex multi-objective optimal control problems. *Computational Optimization and Applications*, 57(3):685–702, 2014. ISSN 15732894. doi: 10.1007/s10589-013-9603-2.

R. L. Keeney and H. Raiffa. *Decisions with multiple objectives: preferences and value trade-offs*. Cambridge university press, 1993.

H. B. Keller. *Numerical methods for two-point boundary-value problems*. Courier Dover Publications, 1968.

B. T. Kien, N. C. Wong, and J. C. Yao. Necessary conditions for multiobjective optimal control problems with state constraints. *Dynamics of Continuous, Discrete and Impulsive Systems Series B: Applications and Algorithms*, 19(3):431–446, 2012. ISSN 14928760.

J. B. Kim and D. Wallace. A goal-oriented design evaluation model. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 80456, page V003T30A014. American Society of Mechanical Engineers, 1997.

J. R. Koza. The genetic programming paradigm: Genetically breeding populations of computer programs to solve problems. *Dynamic, Genetic, and Chaotic Programming*, (June):203–321, 1992. doi: doi:10.1109/TAI.1990.130444.

H. Kwakernaak, R. Sivan, and B. N. D. Tyreus. Linear optimal control systems. 1974.

R. H. Leary. Global Optimization on Funneling Landscapes. *Journal of Global Optimization*, 18 (4):367–383, 2000. ISSN 09255001. doi: 10.1023/A:1026500301312.

E. S. Lee and R. J. Li. Fuzzy multiple objective programming and compromise programming with Pareto optimum. *Fuzzy sets and systems*, 53(3):275–288, 1993.

B. Li, H. Zhang, W. Zheng, and L. Wang. Spacecraft close-range trajectory planning via convex optimization and multi-resolution technique. *Acta Astronautica*, 175(November 2019):421–437, 2020. ISSN 00945765. doi: https://doi.org/10.1016/j.actaastro.2020.05.051.

F. Logist, P. M. Van Erdeghem, and J. F. Van Impe. Efficient deterministic multiple objective optimal control of (bio)chemical processes. *Chemical Engineering Science*, 64(11):2527–2538, 2009. ISSN 00092509. doi: 10.1016/j.ces.2009.01.054.

M. Long, E. Betrán, K. Thornton, and W. Wang. The origin of new genes: Glimpses from the young and old. *Nature Reviews Genetics*, 4(11):865–875, 2003. ISSN 14710056. doi: 10.1038/nrg1204.

M. W. Maier. *The art of systems architecting*. CRC press, 2009.

A. Merlevede, H. Åhl, and C. Troein. Homology and linkage in crossover for linear genomes of variable length. *PLoS ONE*, 14(1):1–16, 2019. ISSN 19326203. doi: 10.1371/journal.pone.0209712.

A. Messac. Physical programming: Effective optimization for computational design. *AIAA Journal*, 34(1):149–158, 1996. ISSN 00011452. doi: 10.2514/3.13035.

A. Messac and C. A. Mattson. Normal constraint method with guarantee of even representation of complete pareto frontier. *AIAA Journal*, 42(10):2101–2111, 2004. ISSN 00011452. doi: 10.2514/1.8977.

E. Mezura-Montes and C. A. C. Coello. Constrained optimization via multiobjective evolutionary algorithms. *Multiobjective Problem Solving from Nature: From Concepts to Applications*, pages 53–75, 2008.

A. Miele. Recent advances in gradient algorithms for optimal control problems. *Journal of Optimization Theory and Applications*, 17:361–430, 1975.

T. N. Ngo and N. Hayek. Necessary conditions of Pareto optimality for multiobjective optimal control problems under constraints. *Optimization*, 66(2):149–177, 2017. ISSN 10294945. doi: 10.1080/02331934.2016.1261349.

S. Ober-Blobaum, M. Ringkamp, and G. Zum Felde. Solving multiobjective Optimal Control problems in space mission design using Discrete Mechanics and reference point techniques. *Proceedings of the IEEE Conference on Decision and Control*, pages 5711–5716, 2012. ISSN 25762370. doi: 10.1109/CDC.2012.6426285.

V. Pareto. *Manuale di economia politica: con una introduzione alla scienza sociale*, volume 13. Società editrice libraria, 1919.

A. Pascoletti and P. Serafini. Scalarizing vector optimization problems. *Journal of Optimization Theory and Applications*, 42(4):499–524, 1984. ISSN 00223239. doi: 10.1007/BF00934564.

M. A. Patterson, W. W. Hager, and A. V. Rao. A ph mesh refinement method for optimal control. *Optimal Control Applications and Methods*, 36(4):398–421, 2015. ISSN 10991514. doi: 10.1002/oca.2114.

E. Polak. An Historical Survey of Computational Methods in Optimal Control. 15(2):553–584, 1973. doi: https://doi.org/10.1137/1015071.

Y. Qi, X. Ma, F. Liu, L. Jiao, J. Sun, and J. Wu. MOEA/D with Adaptive Weight Adjustment. *Evolutionary Computation*, 22(2):231–264, 2014. ISSN 15309304. doi: 10.1162/EVCO_a_00109.

Z. Qiongbing and D. Lixin. A new crossover mechanism for genetic algorithms with variable-length chromosomes for path optimization problems. *Expert Systems with Applications*, 60: 183–189, 2016. ISSN 09574174. doi: 10.1016/j.eswa.2016.04.005.

A. Rao. A Survey of Numerical Methods for Optimal Control. *Solution Methods for Integral Equations*, pages 1–58, 2013. doi: 10.1007/978-1-4757-1466-1_1.

S. S. Rao. *Engineering Optimization: Theory and Practice*, volume 56. 2009. ISBN 9780470183526. doi: 10.1080/03052150500066646.

G. W. Reddien. Collocation at Gauss points as a discretization in optimal control. *SIAM Journal on Control and Optimization*, 17(2):298–306, 1979.

J. L. Ringuest. Compromise programming. *Multiobjective optimization: behavioral and computational considerations*, pages 51–59, 1992.

A. Rodríguez-Molina, J. Solís-Romero, M. G. Villarreal-Cervantes, O. Serrano-Pérez, and G. Flores-Caballero. Path-Planning for Mobile Robots Using a Novel Variable-Length Differential Evolution Variant. *Mathematics*, 9(4):357, 2021. ISSN 22277390. doi: 10.3390/math9040357.

J. Roshanian, A. A. Bataleblu, and M. Ebrahimi. Robust ascent trajectory design and optimization of a typical launch vehicle. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 232(24):4601–4614, 2018. ISSN 20412983. doi: 10.1177/0954406217753460.

I. M. Ross and F. Fahroo. Pseudospectral knotting methods for solving optimal control problems. *Journal of Guidance, Control, and Dynamics*, 27(3):397–405, 2004. ISSN 15333884. doi: 10.2514/1.3426.

I. M. Ross and M. Karpenko. A review of pseudospectral optimal control: From theory to flight. *Annual Reviews in Control*, 36(2):182–197, 2012.

L. d. S. Santos, K. M. F. de Souza, M. R. Bandeira, V. R. Ruiz Ahón, F. C. Peixoto, and D. M. Prata. Dynamic optimization of a continuous gas lift process using a mesh refining sequential method. *Journal of Petroleum Science and Engineering*, 165(September 2017):161–170, 2018. ISSN 09204105. doi: 10.1016/j.petrol.2018.02.019.

M. Schlegel, K. Stockmann, T. Binder, and W. Marquardt. Dynamic optimization using adaptive control vector parameterization. *Computers and Chemical Engineering*, 29(8):1731–1751, 2005. ISSN 00981354. doi: 10.1016/j.compchemeng.2005.02.036.

A. L. Schwartz. *Theory and implementation of numerical methods based on Runge-Kutta integration for solving optimal control problems*. PhD thesis, University of California, Berkeley, 1996. URL http://www2.eecs.berkeley.edu/Pubs/TechRpts/1996/2995.html.

R. E. Steuer. Multiple criteria optimization. *Theory, Computation, and Application*, 1986.

H. J. Sussmann and J. C. Willems. 300 Years of Optimal Control: From The Brachystochrone to the Maximum Principle, 1997. ISSN 1066033X.

N. Tsiantis and J. R. Banga. Using optimal control to understand complex metabolic pathways. *BMC Bioinformatics*, 21(1):1–33, 2020. ISSN 14712105. doi: 10.1186/s12859-020-03808-8.

O. von Stryk and R. Bulirsch. Direct and indirect methods for trajectory optimization. *Annals of Operations Research*, 37(1):357–373, 1992. ISSN 02545330. doi: 10.1007/BF02071065.

D. R. Wallace, M. J. Jakiela, and W. C. Flowers. Design search under probabilistic specifications using genetic algorithms. *Computer-Aided Design*, 28(5):405–421, 1996.

P. Yang and R. Qi. Reentry trajectory optimization for hypersonic vehicle based on improved mesh refinement techniques. *Chinese Control Conference, CCC*, 2016-Augus(61374116): 5473–5478, 2016. ISSN 21612927. doi: 10.1109/ChiCC.2016.7554207.

J. Zhao and S. Li. Modified multiresolution technique for mesh refinement in numerical optimal control. *Journal of Guidance, Control, and Dynamics*, 40(12):3324–3334, 2017. ISSN 15333884. doi: 10.2514/1.G002796.

J. Zhao and S. Li. Mars atmospheric entry trajectory optimization with maximum parachute deployment altitude using adaptive mesh refinement. *Acta Astronautica*, 160(June 2018): 401–413, 2019. ISSN 00945765. doi: 10.1016/j.actaastro.2019.03.027.

J. Zhao and S. Li. Adaptive mesh refinement method for solving optimal control problems using interpolation error analysis and improved data compression. *Journal of the Franklin Institute*, 357(3):1603–1627, 2020. ISSN 00160032. doi: 10.1016/j.jfranklin.2019.11.024.

J. Zhao and T. Shang. Dynamic optimization using local collocation methods and improved multiresolution technique. *Applied Sciences (Switzerland)*, 8(9), 2018. ISSN 20763417. doi: 10.3390/app8091680.

Y. Zhao and P. Tsiotras. Density functions for mesh refinement in numerical optimal control. *Journal of Guidance, Control, and Dynamics*, 34(1):271–277, 2011. ISSN 15333884. doi: 10.2514/1.45852.

Q. Zhou, Y. Wang, S. K. Choi, P. Jiang, X. Shao, and J. Hu. A sequential multi-fidelity metamodeling approach for data regression. *Knowledge-Based Systems*, 134:199–212, 2017. ISSN 09507051. doi: 10.1016/j.knosys.2017.07.033. URL http://dx.doi.org/10.1016/j.knosys.2017.07.033.

Q. J. Zhu. Hamiltonian necessary conditions for a multiobjective optimal control problem with endpoint constraints. *SIAM Journal on Control and Optimization*, 39(1):97–112, 2000. ISSN 03630129. doi: 10.1137/S0363012999350821.

W. Zou, Q. Zhang, Q. Gao, and Z. Feng. Wavelet multi-resolution approximation for multiobjective optimal control. *PLoS ONE*, 13(8):1–13, 2018. ISSN 19326203. doi: 10.1371/journal.pone.0201514.

Multi-Objective Optimal Control with Multi-Resolution Grids 6

6.1 Direct Finite Elements in Time . . . 257
6.2 Dyadic Mesh Refinement . . . 261
6.2.1 Vectorised Bézier Curve Refinement . . . 263
6.3 Variable-Length Evolutionary Algorithms . . . 265
6.4 Multi-Resolution Memetics . . . 268
6.4.1 One-Point Crossover . . . 269
6.4.2 Two-Point Crossover . . . 270
6.4.3 Dyadic Crossover . . . 270
6.4.4 Variable-Length Differential Evolution . . . 271
6.5 The VL-MOCP Algorithm . . . 272
6.6 Computational Experiments: Variable-Length Crossover . . . 273
6.6.1 Chemical Reaction . . . 274
6.6.2 Moon Lander . . . 280
6.6.3 Discussion . . . 286
6.7 Computational Experiments: Grid Resolution . . . 287
6.7.1 Drug Displacement . . . 288
6.7.2 Apollo Re-entry . . . 292
6.7.3 Betts Shuttle Re-entry . . . 298
6.7.4 Discussion . . . 302
6.8 Chapter Summary . . . 303
6.9 Bibliography . . . 309

This chapter continues the philosophy of the research introduced in Chapter 5, presenting a method of reducing the computational burden associated with *Multi-Objective Optimal*

Control Problems (MOCPs) in engineering design. The method focuses on adaptively increasing the resolution at which a set of discretised control solutions, each representing a particular balance between the problem objectives, may be evaluated. This is achieved by concentrating computational resource (in the form of discretisation nodes/grid-points) around areas of high complexity independently for each solution.

A novel aspect of the proposed approach is the simultaneous handling of various grid resolutions within an evolutionary framework, allowing solutions to independently evolve concurrent with their propagation through objective space. This is achieved through a unique social operator, designed to facilitate collaborative action between solutions of variable vector lengths. A complete MOCP may thus be treated as an open-ended evolutionary design problem. This ultimately facilitates a more complete, higher fidelity representation of the trade-off surface in a more numerically efficient manner. The advantages of this approach are summarised as follows:

- Non-smooth/discontinuous regions within any individual MOCP solution vector can be automatically detected, thus the *a priori* specification of both the number and distribution of discretisation elements is not required. Practically, this improves the usability of the proposed algorithm in generalised application.
- A population of individual solution vectors (or *agents*) may thus be initialised using only one finite element and allowed to independently refine across successive generations of an evolutionary framework. Importantly, this means that the size of the NLP problem associated with each agent is relatively small at earlier iterations, thus both computationally cheaper to evaluate and more receptive to global exploration.
- The criterion for deciding the region(s) in which to introduce new grid points is based on the inexpensive relative local error evaluated at element boundaries.
- Due to the multi-level nature of dyadic grid structures, existing nodes remain fixed across successive iterations, which benefits the overall convergence characteristics.
- User-specification of multi-objective scalarisation weights or preference functions is not required, as a population of multi-resolution solutions can be concurrently advanced towards the Pareto front via cooperative social actions designed to promote population diversity.
- Together, these points allow solution-specific non-linear/discontinuous behaviour to be efficiently captured during the evolutionary process. This ultimately allows a more accurate representation of the multi-objective trade-off for reduced computational effort relative to traditional uniform-grid approaches.

The work presented in the following chapter can be found in the following peer-reviewed publications:

- B. Parsonage and C. Maddock. Biased Dyadic Crossover for Variable-Length Multi-Objective Optimal Control Problems. *2024 IEEE Congress on Evolutionary Computation, CEC 2024 - Proceedings*, pages 1–9, 2024a. doi: 10.1109/CEC60901.2024.10611888
- B. Parsonage and C. Maddock. Solving multi-objective optimal control problems using a multiresolution approach. *Journal of Guidance, Control, and Dynamics*, pages 1–14, 2024b. ISSN 1533-3884. doi: 10.2514/1.g008223

6.1 Direct Finite Elements in Time

In this work, nonlinear multi-phase Optimal Control problems are transcribed into finite-dimensional NLP problems using the *Direct Finite Elements in Time* (DFET) transcription method (Vasile and Finzi, 2000).

Finite Elements in Time (FET) for the indirect solution of Optimal Control problems were initially proposed by Hodges and Bless (1991), later developing a discontinuous version. Several possible formulations of FET are possible (Borri et al., 1992), the most promising being those that derive from the semi-discontinuous and bi-discontinuous formulation of the mixed field equations. For example, this includes an unconditionally stable symplectic integration scheme for the bi-discontinuous version, and an unconditionally stable scheme with the asymptotic annihilation property. The latter proves useful within schemes for the integration of stiff systems. Additionally, Bottasso (1997) showed that FET for the forward integration of ordinary differential equations are equivalent to some classes of implicit Runge-Kutta integration schemes. FET can be extended to arbitrary high-order. Vasile (2010a) showed how a direct transcription method based on FET is very robust and allows for a *hp*-adaptive refinement technique (discussed further in Section 6.2). Direct transcription with FET on spectral bases has been successfully used to solve a range of problems, including the design of low-thrust multi-gravity assist trajectories to Mercury (Vasile and Bernelli-Zazzera, 2003), the design of weak stability boundary transfers to the Moon, low-thrust transfers within a restricted three-body problem and optimal landing trajectories to the Moon (Vasile and Finzi, 2000). More recently, DFET have been used to solve *Multi-Objective Optimal Control Problems* (MOCPs) representing spacecraft (Ricciardi et al., 2018a), the ascent trajectories of launchers (Ricciardi et al., 2016), abort trajectories of reusable launch vehicles (Ricciardi et al., 2018b) and the removal of space debris (Ricciardi and Vasile, 2019c).

This section recalls the general approach to the DFET transcription method presented by Vasile (2010b). From Section 6.1, a generic optimal control problem can be expressed in Bolza

form as:

$$\min_{\mathbf{u}\in U} J = \min_{\mathbf{u}\in U} \phi(\mathbf{x}_i, \mathbf{x}_f, t_0, t_f) + \int_{t_0}^{t_f} L(\mathbf{x}(t), \mathbf{u}(t), t)dt \tag{6.1}$$

subject to n_D differential constraints:

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}(t), \mathbf{u}(t), t) \tag{6.2}$$

and n_A algebraic constraints:

$$\mathbf{g}(\mathbf{x}(t), \mathbf{u}(t), t) \leq 0 \tag{6.3}$$

with the transversality conditions:

$$\psi(\mathbf{x}_0, \mathbf{x}_f, t_0, t_f) \leq 0 \tag{6.4}$$

for the time interval given by:

$$t \in [t_0, t_f] \tag{6.5}$$

where the scalar t is time, the functions $\mathbf{x}(t) : [t_0, t_f] \to \mathbb{R}^n$ represent the n-element state vector, the functions $\mathbf{u}(t) : [t_0, t_f] \to \mathbb{R}^p$ represent the p-element control vector and J is the sum of the boundary state cost function $\psi : \mathbb{R}^n \times \mathbb{R}^p \times [t_0, t_f] \to \mathbb{R}$ and of the integral of a running cost function $L : \mathbb{R}^n \times \mathbb{R}^p \times [t_0, t_f] \to \mathbb{R}$. The functions $\mathbf{x}(t)$ belong to the Sobolev space $\mathcal{W}^{l,\infty}$ while the functions $\mathbf{u}(t)$ belong to L^∞. The vector fields that appear in the dynamic, path and boundary constraints, respectively, are $\mathbf{F} : \mathbb{R}^n \times \mathbb{R}^p \times [t_0, t_f] \to \mathbb{R}^n$, $\mathbf{g} : \mathbb{R}^n \times \mathbb{R}^p \times [t_0, t_f] \to \mathbb{R}^s$ and $\boldsymbol{\psi} : \mathbb{R}^{2n+2} \to \mathbb{R}^q$. Recasting the differential constraints in weak variational form gives:

$$\int_{t_0}^{t_f} \mathbf{w}(t)^T(\dot{\mathbf{x}}(t) - F(\mathbf{x}(t), \mathbf{u}(t), t))dt = \mathbf{w}(t)^T(\mathbf{x} - \mathbf{x}^b)|_{t_0}^{t_f} \tag{6.6}$$

where $\mathbf{w}$ are generalised weight functions assuming a value of 0 on either bound and $\mathbf{x}^b$ are the imposed/free state boundary values. Integrating by parts gives:

$$\int_{t_0}^{t_f} (\dot{\mathbf{w}}(t)^T\mathbf{x}(t) + \mathbf{w}(t)^T F(\mathbf{x}(t), \mathbf{u}(t), t))dt - \mathbf{w}_f^T\mathbf{x}_f^b + \mathbf{w}_i^T\mathbf{x}_i^b = 0 \tag{6.7}$$

Of note, in this bi-discontinuous formulation, the value of $\mathbf{x}(t)$ at the boundaries does not coincide with either with $\mathbf{x}_f^b$ or $\mathbf{x}_0^b$. As remarked by Bottasso (1997), the weak form of $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}(t), \mathbf{u}(t), t)$ is a mathematically correct formulation even when $\mathbf{F}(\mathbf{x}(t), \mathbf{u}(t), t)$ is not continuous or differentiable. The time domain $D(t_0, t_f) \subset \mathbb{R}$ is decomposed into N finite elements:

$$D = \bigcup_{\epsilon=1}^{N} D_\epsilon(t_{\epsilon-1}, t_\epsilon) \tag{6.8}$$

and thus the state, control and weight functions may be parametrised on each element D_ϵ as follows:

$$\mathbf{x}_\epsilon(t) = \sum_{s=0}^{l} f_{s\epsilon}(t)\mathbf{x}_{s\epsilon} \tag{6.9}$$

$$\mathbf{u}_\epsilon(t) = \sum_{s=0}^{m} g_{s\epsilon}(t)\mathbf{u}_{s\epsilon} \tag{6.10}$$

$$\mathbf{w}_\epsilon(t) = \sum_{s=0}^{l+1} h_{s\epsilon}(t)\mathbf{w}_{s\epsilon} \tag{6.11}$$

where the basis functions $f_{s\epsilon}(t)$, $g_{s\epsilon}(t)$ and $h_{s\epsilon}(t)$ are chosen within the space of polynomials of order l, m and $l+1$ respectively:

$$f_{s\epsilon} \in P^l(D_\epsilon) \tag{6.12}$$

$$g_{s\epsilon} \in P^m(D_\epsilon) \tag{6.13}$$

$$h_{s\epsilon} \in P^{l+1}(D_\epsilon) \tag{6.14}$$

and the vectors $\mathbf{x}_{s\epsilon}$, $\mathbf{u}_{s\epsilon}$ and $\mathbf{w}_{s\epsilon}$ are weights given to each polynomial in each element. Each element D_ϵ can be seen as a smooth image of the reference time interval $[-1, 1]$, where the reference time τ is defined as:

$$\tau = 2\frac{t - \frac{t_\epsilon - t_{\epsilon-1}}{2}}{t_\epsilon - t_{\epsilon-1}} \tag{6.15}$$

In this way, the domain of the basis function is constant and irrespective of the size of the element. Furthermore, the domain thus overlaps with the interval of the Gauss nodes that will be employed for the integration of the dynamics. Substituting the definitions of the polynomials into the objective functions and integrating with Gauss quadrature formulas with q nodes leads to:

$$\tilde{J} = \phi(\mathbf{x}_0^b, \mathbf{x}_f^b, t_0, t_f) + \sum_{\epsilon=1}^{N}\sum_{k=1}^{q} \sigma_k L(\mathbf{x}_\epsilon(\tau_k), \mathbf{u}_\epsilon(\tau_k), \tau_k)\frac{\Delta t}{2} \tag{6.16}$$

and the variational constraints for each element j become:

$$\sum_{k=1}^{q} \sigma_k \left[\dot{\mathbf{w}}_\epsilon(\tau_k)^T\mathbf{x}_\epsilon(\tau_k) + \mathbf{w}_\epsilon(\tau_k)^T\mathbf{F}_\epsilon(\tau_k)\frac{\Delta t}{2}\right] - \mathbf{w}_\epsilon(t_f)^T\mathbf{x}_\epsilon^b + \mathbf{w}_\epsilon(t_0)^T\mathbf{x}_{\epsilon-1}^b = 0 \tag{6.17}$$

where τ_k and σ_k are the Gauss nodes and weights, $\mathbf{F}_\epsilon(\tau_k)$ is the shorthand notation for $\mathbf{F}(\mathbf{x}_\epsilon(\tau_k), \mathbf{u}_\epsilon(\tau_k), \tau_k)$, and $\mathbf{x}_\epsilon^b$ and $\mathbf{x}_{\epsilon-1}^b$ denote the boundary values of element ϵ at $t = t_\epsilon$ and $t = t_{\epsilon-1}$ respectively.

In the DFET literature, the basis is typically generated through a Lagrange interpolation on the quadrature nodes, usually either Gauss-Legendre or Gauss-Lobatto:

$$f_{sj}(\tau) = \prod_{k=0,k\neq s}^{l} \frac{\tau - \tau_k}{\tau_s - \tau_k} \tag{6.18}$$

$$g_{sj}(\tau) = \prod_{k=0,k\neq s}^{m} \frac{\tau - \tau_k}{\tau_s - \tau_k} \tag{6.19}$$

$$h_{sj}(\tau) = \prod_{k=0,k\neq s}^{l+1} \frac{\tau - \tau_k}{\tau_s - \tau_k} \tag{6.20}$$

The test functions w must assume a value of 0 on either bound, thus the nodes for the construction of test functions with Lagrange interpolation must be of Lobatto type. Readers are directed towards (Ricciardi and Vasile, 2019a) for further discussion on the choice of basis functions for this class of problem. In this chapter, Bernstein polynomials are employed as basis functions in accordance with the work by Ricciardi and Vasile (2019a). A Bernstein basis of order n is defined as:

$$B_{v,n}(t) = \binom{n}{v} t^v (1-t)^{n-v} \quad 0 \leq v \leq n,\ 0 \leq t \leq 1 \tag{6.21}$$

As shown in Fig. 6.1, Bernstein polynomials either assume a value of 0 on both boundaries or a value of 1 on one boundary and of 0 on the other, thus may be used as test functions. Since the resulting curves will be integrated through Gauss quadrature, which requires nodes defined on $[-1, 1]$, Bernstein polynomials must also be defined on the interval $[-1, 1]$:

$$\tilde{B}_{v,n}(\tau) = B_{v,n}(t), \quad \tau = 2t - 1 \tag{6.22}$$

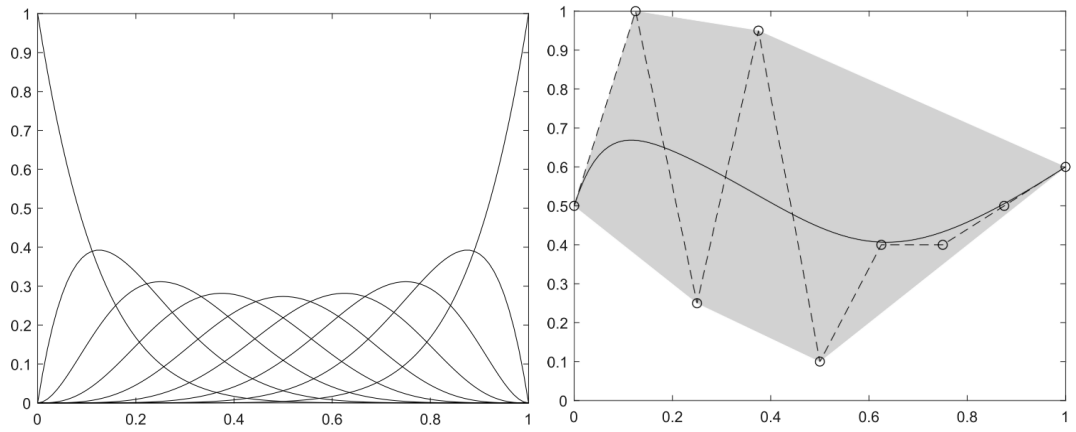
In the following, for clarity of notation, the subscript n to denote the order of the Bernstein polynomials will be dropped. That degree will equal l, m or $l + 1$ depending on whether the polynomials will be used to describe the states, controls or test functions respectively. Substituting:

$$\mathbf{x}_\epsilon(t) = \sum_{s=0}^{l} \tilde{B}_{s\epsilon}(\tau)\mathbf{x}_{s\epsilon} = \mathcal{B}_\epsilon(\tau) \tag{6.23}$$

$$\mathbf{u}_\epsilon(t) = \sum_{s=0}^{m} \tilde{B}_{s\epsilon}(\tau)\mathbf{u}_{s\epsilon} = \mathcal{C}_\epsilon(\tau) \tag{6.24}$$

$$\mathbf{w}_\epsilon(t) = \sum_{s=0}^{l+1} \tilde{B}_{s\epsilon}(\tau)\mathbf{w}_{s\epsilon} = \mathcal{D}_\epsilon(\tau) \tag{6.25}$$

Thus, if the basis is made of Bernstein polynomials, the resulting state curves $\mathcal{B}_\epsilon(\tau)$, control curves $\mathcal{C}_\epsilon(\tau)$ and test curves $\mathcal{D}_\epsilon(\tau)$ are, by definition, Bézier curves. The reader is referred to Ricciardi and Vasile (2019a) for a more in-depth discussion of the properties of Bézier curves that support their use in this application.



(a) (b)

FIGURE 6.1: 8th order (a) Bernstein basis and (b) Bézier curve (Ricciardi and Vasile, 2019a).

6.2 Dyadic Mesh Refinement

The collection of discretisation nodes τ that represent the interface between each finite element D of a solution vector $\mathbf{x}$ over a normalised time interval may be considered a uniform dyadic mesh $\mathcal{V}_{j,N}$ conforming to the general form:

$$\mathcal{V}_{j,N} = \left\{ \tau_{j,k} \in [0,1] : \tau_{j,k} = \frac{k}{2^j N}, \;\; 0 \leq k \leq 2^j N, \;\; 0 \leq j \leq J_{max} \right\} \tag{6.26}$$

where $\tau_{j,k}$ are the time coordinates of the interface nodes with corresponding spatial index k. The positive integers j and J_{max} represent the current and maximum allowable resolution level and N is the number of elements present in the initial uniform mesh discretisation. By definition, a dyadic grid is obtained by successively subdividing a uniform initial grid such that $\mathcal{W}_{j,N}$ denotes the set of grid points belonging to $\mathcal{V}_{j+1,N}$ but not $\mathcal{V}_{j,N}$; that is,

$$\mathcal{W}_{j,N} = \left\{ \hat{\tau}_{j,k} \in [0,1] : \hat{\tau}_{j,k} = \frac{2k+1}{2^{j+1} N}, \; 0 \leq k \leq 2^j N - 1, \;\; 0 \leq j \leq (J_{max} - 1) \right\} \tag{6.27}$$

Hence, $\tau_{j+1,k} \in \mathcal{V}_{j+1,N}$ if:

$$\tau_{j+1,k} = \begin{cases} \tau_{j,k/2}, & k \text{ is even} \\ \hat{\tau}_{j,(k-1)/2} & \text{otherwise} \end{cases} \tag{6.28}$$

N can be any positive integer for the generalised dyadic grid, which is more convenient for optimisation[1]. An example of a uniform dyadic grid with $N = 1$ and $J_{max} = 7$ is shown in Fig. 6.2, where only the nodes corresponding to $\mathcal{V}_{0,N}$ and $\mathcal{W}_{j,N}$, $0 \leq j \leq (J_{max} - 1)$ have been included for clarity.

[1]Throughout this work however, $N = 1$.

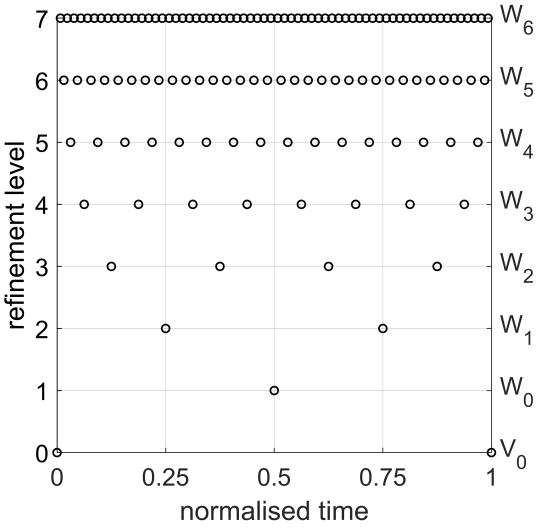


FIGURE 6.2: A uniform dyadic grid conforming to Eqs (6.26) and (6.27), where $N = 1$ and $J_{max} = 7$.

The subspaces $\mathcal{V}_{j,N}$ are nested;

$$\mathcal{V}_{0,N} \subset \mathcal{V}_{1,N} \subset \cdots \subset \mathcal{V}_{J_{max},N} \tag{6.29}$$

with

$$\lim_{J_{max}\to\infty} \mathcal{V}_{J_{max},N} = [0,1] \tag{6.30}$$

The sequence of subspaces $\mathcal{W}_{j,N}$ satisfies the property $\mathcal{W}_{j,N} \cap \mathcal{W}_{l,N} = \varnothing$ for all $j \neq l$. The sets $\mathcal{V}_{j,N}$ and $\mathcal{W}_{l,N}$ for $j \leq l$ are orthogonal to each other, as are the sets $\mathcal{W}_{j,N}$ and $\mathcal{W}_{l,N}$ for $j \neq l$ (see Fig. 6.2). If element subdivision is performed in a non-uniform manner, (6.26) can be modified to represent the resulting non-uniform grid:

$$\begin{aligned} \mathcal{G} &= \left\{ \tau_{j_i,k_i} \in [0,1] : 0 \leq k_i \leq 2^{j_i} \right\} \\ & J_{min} \leq j_i \leq J_{max} \text{ for } i = 1,\ldots,R \\ & \tau_{j_i,k_i} < \tau_{j_{i+1},k_{i+1}} \text{ for } i = 1,\ldots,R-1 \end{aligned} \tag{6.31}$$

where $\mathcal{G} \subset \mathcal{V}_{J_{max},N}$ and R represents the number of available resolution levels. Of note, working with dyadic grids is essentially equivalent to using interpolating wavelets for the analysis of the underlying function (Jain and Tsiotras, 2008), and retains one of the major benefits of wavelet-based analyses: multi-resolution functional representations (Harten, 1996; Mallat, 1989).

Grid adaptation is accomplished by first assessing the mesh quality. Temporal finite elements exhibit spectral convergence of the boundary values as the number of nodes increases and a slower convergence of the discontinuity at the boundaries (Vasile, 2010b). Hence, controlling

the discontinuities at element boundaries guarantees a control of the global error. By defining the discontinuity at each interface node τ as:

$$\boldsymbol{\delta}_\tau = |\mathbf{x}^l - \mathbf{x}^r| \tag{6.32}$$

where superscripts l and r denote the values to the left and right of interface node τ respectively[2], one can thus obtain the *relative local error* as an indicator of grid quality:

$$\boldsymbol{\epsilon}_\tau = \frac{\boldsymbol{\delta}_\tau}{\max[\mathbf{x}_{UB} - \mathbf{x}_{LB}, |\mathbf{x}_{UB}|, |\mathbf{x}_{LB}|]} \tag{6.33}$$

where $\mathbf{x}_{UB}$ and $\mathbf{x}_{LB}$ are respectively the user-defined upper and lower bounds of the state vector $\mathbf{x}$.

Using this approach, one does not require any additional procedure or integration as sufficient information may be derived from the discontinuities at the boundaries of each finite element. A given grid may therefore be modified by inserting additional interface nodes at the midpoints of all finite elements not satisfying $\epsilon_\tau \leq \epsilon_{tol}$, where ϵ_{tol} is a user-specified tolerance. This then preserves the non-uniform dyadic structure according to (6.31).

This refinement procedure requires the associated basis functions spanning each temporal element be split in the same manner. A useful property of Bernstein polynomials in this regard is that any Bézier curve (sequential collection of Bernstein polynomials) can be exactly represented by two, smaller subcurves of the same order, as depicted graphically in Fig. 6.3. Alternative basis functions, such as Chebyshev polynomials, would have to be refit after each grid adaptation step.

6.2.1 Vectorised Bézier Curve Refinement

In this study, a vectorised, non-destructive, recursive subdivision scheme (Pomax, 2023) is employed to refine Bernstein Polynomials. Recall from Section 6.1, that a basis comprised of n-th order Bernstein polynomials (6.21) defines a Bézier curve according to the general form:

$$\mathcal{B}(t) = \sum_{v=0}^{n} B_{v,n}(t)\mathbf{x}_v \tag{6.34}$$

which can be represented in matrix form as:

$$\mathcal{B}(t) = [\mathbf{T}]\ [\mathbf{M}]\ [\mathbf{X}] \tag{6.35}$$

where $\mathbf{T}$ represents the polynomial basis $\mathbf{T} = [t^0, \ldots, t^n]$, $\mathbf{X}$ is a column vector of control points $\mathbf{X} = [x_1, \ldots, x_m]^T$, where $m = n + 1$, and $\mathbf{M} : \mathbb{R}^{m \times m}$ is a triangular coefficients

[2] In other words, $\mathbf{x}^l$ represents the final values of the state basis function corresponding to the element immediately preceding τ, and $\mathbf{x}^r$ represents the initial value of the state basis function corresponding to the element following τ.

matrix that may be generalised (in accordance with binomial theorem) as:

$$\mathbf{M} = \mathcal{P}^T \text{diag}([\mathcal{P}_{m,1}, \ldots, \mathcal{P}_{m,m}]) \tag{6.36}$$

where $\mathcal{P} : \mathbb{R}^{m \times m}$ is the lower triangular Cholesky factor (up to the signs of the columns) of the Pascal matrix.

If $\mathcal{B}$ is to be split exactly in half, we define the bisection node $z = 0.5$. To find the coordinates $\tilde{\mathbf{X}}_1$ of the first sub-curve $\tilde{\mathcal{B}}_1$ bounded between $[0,\ z]$, the matrix $\mathbf{T}$ becomes:

$$\mathbf{T} = [(0 + z\ t)^0, \ldots, (0 + z\ t)^n] \tag{6.37}$$

which may be decomposed to provide the following matrix representation of subcurve $\tilde{\mathcal{B}}_1$:

$$\tilde{\mathcal{B}}_1(t) = \mathbf{T}\ \mathbf{Z}\ \mathbf{M}\ \mathbf{X} \tag{6.38}$$

where

$$\mathbf{Z} = \text{diag}(\mathbf{z}) \tag{6.39}$$

with $\mathbf{z} = [z^0, \ldots, z^n]$. By inserting the identity matrix $\mathbf{M}\ \mathbf{M}^{-1}$, (6.43) can be represented as:

$$\tilde{\mathcal{B}}_1(t) = \mathbf{T}\ \mathbf{M}\ \ (\mathbf{Q}\ \mathbf{X}) \tag{6.40}$$

where

$$\mathbf{Q} = \mathbf{M}^{-1}\ \mathbf{Z}\ \mathbf{M} \tag{6.41}$$

Thus by comparing (6.35) and (6.40), it can be seen that the control points representing the first split segment $\tilde{\mathcal{B}}_1$ bounded between $[0,\ z]$ are given by $\tilde{\mathbf{X}}_1 = [\mathbf{Q}\ \mathbf{X}]$. To evaluate the control points defining the second segment $\tilde{\mathcal{B}}_2$, this process is repeated for the interval $[z,\ 1]$. The matrix $\mathbf{T}$ is thus written as:

$$\mathbf{T} = [(z + (1 - z)\ t)^0, \ldots, (z + (1 - z)\ t)^n] \tag{6.42}$$

which may be decomposed to give:

$$\tilde{\mathcal{B}}_2(t) = \mathbf{T}\ \mathbf{Z}^*\ \mathbf{M}\ \mathbf{X} \tag{6.43}$$

By defining $w = 1 - z$, the general form of $\mathbf{Z}^*$ may be written as:

$$\mathbf{Z}^* = |\mathcal{P}^T|\ \mathbf{w}^T\ \mathcal{T}_U(\mathbf{z}) \tag{6.44}$$

where $\mathbf{w} = [w^0, \ldots, w^n]$ and $\mathcal{T}_U(\mathbf{z})$ is the upper triangular of a symmetrical Toeplitz matrix with the first row given by $\mathbf{z} = [z^0, \ldots, z^n]$. The second subcurve $\tilde{\mathcal{B}}_2$ can therefore be written as:

$$\tilde{\mathcal{B}}(t) = \mathbf{T}\ \mathbf{M}\ \mathbf{Q}^*\ \mathbf{X} \tag{6.45}$$

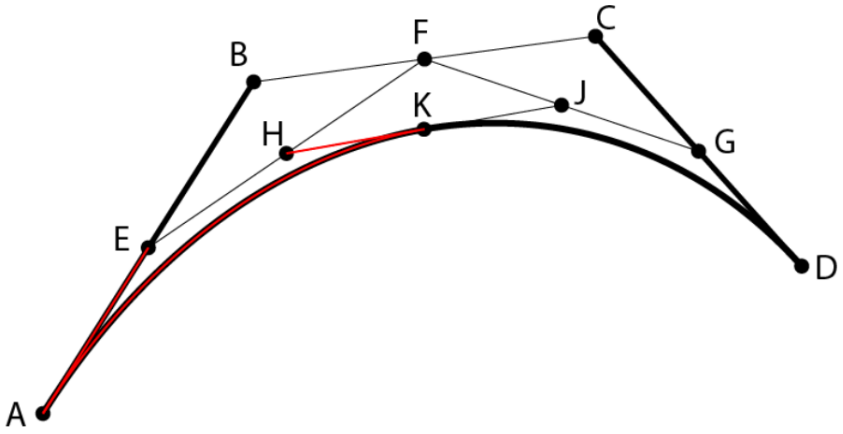


FIGURE 6.3: The Bézier curve defined by nodes ABCD may be perfectly bisected into the curves AEHK and KJGD.

where

$$\mathbf{Q}^* = \mathbf{M}^{-1}\,\mathbf{Z}^*\,\mathbf{M} \tag{6.46}$$

Thus the control points representing the second subcurve $\tilde{\mathcal{B}}_2$ bounded between $[z, 1]$ are given by the expression: $\tilde{\mathbf{X}}_2 = [\mathbf{Q}^*\ \mathbf{X}]$.

As an example, Figure 6.4 depicts two extremal solutions to a classic MOCP sourced from the literature (Betts and Huffman, 1998; Citron, 1969; Jain and Tsiotras, 2008), solved using the multi-resolution refinement approach of this chapter. In each case it can be seen that grid nodes have been successfully concentrated to an appropriate degree within different areas of the time domain.

6.3 Variable-Length Evolutionary Algorithms

Homology is defined in the field of evolutionary biology as "possessing a common evolutionary origin" (Reeck et al., 1987). This is as opposed to merely similarity of sequence, function or location. Considering linear genomes of constant vector length, genetic crossover is performed by exchanging homologous sequences between two parental genomes. The most prevalent methods are typically differentiated by the number n of crossover point-pairs selected to define such sequences (e.g. $n = 1, 2, \ldots, k$) (Merlevede et al., 2019). Crossover points are often selected assuming unambiguous homology across successive generations. That is, each element or sequence remains fixed in position within a particular genome. Clearly, this assumption cannot hold for variable-length structures. As in nature, variable-length crossover must be guided to exchange sensible or equivalent sequences regardless of the respective lengths of the parent genomes. Practically, this may be achieved (for example) by embedding

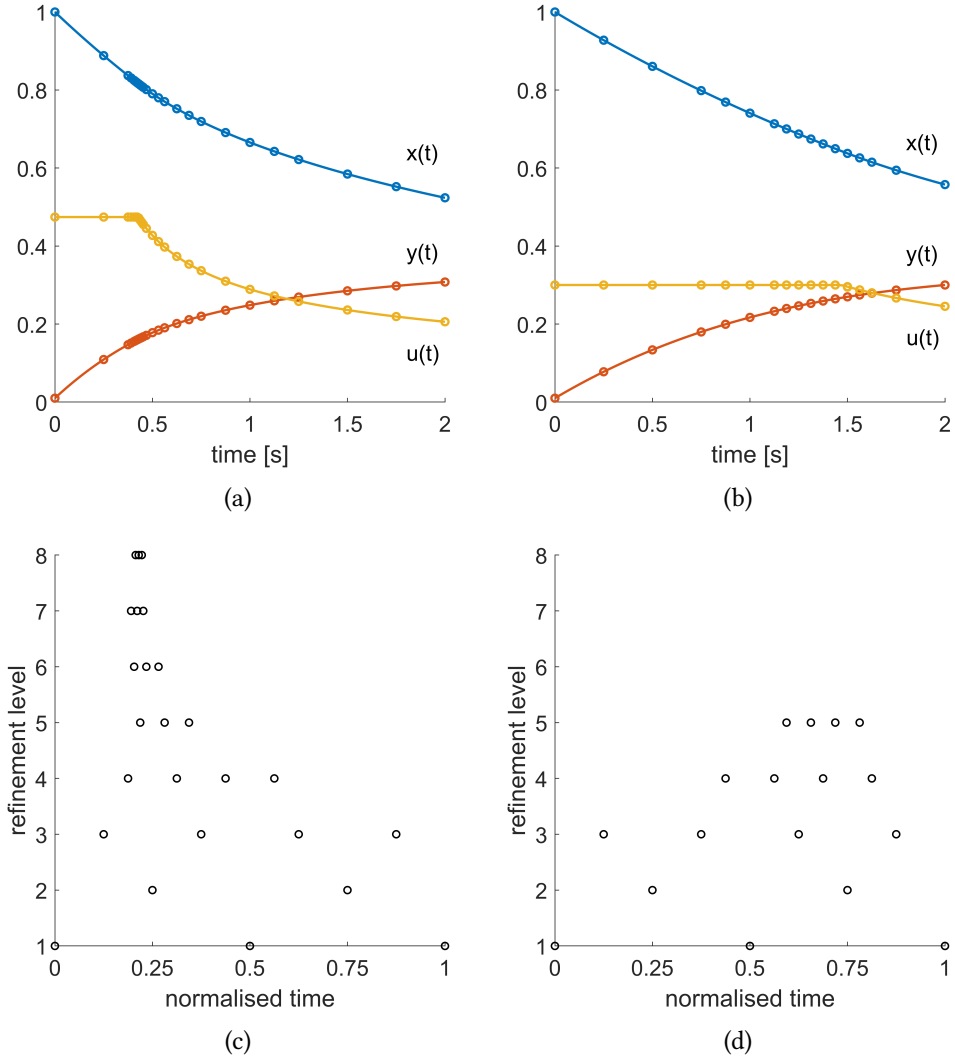


FIGURE 6.4: Optimal solutions to the problem presented in Betts and Huffman (1998); Citron (1969); Jain and Tsiotras (2008), where (a) and (b) represent the state and control profiles for $\min_{t_f,\mathbf{u}}[J_1] = -y$ and $\min_{t_f,\mathbf{u}}[J_2] = u_{max}$ respectively. The circle markers in (c) and (d) represent the grid nodes (element boundaries) for corresponding solutions (a) and (b).

metadata within the genome (Zebulum et al., 2000a), or by inferring homology via sequence similarity or some other relevant heuristic (Pearson, 2013).

An example of the latter is the *Species Adaptation Genetic Algorithm* (SAGA) (Harvey, 1992). SAGA utilises a similarity metric, the *Longest Common Sub-Sequence* (LCSS), designed to favour the exchange of similar segments during recombination. Similarly to Goldberg's *Messy-GA* (mGA) (Goldberg, 1989), an initial crossover point is randomly selected from the first parent genome. However, where mGA randomly samples a corresponding point in the second parent genome, SAGA sequentially tests each possible matching point in the second parent. The pairing with the maximum sum of LCSS (on both the left and right split segments of each genome) is selected. Where there are multiple candidates with identical scoring, one is selected at random. A similar approach is adopted in the *Virtual Virus* (VV) algorithm (Burke et al., 1998), where crossover probability is regulated by the degree of local similarity between parent genomes, given by the number of matches within a specified fixed size window. However, the SAGA crossover approach explicitly attempts to preserve a complete genetic sequence using the parent genomes as a guide, and thus remains superior in this respect.

Another example is the use of *Hidden Gene* GAs (Abdelkhalik, 2013; Zebulum et al., 2000b). Hidden-Gene GAs represent solutions using two structures, one including all design parameters of a particular problem and another including a binary 'tag' vector that indicates which parameters are 'active' in a particular solution. The so-called *global* vector therefore contains the maximum number of variables that can be used to characterise the current problem. If the binary tag associated to a particular element of the global vector is not active, that variable will not be included in the evaluation of that solution.

These methods still fundamentally assume the genome itself is an inflexible, rigid array of data, allowing unbiased selection of the initial crossover point. As such, this selected point may lie within any unique structure or subsequence that has a particularly strong effect on the fitness of the genome as a whole, thus is less likely to have a complimentary (or even appropriate) pair in the second genome. This can result in a largely disruptive crossover, potentially compromising the exchange and/or preservation of data. In the work (Hutt and Warwick, 2007), a softer approach is demonstrated, more in line with biological crossover. The *Synapsing Variable-Length Crossover* (SVLC) algorithm selects the initial crossover point biased by the physical structure of *both* parent genomes. Parents are *synapsed* together at points of similarity (again determined using the LCSS approach, with longer sequences given priority). Crossover is then performed only within synapsed regions, producing offspring that inherit the entirety of the common parental subsequences. This potential for smoother, non-destructive crossover is particularly appealing given the findings of (Harvey, 1992), who used the concepts of *epistasis* and *fitness landscapes* to show that propagation through a search space of indefinite dimensionality may only be feasible through a *gradual* increase in genotype

length.

It may thus be suggested that the inference of homology via the genetic structure of parental genomes is more likely to facilitate meaningful, non-destructive crossover. In this manner, homologous segments may be both adaptively selected and compared throughout an open-ended optimisation process. It is via this principle that the following methods are proposed and applied within the context of evolutionary MOCPs, where the structure of unique solutions is inherently linked to the NLP transcription approach.

6.4 Multi-Resolution Memetics

For the purposes of this work, the proposed grid refinement approach is used to augment the *Multi-Agent Collaborative Search* (MACS) algorithm, a dominance-based memetic solver for *Multi-Objective Optimisation* (MOO) problems under development at the University of Strathclyde (Vasile and Zuiani, 2011).

MACS initialises a population of virtual agents at random locations in the search space, where an 'agent' comprises an optimal control solution vector and corresponding objective function and constraint function values. Every agent in turn explore it's local neighbourhood through a series of individualistic actions (*inertia*, *pattern search* and *differential evolution*). The population as a whole then performs a set of collaborative, social actions to concurrently advance towards a complete representation of the Pareto-front. At every user specified number of iterations (and again at the final iteration), a gradient-based local optimisation utilising Pascoletti-Serafini scalarisation Pascoletti and Serafini (1984) is performed on each agent.

The MACS algorithm has been extensively described in the works by Vasile and Zuiani (2011); Ricciardi (2019); Ricciardi and Vasile (2019b). However, the current formulation of MACS assumes all solution vectors conform to a predetermined structure of fixed length, thus is incompatible with adaptive problems. In order for the proposed grid refinement procedure to be integrated into MACS, a new social operation, one that allows the handling of variable-length discretisation grids and therefore enables open-ended evolution, is required.

The individualistic actions within MACS operate in isolation for each member of the current population, and thus need only minimal modification to incorporate continuously adapting solutions. However, considering the collaborative actions (operations that combine information from multiple members of a population, e.g., memetic crossover), accounting for open-ended evolution implies the need to identify common points or sequences between two individuals without assuming one-to-one correspondence between vector components. This type of identification is utilised within *Variable-Length-Chromosome Genetic Algorithmss* (VLC-GAs) (Smith, 1980; Koza, 1992), a subclass within the wider field of *Genetic Algorithms* (GAs) (Holland,

1992). Munetomo et al. (1998) present a crossover mechanism for VLC-GAs called same-point crossover, involving two operators: one-point (1P) and two-point (2P) crossover. Partial chromosomes[3] are thus exchanged at the loci of identical genes between both parent chromosomes. This mechanism has since been inherited by most (if not all) subsequent VLC-GAs. Among other well known methods for variable-length crossover include Goldberg's Messy-GA (Goldberg, 1989), Species Adaptation Genetic Algorithms (Harvey, 1992), Synapsing Variable-Length Crossover (Hutt and Warwick, 2007), the Virtual Virus method (Burke et al., 1998) and the more recently proposed Same-Adjacency crossover (Qiongbing and Lixin, 2016).

Representing a novel application of *Variable-Length Chromosome* (VLC) methodology, this work proposes to leverage the imposed dyadic grid structures described in Section 6.2 to allow a MOCP to be redefined as an open-ended evolutionary design problem. More specifically, given that the nodal intersection between two or more unique dyadic grids can be easily computed, common vector components between alternative solutions can be quickly identified. This then facilitates direct comparison and/or recombination for the purpose of social action.

It is important to note that given an initially uniform discretisation, the proposed dyadic refinement scheme (Section 6.2) implies that any segment bounded by common intersection nodes of any two solutions must by definition share common ancestry. With reference to the previous definition given in Section 6.3, *homology* in this context additionally infers that such segments represent the same portion of the solution in normalised time, regardless of the number of intermediate nodes enclosed within.

The precise manner in which this may be performed is now discussed.

6.4.1 One-Point Crossover

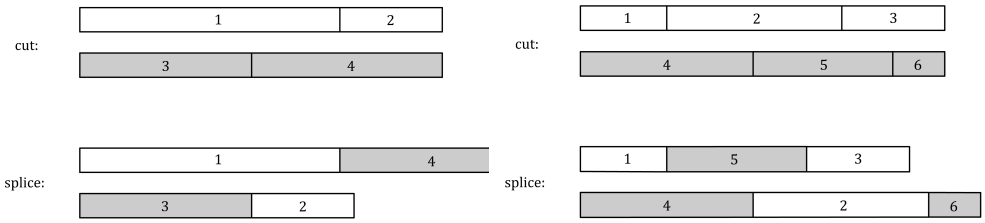
Consider two individual solution vectors of respective lengths L and K, $\mathbf{x} = [x_0, x_1, \ldots, x_i, \ldots, x_L]$ and $\mathbf{x}^* = [x_0^*, x_1^*, \ldots, x_j^*, \ldots, x_K^*]$ with corresponding multi-resolution grids $\mathcal{G}^{(\mathbf{x})}$ and $\mathcal{G}^{(\mathbf{x}^*)}$. Given that each agent within the population is initialized upon an identical uniform grid discretization, i.e. $N^{(\mathbf{x})} = N^{(\mathbf{x}^*)}$, and that individual grid elements can only be perfectly bisected, the intersection between $\mathbf{x}$ and $\mathbf{x}^*$ will include at least one intermediate node (excluding the trivial case of common initial and final nodes) such that:

$$\mathcal{G}^{(\mathbf{x})} \cap \mathcal{G}^{(\mathbf{x}^*)} \neq \varnothing, \quad 2 \leq \mathbb{R}^{(\mathbf{x},\mathbf{x}^*)} \leq 2^{J_{max}} \tag{6.47}$$

The simplest variable-length social action that can thus be considered is that of classical one-point (1P) crossover. In this case, a single intersection node is selected at random, $x_i = x_j^*$, with all components following x_i and x_j^* exchanged between the parent solutions to create two

[3] In this context, 'chromosome' is taken as equivalent to the problem solution vector.

new individuals (see Fig. 6.5a). This is functionally equivalent to the *cut* and *splice* operations introduced in Goldberg's Messy-GA (Goldberg, 1989).



(a) (b)

FIGURE 6.5: Two parent vectors are split at (a) one or (b) two appropriate point-pair(s), with all elements following (or between) the split(s) exchanged between parents to form two new vectors.

6.4.2 Two-Point Crossover

One-point crossover has the feature that solution vector components nearer the extremes of each individual are much more likely to be separated than components nearer the middle (Harvey, 1992). Intuitively then, two-point (2P) crossover may avoid this inherent bias by instead randomly selecting two intermediate nodes from the set of intersections between $\mathbf{x}$ and $\mathbf{x}^*$, exchanging the enclosed subsequence (see Fig. 6.5b). In this case, grid point intersections conform to:

$$\mathcal{G}^{(\mathbf{x})} \cap \mathcal{G}^{(\mathbf{x}^*)} \geq 2, \quad 2 \leq \mathbb{R}^{(\mathbf{x},\mathbf{x}^*)} \leq 2^{J_{max}} \tag{6.48}$$

6.4.3 Dyadic Crossover

A known flaw within both preceding methods is the potential for lack of offspring diversity should the crossover point(s) be chosen too close to the initial and/or final nodes (Qiongbing and Lixin, 2016). In such cases, this is equivalent to saying the length of the exchanged segment is comparable to the lengths of the parent vectors themselves. An interesting consideration within the context of this study is that as the population is progressively evolved, the number of intersections between similar solutions is likely to increase. This therefore allows a progressively finer level of crossover to occur between similar solutions, thus reducing the overall difference between agents comprising successive generations, which itself has implications for the convergence characteristics of the algorithm as a whole.

However, while the potential for smoother social cooperation is introduced, crossover points are still only selected at random, thus there is no guarantee that crossover at a higher level of resolution will actually occur. Indeed, Harvey (1992) provides a modification to classical two-point crossover in order to choose the second point such that the similarity between the

exchanged segments is maximized. *'Similarity'* in this sense is measured according to the longest common subsequence, defined as the "longest uninterrupted matching substring of gene values (alleles) found between two strings of arbitrary length" (Harvey, 1992).

In a similar vein, the proposed method in this work is designed to instead favour the exchange of *higher resolution* segments. That is, segments that intersect at higher resolution levels J. In this manner, as the number of intersection points between solutions increases, information exchange will occur at as high a resolution level as is possible. This aims to minimize the potential disruptive effects caused by blind exchange of larger subsequences (a not uncommon feature of unbiased 1P/2P crossover) and thus promotes a more gradual convergence in objective space. Furthermore, if the selection of crossover points is allowed to include either the initial or final nodes, the biased crossover operation (referred to now as *dyadic crossover*), can effectively switch freely between 1P and 2P classifications, guided by the resolution of the parent individuals.

6.4.4 Variable-Length Differential Evolution

A final consideration must be given to the use of *Differential Evolution* (DE) (Madavan, 2002), one of the default social operations within the original MACS algorithm. When a particular agent $\mathbf{x}_p$ is selected for DE, a displacement vector $\mathbf{dx}_p$ is defined through combination with three randomly selected individuals, $\mathbf{x}_{p_1}$, $\mathbf{x}_{p_2}$ and $\mathbf{x}_{p_3}$, from the population:

$$\mathbf{dx}_p = \alpha \mathbf{e} \left((\mathbf{x}_p - \mathbf{x}_{p_1}) + F \left(\mathbf{x}_{p_2} - \mathbf{x}_{p_3} \right) \right) \tag{6.49}$$

where α is a random number between 0 and 1, F is a user specified constant and $\mathbf{e}$ is a mask vector of elements of either 0 or 1 according to:

$$e_j = \begin{cases} 1, & \text{if } \alpha_2 < CR \\ 0, & \text{otherwise} \end{cases} \tag{6.50}$$

where α_2 is another random number between 0 and 1, and the crossover ratio CR is another user specified constant. The resulting trial position thus reads:

$$\mathbf{x}_{trial} = \mathbf{x}_p + \mathbf{dx}_p \tag{6.51}$$

To ensure $\mathbf{x}_{trial}$ lies within the search domain, some components of $\mathbf{dx}_p$ are suppressed whilst α is reduced. If $\mathbf{x}_{trial}$ is determined to represent an improvement with respect to the problem objectives, it is included into the population.

When considering that the agents $\mathbf{x}_p$, $\mathbf{x}_{p_1}$, $\mathbf{x}_{p_2}$ and $\mathbf{x}_{p_3}$ may have inconsistent vector lengths, it is clear that (6.49) cannot be directly evaluated. Instead, the grid intersection between all individuals is evaluated and used to construct directly comparable segments of constant vector length, to which (6.49) may then be successfully applied. In principle then, (6.49) is

applied only to masked sub-sequences of the selected individuals representing mutually shared segments, identified via the computation of dyadic grid point intersections in normalised time. Similarly as for the crossover operators, this modification implies that the effects of the new DE operation will naturally strengthen proportionally to the level of nodal intersection observed between solutions.

6.5 The VL-MOCP Algorithm

The components introduced in this work, namely the dyadic grid refinement (Section 6.2) and the variable-length crossover operations (Section 6.4), are applied as an extension to the existing MACS algorithm. The methods themselves are relatively general, thus MACS is here used primarily as a convenient test-bed with which to demonstrate their usability.

Figure 6.6 illustrates the complete process flow, with the proposed modifications to the existing MACS algorithm structure indicated with red dotted boxes. The complete algorithm operates as follows. A population of agents is initialised, where each represents a solution to a MOCP discretised upon a uniform initial grid using DFET (Vasile, 2010b; Ricciardi and Vasile, 2019a). Each agent is first modified in isolation using the individualistic actions of the original algorithm. Next, each agent is modified further using a selected variable-length crossover action, using grid-point intersections as a basis for identifying suitable crossover points. A bi-level optimisation structure (Ricciardi, 2019) is employed, where solutions produced by both the individualistic and social operations (i.e. comprising the outer-level) are initially considered only candidate solutions and passed to an inner-level local optimisation step. This local search operation simply finds the nearest solution to the candidate in parameter space that meets the current constraint tolerances, thus ensuring that the returned solution is fully feasible with minimal alteration to the original candidate. If the resulting solution then represents an improvement, it is included in the current population, otherwise discarded. At every user-specified number of iterations, the gradient-based step of the original algorithm is performed and immediately supplemented by the proposed dyadic grid refinement step. By doing so, local optimality may be guaranteed prior to the mesh refinement step. Given the useful properties of Bernstein basis functions described in Section 6.2, no further treatment of the population is necessary following grid refinement. This whole process repeats until either a maximum budget of function evaluations or total iterations is met.

By integrating these new steps into the original algorithm, the population is allowed to evolve into a set of multi-resolution solutions defined upon independent non-uniform discretisation grids. This then allows complex, nonlinear/discontinuous behaviour otherwise unique to any particular solution to be efficiently identified and captured as part of an evolutionary process.

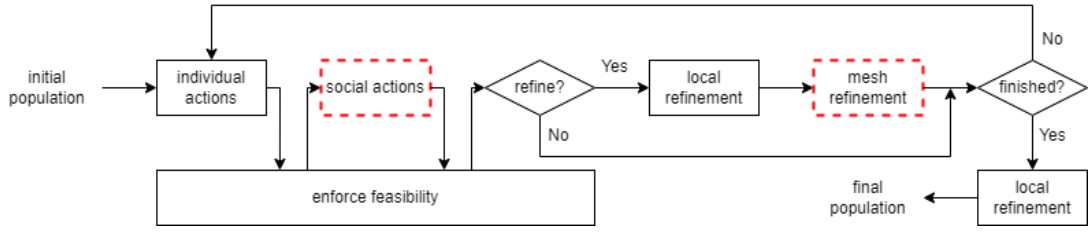


FIGURE 6.6: Simplified process flowchart of the modified multi-resolution MACS algorithm. Red dotted boxes indicate new/modified subroutines relative to the original MACS implementation. At each iteration, the population is subject to *individual* and *social* actions in turn, with a simplified local search applied after each to ensure the feasibility of each agent. If a refinement step in triggered, the population undergoes first a more extensive gradient-based local search to ensure local optimality, followed by the mesh refinement subroutine. This processes iterates until a stopping criteria is met, at which point the algorithm ends with a final local search applied to each agent.

6.6 Computational Experiments: Variable-Length Crossover

This section presents a computational test case intended to explore the effectiveness of the proposed VLC approach. An example MOCP is transcribed using DFET and solved with the modified MACS algorithm (incorporating the bi-level optimisation structure, as described in Section 6.5), where the original social operations are replaced with each crossover method described in Section 6.4, namely *1-Point Crossover* (1PCO), *2-Point Crossover* (2PCO) and *Dyadic Crossover* (DCO) in turn.

Two additional methods are included for reference. The first exchanges only the static parameters between the two selected parent solution vectors. In general, the static portion (e.g. design variables) of an Optimal Control solution vector is greatly outnumbered by the dynamic parameters (representing the system states and control law). As such, the lengths of the exchanged segments are comparable to those of the parent vectors themselves, thus this method is referred to here as *Near-Uniform Crossover* (NUCO). The second demonstrates an extreme case in which parents are simply swapped prior to the inner-level optimisation step, thus named *No Crossover* (NCO). In effect, this action represents a local optimisation step along a search direction different to that originally assigned to that particular agent[4]. The complete set of examined crossover methods thus demonstrates a clear hierarchy of complexity, whilst including the most prominent fundamental methods from the literature (i.e. 1PCO and 2PCO). This allows the individual effects of each level to be more explicitly observed.

Given the prevalence of additional heuristics within MACS however, it can be difficult to disseminate the global effects of any particular operation individually. For example, deficiencies

[4]The reader is referred to (Ricciardi, 2019) for a description of how the search direction is defined for each agent within the context of the MACS algorithm.

in one may be compensated for by another, for a particular problem. Furthermore, the deliberate isolation of heuristics for testing purposes may not give an accurate representation of their behaviour when employed in collaboration. In an attempt to provide as unbiased a performance metric as practical, the number of dominating solutions produced by the crossover action (including the inner-level gradient-based optimisation step of the bi-level approach) with respect to the current population is recorded as a measure of effectiveness. The likelihood of each approach to produce meaningful, globally exploratory offspring may thus be tracked in a relatively unbiased manner. Additionally, the quality and spread of the final Pareto-front is assessed against a reference Pareto-optimal set of solutions via the *Inverted Generational Distance* (IGD) and *Averaged Hausdorff Distance* (AHD)[5] metrics (Schütze et al., 2012).

Due to the inherent stochastic nature of the employed algorithms, 100 independent analyses were performed to track both the mean and standard deviation of relevant metrics. Additionally, to objectively indicate statistical significance, a one-way *Analysis of Variance* (ANOVA) was performed for each metric throughout the computation, where the significance level required to reject the null-hypothesis is taken as $\alpha = 0.05$. To subsequently assess the extent of the variation between individual pairs of crossover methods, a multiple comparison test was performed using the ANOVA output (Hochberg and Tamhane, 1987).

All computations in this section were performed using the ARCHIE-WeSt High Performance Computer. Individual runs utilised a single Intel Xeon Gold 6138 20 core 2.0GHz CPU, where 40 nodes comprise a standard compute node (on Lenovo SD30 servers) of 192GB RAM per node.

6.6.1 Chemical Reaction

The following problem is derived from the control of a chemical process, as presented in (Betts and Huffman, 1998; Citron, 1969; Jain and Tsiotras, 2008). The amount of chemical product y is determined via the two-stage reaction $x \rightarrow y \rightarrow z$, where z represents the waste product. The reaction rates for x and y are dependant on the rate coefficient $u(t)$ according to the following relationships:

$$\dot{x} = -ux \tag{6.52}$$

$$\dot{y} = ux - \rho u^k y \tag{6.53}$$

where the positive constants $\rho = 2.5$ and $k = 1.5$. A MOCP may be defined where the objective is thus to maximise the final amount of product y whilst minimising the peak reaction rate

[5] As remarked by Schütze et al. (2012), the IGD metric is sensitive to the number of elements in the reference Pareto front and in the computed one, hence the inclusion of the AHD in the presented comparisons.

u_{max}.

$$\min_{t_f,\mathbf{u}}[J_1, J_2]^T = [-y, u_{max}]^T \tag{6.54}$$

$$s.t. \tag{6.55}$$

$$0.1 \leq u(t) \leq u_{max} \tag{6.56}$$

where the static parameter $u_{max} \in [0.3, 0.5]$. The initial conditions are:

$$x(0) = 1, \quad y(0) = 0.01 \tag{6.57}$$

and the final time is fixed at $t_f = 2$.

This problem was solved by initialising a population of 10 individual agents, where each is a single continuous time phase discretised upon a uniform mesh of resolution level $J = 0$ (i.e. a single finite element). States and controls are represented by element specific 4th order Bernstein polynomials. The target local error threshold imposed at element boundaries was set to $\epsilon = 10^{-7}$, with a refinement step performed every 10 iterations. Elements not meeting this criteria are split at a frequency chosen to coincide with the gradient-based local searches (in this case, every 10 iterations). In this manner, local optimality may be guaranteed prior to each mesh refinement step. Local (inner-level) optimisation was completed using the fmincon solver with a SQP algorithm with function, constraint violation and step tolerances set to 1e-6, 1e-16 and 1e-9 respectively. The optimisation process was set to terminate upon reaching a maximum number of outer-loop iterations $n_{iter} = 200$. Figure 6.7 shows an example representation of the Pareto front obtained using the proposed multi-resolution approach, where the extremal solutions correspond to the profiles shown in Figure 6.4.

Figure 6.8 shows the accumulation of dominating solutions with respect to the current archive using each of the described crossover approaches. Several interesting features can be observed. Firstly, the 1PCO and 2PCO methods exhibit very similar rates of change, with the former displaced in the positive vertical direction due to an initially superior accumulation rate during early iterations. The gradual plateauing of each may be attributed to the associated evolution of NLP size (shown in Figure 6.9) which causes an increase in available grid point intersections and thus a decrease in the effectiveness of random crossover point pair selection. The effect of introducing the resolution-based bias can be clearly seen in the comparison between the 2PCO and DCO results, where the latter maintains an almost linear accumulation even as NLP size (and thus also the available grid intersections) begins converging. As expected, the NCO approach represents a performance baseline where no social manipulation is actually present, only a gradient-based local search in a different direction to that originally associated with that particular agent. The generality of this approach may go to explain the linear accumulation (albeit at a reduced rate) of dominating solutions irregardless of NLP size convergence. A slightly more interesting response is seen from the NUCO approach, with a larger accumulation

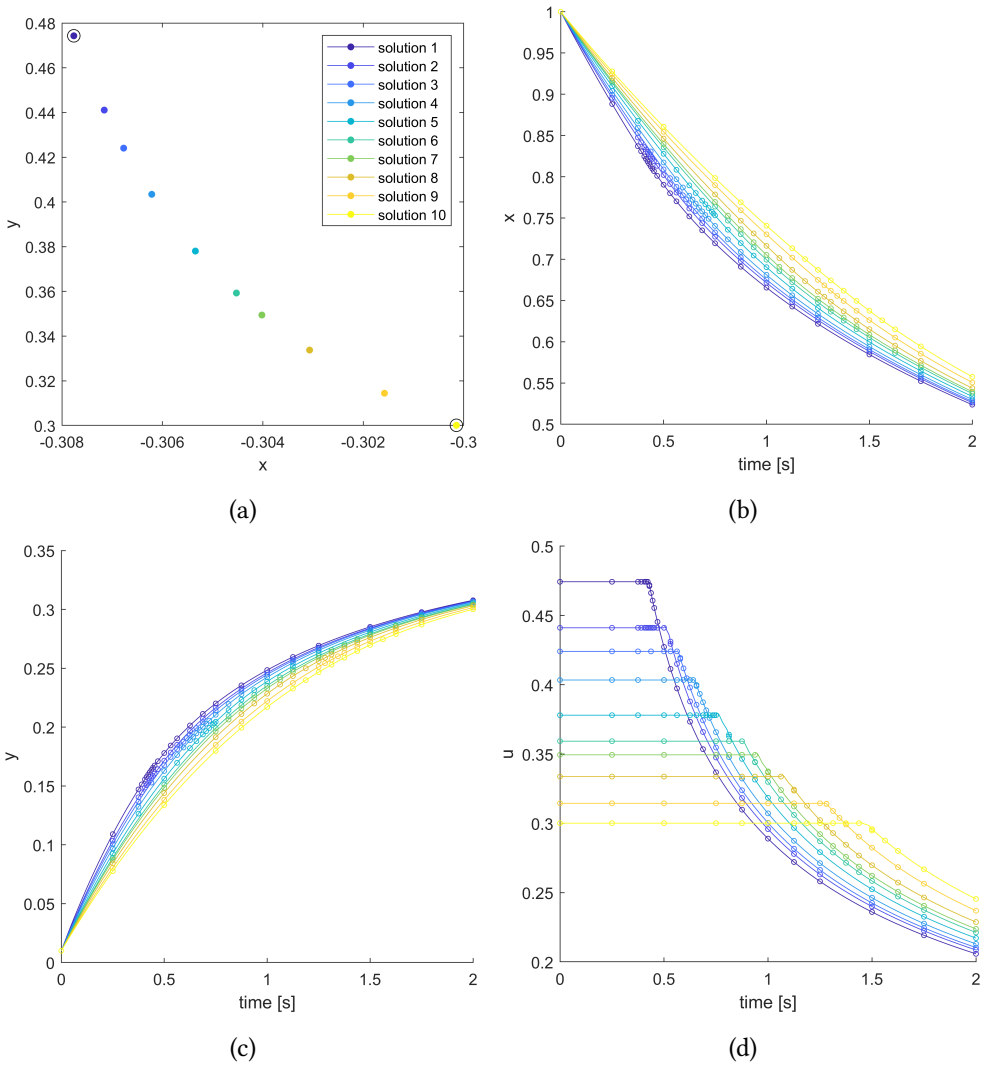


FIGURE 6.7: Example multi-objective solutions to Problem 6.6.1 (Chemical Reaction), solved using the multi-resolution approach with the DCO operator. Included is the (a) Pareto front with corresponding (b-c) state and (d) control profiles.

rate for $iter < 50$, followed by a reduced gradient for $iter > 50$, almost matching that of NCO. Given that NUCO is only exchanging the static parameters between parents, the initially larger accumulation rate may be attributed to the more exploratory nature of the earlier iterations, where the choice in static parameters plays a more definitive role in the positioning of agents within objective space. As the distribution of static parameter values across the population stabilises, the performance of NUCO reduces to the baseline level demonstrated by the NCO approach, interesting in the sense that the local search direction for these cases *is* in the originally associated direction.

To generally summarise these results, NLP size l_x and the number of dominating solutions

Table 6.1: Values of total execution time t_f, NLP size l_x and number of dominating solutions n_{soc} produced by each crossover operation at the final iteration $n_{iter} = 200$ for Problem 6.6.1 (Chemical Reaction).

	t_f [s]		l_x		n_{soc}	
1P	2948	(490)	321	(75)	2.3196	(1.26)
2P	3199	(581)	324	(75)	2.2717	(1.23)
D	2881	(606)	334	(89)	3.6932	(1.50)
N	1727	(327)	322	(77)	0.7010	(0.77)
NU	2827	(403)	320	(68)	1.1596	(0.99)

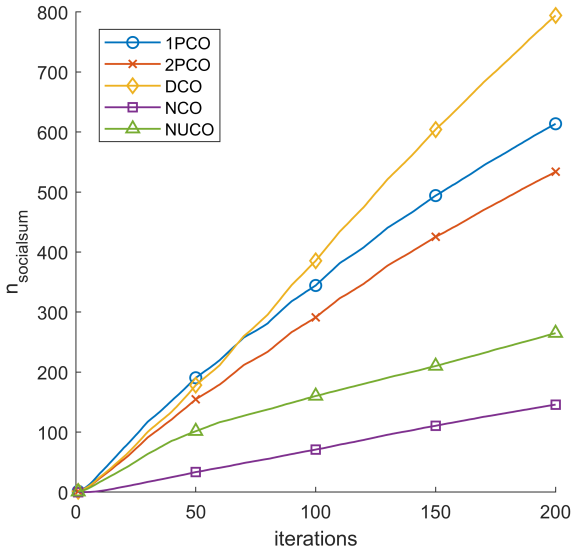
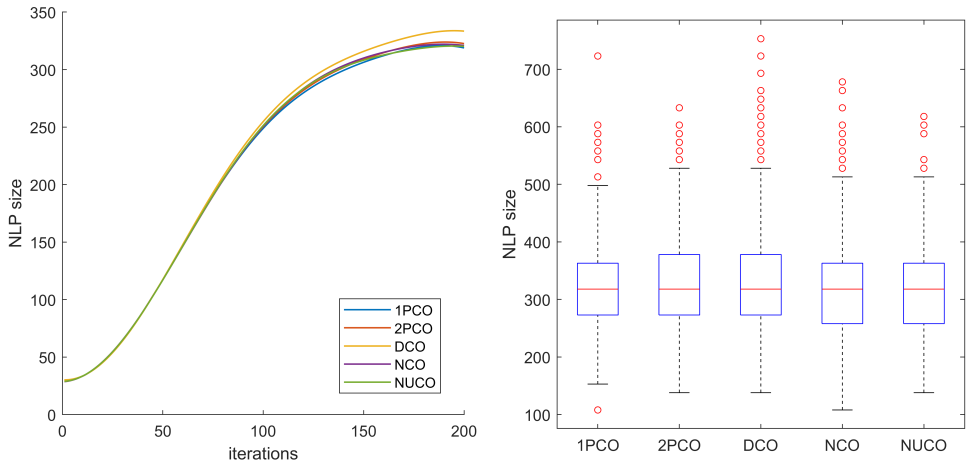


FIGURE 6.8: Total number of dominating solutions accumulated by each crossover method for Problem 6.6.1 (Chemical Reaction).

produced by the crossover operation n_{soc} at the final iteration of each approach are shown in Table 6.1, with the addition of the total recorded execution time t_f. Reported are the mean values with standard deviation included in brackets.

An acknowledged risk with variable-length representations in evolutionary algorithms is the uncontrolled growth of genome size (also referred to as "bloat" or "fluff") (Merlevede et al., 2019; Brameier, 2004; Langdon and Poli, 1998). It may be tempting then to attribute slight increase in average NLP size exhibited by the dyadic crossover approach to this excessive growth. However, by examining the distribution of NLP sizes recorded at the final iteration for each approach, shown in Figure 6.9, little difference between the methods can be discerned.

Ultimately this suggests that in this case the choice of crossover does not significantly influence the rate of evolution or the quality of the final population, with the reported difference in likelihood of producing new, dominating solutions largely mitigated by the bi-level approach.



(a) (b)

FIGURE 6.9: NLP size progression recorded using each crossover method for Problem 6.6.1 (Chemical Reaction), where (a) shows the fourier-fit trend of the averages across all iterations, and (b) shows the distributions recorded at the final iteration $n_{iter} = 200$.

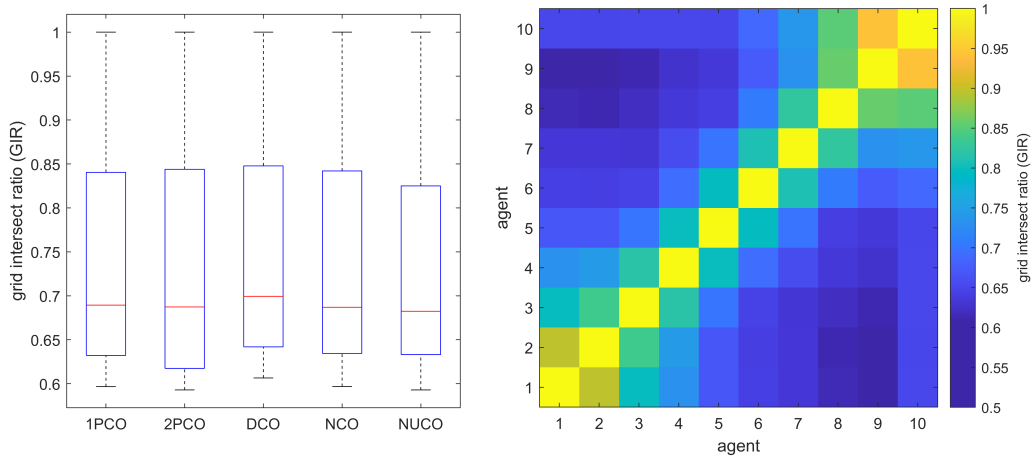
Table 6.2: Mean values (standard deviation in brackets) of IGD and AHD for each crossover method used to solve Problem 6.6.1 (Chemical Reaction).

	IGD		AHD	
1P	0.0030	(2.21e-3)	0.0262	(2.40e-3)
2P	0.0038	(2.69e-3)	0.0265	(2.17e-3)
D	0.0034	(2.99e-3)	0.0265	(2.29e-3)
N	0.0035	(2.42e-3)	0.0264	(2.45e-3)
NU	0.0028	(2.02e-3)	0.0262	(2.10e-3)

In part this suggests the problem is relatively simple for the optimiser to solve using a gradient based approach (albeit assuming some initial evolutionary spreading in objective space), or at least is computationally inexpensive enough for the multi-resolution approach to run for a sufficient number of iterations relative to the complexity of the problem. Whilst as expected given the additional heuristics within MACS (see Section 6.4), this is nonetheless supported by the general agreement in IGD and AHD values for each method, reported in Table 6.2.

Perhaps a more surprising aspect of Figure 6.9(b) is the range in final NLP size. Indeed, for each method the observed range is a significant proportion of the maximum observed value. To account for this, a closer inspection of the final multi-resolution population is necessary. As stated earlier, an important aspect of the original mesh refinement problem is that individual solutions distributed across the Pareto-front may require unique discretisation grids to fully represent the corresponding state and/or control profiles (this can partially be seen in the distribution of grid nodes in Figure 6.7). To quantitatively measure the extent

of this requirement, we define the *Grid Intersect Ratio* (GIR) as the fraction of intersection points between any two individuals relative to the total number of possible intersections (the latter in effect is equivalent to the total number of nodes of the smaller parent). Figure 6.10 shows an example of the pairwise correlation in GIR recorded by the dyadic crossover method at the final iteration. In this figure, the agent numbering corresponds with the numbering shown in Figure 6.7(a), thus agents with consecutive numbering also lie next to each other in objective space. The concentration of high intersect values across the diagonal of the agent pairing matrix supports the conjecture that "similar" solutions in objective space share a higher fraction of common grid points. Furthermore, the lowest observed value of GIR indicates that only around 60% of possible nodes are shared between the 'least similar' individuals of the final population, which then supports the large range in final NLP size seen in Figure 6.9.



(a) (b)

FIGURE 6.10: *Grid Intersect Ratio* (GIR) recorded at final iteration of Problem 6.6.1 (Chemical Reaction), where (a) shows the distribution of GIR relative to each crossover approach, and (b) shows the pairwise GIR between individuals within an example population solved using the dyadic crossover approach. Agent numbering corresponds to the numbering shown in Figure 6.7(a).

As an indication on how GIR varies between each of the presented crossover methods, Figure 6.10(a) shows the distribution of grid intersect fraction recorded at final iteration for each. An initial consideration was that the GIR could potentially be used as a measure of diversity across the population. For example, if a particular crossover method facilitated the generation/propagation of unique sequences throughout the optimisation process, the resulting population would exhibit on average a lower number of pairwise intersections. However, as seen earlier for this problem, little variation is seen between alternative methods, with the DCO method achieving only a slightly higher value of average GIR. The agreement in range suggests a similar distribution of pairwise GIR as observed for the DCO approach in Figure 6.10(b). Again, this ultimately suggests that the problem is relatively simple, with the

major differences between the described crossover methods largely mitigated by the bi-level approach.

6.6.2 Moon Lander

The following MOCP represents the controlled descent of a Lunar-landing vehicle, based on the formulation originally presented by Meditch (1964). The problem is to determine the control law $\mathbf{u}$ that minimises the total impulse acting on the vehicle and the maximum allowable control force:

$$\min_{t_f,\mathbf{u}}[J_1, J_2]^T = \left[\int_{t_0}^{t_f} u \, \mathrm{d}t, u_{max}\right]^T \tag{6.58}$$

The system state dynamics are given by:

$$\dot{h} = v \tag{6.59}$$

$$\dot{v} = -g + u \tag{6.60}$$

where h [m] is altitude, v [m/s] is velocity, g [m/s^2] is the gravity force and u [m/s^2] is the upwards control force. Boundary conditions are given as:

$$\begin{aligned} h(0) = h_0 = 10, \quad v(0) = v_0 = 2 \\ h(t_f) = h_f = 0, \quad v(t_f) = v_f = 0 \end{aligned} \tag{6.61}$$

with a path constraint on the control force defined by:

$$0 \leq u \leq u_{max} \tag{6.62}$$

The gravity of the Moon is taken as $g = 1.5$ m/s^2, the final time t_f is free and the maximum control force is an optimisable static parameter $u_{max} \in [0, 3]$.

This problem was solved by initialising a population of 10 individual agents, where each is a single continuous time phase discretised upon a uniform mesh of resolution level $J = 0$ (i.e. a single finite element). States and controls are represented by element specific 2nd order Bernstein polynomials. The target mesh error imposed at element boundaries was set to $\epsilon = 10^{-5}$, with a refinement step performed every 10 iterations. Elements not meeting this value were split at a frequency coinciding with the gradient-based local searches (every 10 iterations). In this manner, local optimality may be guaranteed prior to each mesh refinement step. Local (inner-level) optimisation was completed using the MATLAB® fmincon solver with a SQP algorithm where function, constraint violation and step tolerances set to 1e-6, 1e-12 and 1e-9 respectively. The MACS algorithm was set to terminate upon reaching a maximum number of outer-loop iterations $n_{iter} = 200$.

Figure 6.11 shows an example representation of the Pareto front obtained using the proposed approach. Included are the corresponding state and control profiles, where circle markers

represent element boundaries. This problem was chosen in part due to the inherent simplicity in formulation, whilst demonstrating several features of interest to the current study. For example, in Fig. 6.11 it can be seen that all solutions exhibit one clear discontinuity, corresponding to a *bang-bang* control structure and therefore suitable for analysis via an adaptive mesh approach. Furthermore, the location of this discontinuity in time changes as solutions move across one end of the Pareto-front to the other, a feature that should be captured by an effective variable-length crossover. This problem therefore represents a suitable platform to demonstrate and assess the *differences* between the included crossover approaches.

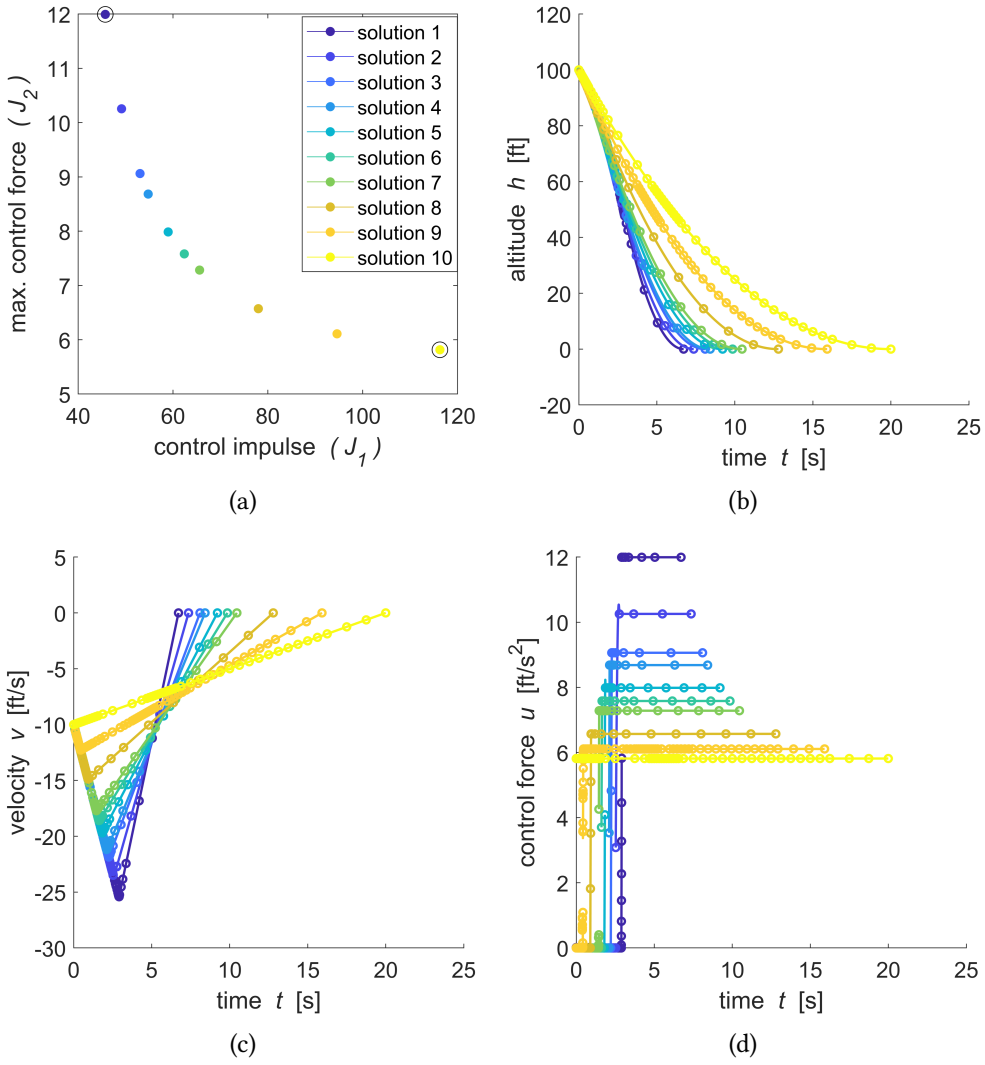


FIGURE 6.11: Multi-objective solutions, including (a) Pareto front and corresponding (b-c) state and (d) control profiles.

Figure 6.12 shows the accumulation of dominating solutions generated by each of the described crossover approaches. Several interesting features can be observed. Firstly, 1PCO and 2PCO exhibit rather similar rates of change. The gradual plateauing of each may be attributed to the

associated evolution of NLP size (shown in Fig. 6.13) which causes an increase in available grid point intersections and thus a decrease in the effectiveness of random point-pair selection. The effect of introducing the resolution-based bias can be clearly seen in the comparison between 2PCO and DCO, where the latter maintains an almost linear accumulation even as NLP size (and thus also the available grid intersections) begins converging. This may be due to the pronounced discontinuities seen in Fig. 6.11, in addition to the free (rather than fixed) final time t_f. The latter is important as the subsequent exchange of homologous segments identified in *normalised* time implies the exchange of *form* rather than simple time-equivalent numeric magnitudes. In this case then, the exchange of information via intersection-based crossover may have a higher potential of preserving/propagating important subsequences throughout the population. As expected, the NCO approach represents a performance baseline where no social manipulation is actually present. The generality of NCO may go to explain the linear accumulation (albeit at a reduced rate) of dominating solutions proportional to NLP size. Particularly interesting behaviour is seen from NUCO, initially accumulating non-dominated solutions at a higher rate than DCO, before exhibiting a plateauing similar to that of 1PCOand 2PCO. Given that NUCO is only exchanging static parameters (e.g. the free final time t_f), the initially larger accumulation rate may be attributed to the more exploratory nature of earlier iterations, where the choice in static parameters plays a more definitive role in the positioning of agents within objective space. As the distribution of static parameter values across the population stabilises, the rate of accumulation reduces towards the baseline level demonstrated by the NCO approach, interesting in the sense that the local search direction for NUCO *is* in the originally associated search direction.

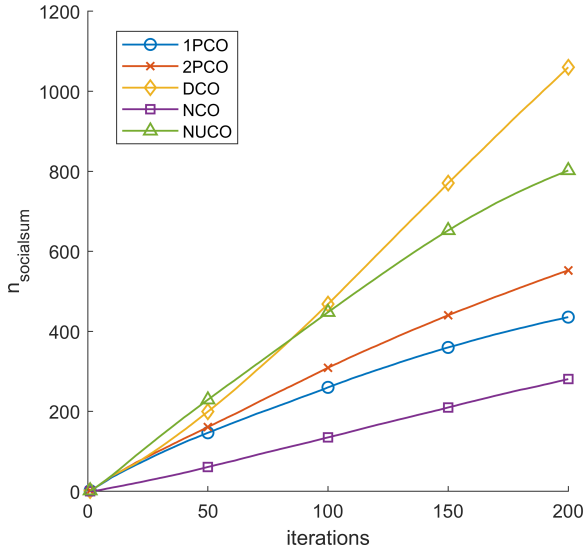


FIGURE 6.12: Total number of dominating solutions accumulated by each crossover method.

The NLP size l_x and the number of dominating solutions (with respect to the current population) produced by the crossover operation n_{soc} at the final iteration ($n_{iter} = 200$) of each approach are shown in Table 6.3, as well as the total recorded execution time t_f. Reported are the mean values across 100 separate analyses (standard deviation included in brackets). A clear hierarchy can be seen, where the DCO approach seems to offer significant improvement relative to the others. Indeed, DCO exhibits execution times approaching the values of NCO, which itself could be taken as the theoretical minimum given that no recombination actually occurs.

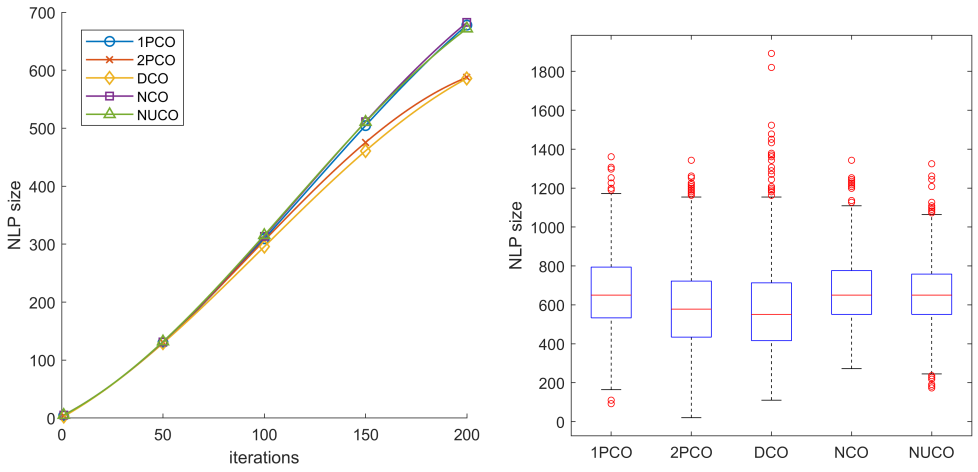
Specifically considering the evolution of NLP size, two distinct groups are present (see Fig. 6.13). Interestingly, the first includes the single-point methods[6] and the second includes the multi-point methods. Whilst a lower average NLP size may suggest a generally more efficient distribution of grid nodes using DCO, the larger quantity and spread of outliers nonetheless suggests a remaining potential for excessive NLP 'bloat' (an acknowledged risk with variable-length representations in evolutionary algorithms (Merlevede et al., 2019)). Of note here is that in this work, NLP bloat is treated only indirectly through the user-specification of mesh-error tolerance. That is, an appropriate tolerance value prevents the algorithm from actively adding excessive new nodes. However, it is acknowledged that this does not prevent heavily refined segments from being relocated by the crossover mechanism to areas within the solution vector that do not require such levels of detail. The explicit identification and removal of unnecessary nodes is a complex problem and thus originally considered outwith the scope of this study, though remains an important subject of future work.

Table 6.3: Values of total execution time t_f, NLP size l_x and number of dominating solutions n_{soc} produced by each crossover operation at the final iteration $n_{iter} = 200$.

Method	t_f [s]		l_x		n_{soc}	
1P	10496	(3.11e+3)	669	(196)	1.22	(1.03)
2P	8944	(2.67e+3)	584	(218)	2.19	(1.31)
D	7761	(2.55e+3)	582	(238)	5.82	(1.58)
N	7572	(2.90e+3)	673	(177)	1.34	(0.82)
NU	9754	(2.58e+3)	663	(177)	2.49	(1.30)

Perhaps a more surprising aspect of Fig. 6.13(b) is the range in final NLP size. Indeed, for each method the observed range is a significant proportion of the maximum observed value. As stated earlier, an important feature of the original mesh refinement problem is that individual solutions distributed across the Pareto-front may require unique discretisation grids to fully represent the corresponding state and/or control profiles (see Fig. 6.11). To quantitatively measure the extent of this requirement, we define the GIR as the fraction of intersection points between any two individuals relative to the total number of possible intersections. Figure 6.14(b) shows an example of the pairwise correlation in GIR recorded by the DCO method

[6]For the purpose of comparison, NCO can be considered a single-point method assuming the selected point is either of the extremal vector loci.



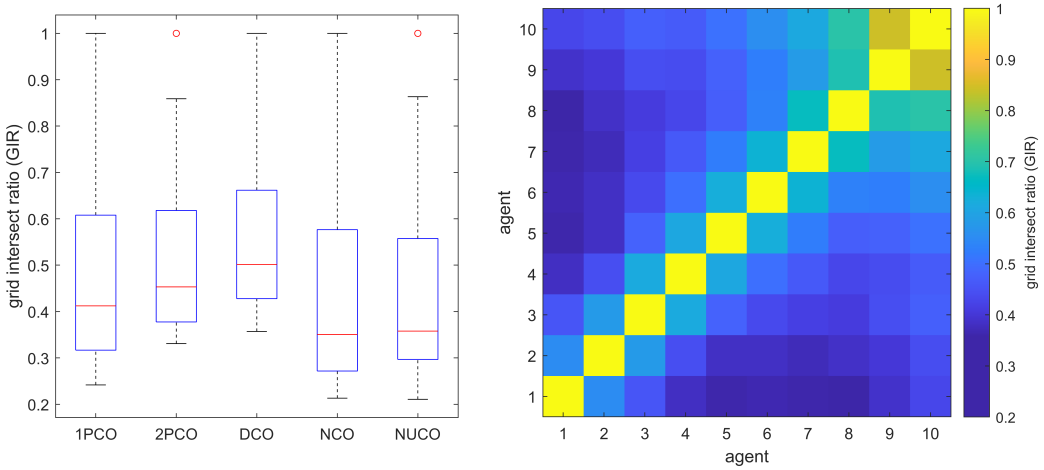
(a) (b)

FIGURE 6.13: NLP size progression recorded during 100 separate runs using each crossover method, where (a) shows the fourier-fit trend of the averages across all iterations, and (b) shows the distributions at the final iteration.

at the final iteration. As in the previous example, agent numbering corresponds with the numbering shown in Figure 6.11(a). The concentration of higher intersect values surrounding the diagonal of the agent pairing matrix supports the conjecture that similar solutions in objective space share a higher fraction of common nodes. Furthermore, the lowest observed value of GIR indicates that only around 30% of possible nodes are shared between the 'least similar' individuals of the final population, which may support the large range in final NLP size seen in Fig. 6.13(b).

To indicate how GIR varies between each method, Fig. 6.14(a) shows the respective distributions recorded at ($n_{iter} = 200$). One might expect a higher GIR to suggest a decrease in structural diversity across the population. That is, if two grids share a larger number of nodes, they are inherently less unique (a GIR of 1 indicates one is a complete subset of the other). However, a complicating factor is the associated NLP size. If a population exhibits a lower GIR but a larger average NLP size, this could suggest over-refinement. Conversely, a higher GIR coupled with a lower NLP size may suggest under-refinement, in that the population may not yet have achieved a genetic diversity complementary to the problem complexity. In part these results suggest the problem is relatively simple for the optimiser (albeit assuming some initial spreading in objective space), or at least is inexpensive enough to complete a sufficient number of iterations. Whilst as expected given the additional heuristics within MACS, this is nonetheless supported by the general agreement in IGD and AHD values between methods, see Table 6.4.

This agreement and it's relation to the complexity of the example problem can be further



(a) (b)

FIGURE 6.14: *Grid Intersect Ratio* (GIR) recorded at the final iteration ($n_{iter} = 200$), where (a) shows the distribution of average values for each crossover method, and (b) shows the pairwise GIR across an example population solved using the DCO method. Agent numbering corresponds with the numbering shown in Figure 6.11(a).

Table 6.4: Mean values (standard deviation in brackets) of IGD and AHD for each crossover method.

Method	IGD		AHD	
1P	0.0927	(3.32e-2)	0.5290	(1.11e-1)
2P	0.0926	(2.90e-2)	0.5367	(1.06e-1)
D	0.0940	(2.81e-2)	0.5298	(9.40e-2)
N	0.0921	(2.50e-2)	0.5029	(1.08e-1)
NU	0.1016	(4.18e-2)	0.5505	(1.16e-1)

understood by considering the results of the statistical analyses.

An *Analysis of Variance* (ANOVA) test compares the mean values of several sets of data (Hochberg and Tamhane, 1987). The 'null' hypothesis that each set has been drawn from the same distribution is directly tested against the competing hypothesis that they are not. A small p-value indicates that the differences in mean values between sets are significant. However, ANOVA results by themselves do not indicate which specific sets this applies to, thus a pairwise *multiple comparisons* test can be employed to identify the specific sets that are likely to be drawn from different distributions. For the purposes of this study, this methodology can be used to determine which crossover approach results in statistically significant changes in behaviour of the social operator. Table 6.5 shows the results of a multiple comparisons test (using Tukey's honestly significant difference criterion), where the first column indicates the specific pair of methods being directly compared through ANOVA, and the remainder displaying the p-values for each performance metric at $n_{iter} = 200$. A p-value smaller than

the user-selected significance level α (commonly 0.05 or 0.01) indicates that at least one of the sample set means is significantly different from the others. In this case, $\alpha = 0.05$.

Firstly, it can be seen that the numbers of non-dominated solutions n_{soc} produced by the DCO method is statistically significant relative to every other method (indicated by p-values less than 0.05). Indeed, it was found that only 40 intermediate iterations were required for this to be true. Secondly, the p-values reported for NLP size support the grouping seen in Fig. 6.13(a), in that the null-hypothesis can only be rejected between a single-point method and a multi-point method. A similar distinction is seen for GIR. Finally, and perhaps most importantly, the p-values of AHD and IGD suggest that there is no significant difference between the quality of final population obtained using any of the included crossover methods. When considered as a whole, these results ultimately suggest that the DCO method is exhibiting a meaningful improvement in the exploration of the design space, but not enough to affect the final convergence of the algorithm in the presence of the additional heuristics. This again highlights the need for more complicated test problems to assess whether the observed effectiveness of DCO (even if strictly internal in this case) is maintained.

Table 6.5: Results of an ANOVA multiple-comparisons test between each combination of the presented crossover methods. The p-values for each metric are shown for each pairwise combination of the examined methods. Statistical significance is indicated where $p < 0.05$).

Methods		p-value ($n_{iter} = 200$)				
		n_{soc}	NLP size	GIR	AHD	IGD
1P	2P	3.67e-6	5.50e-19	0.00	0.96	0.76
1P	D	0	1.48e-20	6.16e-9	0.51	0.37
1P	N	0.90	0.99	0.95	0.95	1.00
1P	NU	2.96e-8	0.96	0.78	0.71	0.98
2P	D	0	1.00	9.29e-2	0.89	0.97
2P	N	2.77e-4	1.53e-21	9.84e-5	1.00	0.80
2P	NU	0.92	5.60e-16	1.24e-5	0.97	0.97
D	N	0	2.59e-23	4.47e-11	0.92	0.41
D	NU	0	2.49e-17	1.73e-12	1.00	0.74
N	NU	4.84e-6	0.82	0.99	0.98	0.99

6.6.3 Discussion

In general, it is found that the proposed approaches successfully promote offspring diversity throughout an open-ended evolutionary process. This is achieved by allowing a progressively finer level of social action to occur between similar solutions, thus reducing the overall genetic difference between successive generations. A positive effect on the execution time of the algorithm is observed, as well as a reduction in the average NLP size across the population for the same solution quality. It is found that a bias towards exchanging higher resolution segments consistently produces more dominating solutions than traditional, unbiased crossover. However, the effects of these improvements on the final solution are found to be negligible,

likely due to the simplicity of the example problems. Future work should include more complex scenarios such that crossover effectiveness becomes more important to the host algorithm. Excessive NLP growth remains a pressing issue outwith passive/indirect mitigation. Future study could include an adaptive approach to validate growth, detect over-refinement and ultimately remove unnecessary nodes without incurring additional expense or loss of crucial information.

A significant factor in these findings is the influence of additional heuristics within the host algorithm. For example, even if the crossover operation successfully produces a dominating solution with respect to the current population, inclusion into the population is not guaranteed without first demonstrating (via substitution with an existing archived individual) a collective improvement as per the archiving strategy implemented within MACS (Ricciardi and Vasile, 2019b). This further reliance on the *positioning* of candidate solutions within objective space, as opposed to merely the diversity or quality of the genetic information contained within, may significantly dampen the exchange/redistribution of genetic features across successive generations. This in turn affects the genetic material available for redistribution during the next iteration. Whilst partly analogous to a natural evolutionary process, this fundamentally makes it difficult to quantitatively isolate and evaluate the global effects of any particular operation, including the proposed crossover methods. It is of course true that this issue may not be fully due to the inherent complexity of the host algorithm. To maximise the efficiency of an evolutionary optimisation, high-fidelity is necessary to preserve homologous information. However, this feature of crossover is common, and can indeed be achieved even by simple algorithms (Merlevede et al., 2019). We may conclude then that despite these misgivings, the results show that the inference of homology via the genetic structure of parental genomes is at least more likely in this case to allow and facilitate meaningful, non-destructive crossover through the redistribution of genetic information. The presented methodology may offer comparable savings in computational expense as would normally be associated with discipline-based multi-fidelity optimisation. Specific investigations into this aspect remains the subject of future work.

6.7 Computational Experiments: Grid Resolution

This section presents three test cases intended to demonstrate the usability of the proposed multi-resolution refinement procedure for MOCPs. For each examined case, the proposed approach is compared with a traditional uniform grid approach at successive levels of resolution. Algorithm performance is measured by tracking the mean and standard deviation of the maximum relative local mesh error, overall execution time, number of function evaluations and minimum achieved objective function values at each iteration. The quality and spread of the final Pareto-front approximation in each case is assessed against a reference Pareto-optimal

set of solutions, again via the IGD and AHD metrics.

All computations in this section were performed using the ARCHIE-WeSt High Performance Computer. Individual runs utilised a single Intel Xeon Gold 6138 20 core 2.0 GHz CPU, where 40 nodes comprise a standard compute node (on Lenovo SD30 servers) of 192 GB RAM per node. All results for each case are averaged over 10 independent analyses.

6.7.1 Drug Displacement

This example considers the optimisation of a continuous chemical process representing the interaction of drugs *warfarin* and *phenylbutazone* in a patient's bloodstream (Maurer and Wiegand, 1992). The problem is to control the rate of *phenylbutazone* infusion such that the desired level of the two drugs can be reached in the minimum time provided that the concentration of *warfarin* does not rise above a certain toxic level. To examine the inherent trade-off between time and toxicity, a MOCP may be formulated where the maximum allowable toxicity level to be reached during the dynamic process is itself a static variable, α, to be minimised. The problem may thus be written as:

$$\min_{t_f,\mathbf{u}}[J_1,\ J_2]^T = [t_f,\ \alpha]^T \tag{6.63}$$

subject to the state dynamics:

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = \frac{C_1^2[C_3(0.02 - x_1) + 46.4x_1(u - 2x_2)]}{C_4} \tag{6.64}$$

$$\frac{\mathrm{d}x_2}{\mathrm{d}t} = \frac{C_1^2[C_2(u - 2x_2) + 46.4x_2(0.02 - x_1)]}{C_4} \tag{6.65}$$

where the state variables x_1 and x_2 represent the concentrations of *warfarin* and *phenylbutazone* respectively, u is the rate of *phenylbutazone* infusion, and C_1, C_2, C_3, and C_4 are defined as follows:

$$C_1 = 1 + 0.2x_1 + 0.2x_2 \tag{6.66}$$

$$C_2 = C_1^2 + 232 + 46.4x_2 \tag{6.67}$$

$$C_3 = C_1^2 + 232 + 46.4x_1 \tag{6.68}$$

$$C_4 = C_2C_3 - 46.4^2x_1x_2 \tag{6.69}$$

Initial and final conditions are specified as:

$$\begin{aligned} x_1(0) &= 0.02, \quad x_2(0) = 0 \\ x_1(t_f) &= 0.02, \quad x_2(t_f) = 2 \end{aligned} \tag{6.70}$$

and the following control bounds are imposed:

$$0 \leq u(t) \leq 8 \tag{6.71}$$

The concentration of *warfarin* is constrained according to:

$$x_1(t) \leq \alpha \tag{6.72}$$

where the static variable $\alpha \in [0, 0.03]$ is the allowable upper limit.

States and controls are represented by element specific 7th order Bernstein polynomials. 10 individual agents were initially discretised upon a uniform grid at resolution level $J_{min} = 0$, and again evolved according to the proposed multi-resolution modification to the MACS algorithm. A relative mesh error of $\epsilon = 10^{-6}$ was imposed at element boundaries, with a mesh refinement frequency set to coincide with the gradient based local search operation (i.e. every 10 iterations). Local (inner-level) optimisation was completed using the MATLAB fmincon solver with a SQP algorithm where function, constraint violation and step tolerances were set to 1e-6, 1e-12 and 1e-9 respectively. The optimisation was terminated upon reaching a specified maximum number of iterations, $n_{iter,max} = 200$.

Fig. 6.15 shows the time histories of the states x_1 and x_2, and the control u for an example set of solutions, where circle markers represent element boundaries. Fig. 6.16 shows the corresponding dyadic distributions of collocation points for both orthogonal solutions with respect to resolution level J across a normalised time domain. This case features two distinct control discontinuities that can be seen to move further apart in normalised time across the Pareto front. The proposed approach can be seen to concentrate the additional collocation points around both discontinuities throughout the population. In this manner, the resultant effects on the system state are more accurately captured. Extremal objective function values correspond to those reported in Zhao and Shang (2018); Maurer and Wiegand (1992).

Table 6.6 presents the performance comparisons between the multi-resolution approach and a globally dense uniform grid approach, the latter completed at successive levels of resolution. The multi-resolution approach is seen to temper the inherent trade-off between solution accuracy and computational cost demonstrated by the uniform grid examples. To illustrate, the highest level of dense grid considered here, $J = 5$, was unable to reach the same level of maximum relative mesh integration error as the multi-resolution approach within the fixed number of iterations, despite the latter terminating with 44.5% of the computational cost. Furthermore, as in the previous example, there seems to be no strong benefit to the dense grid approach when considering the orthogonal objective function solutions, with only a 0.81% improvement in J_1 and an identical J_2 result.

The mean values and corresponding standard deviations of IGD and AHD for each method are reported in Table 6.7. In this case, the reference Pareto-front is taken directly from Maurer and Wiegand (1992). A reasonable agreement between all methods is observed for both metrics, with the multi-resolution approach achieving a slightly superior IGD (mean and variance),

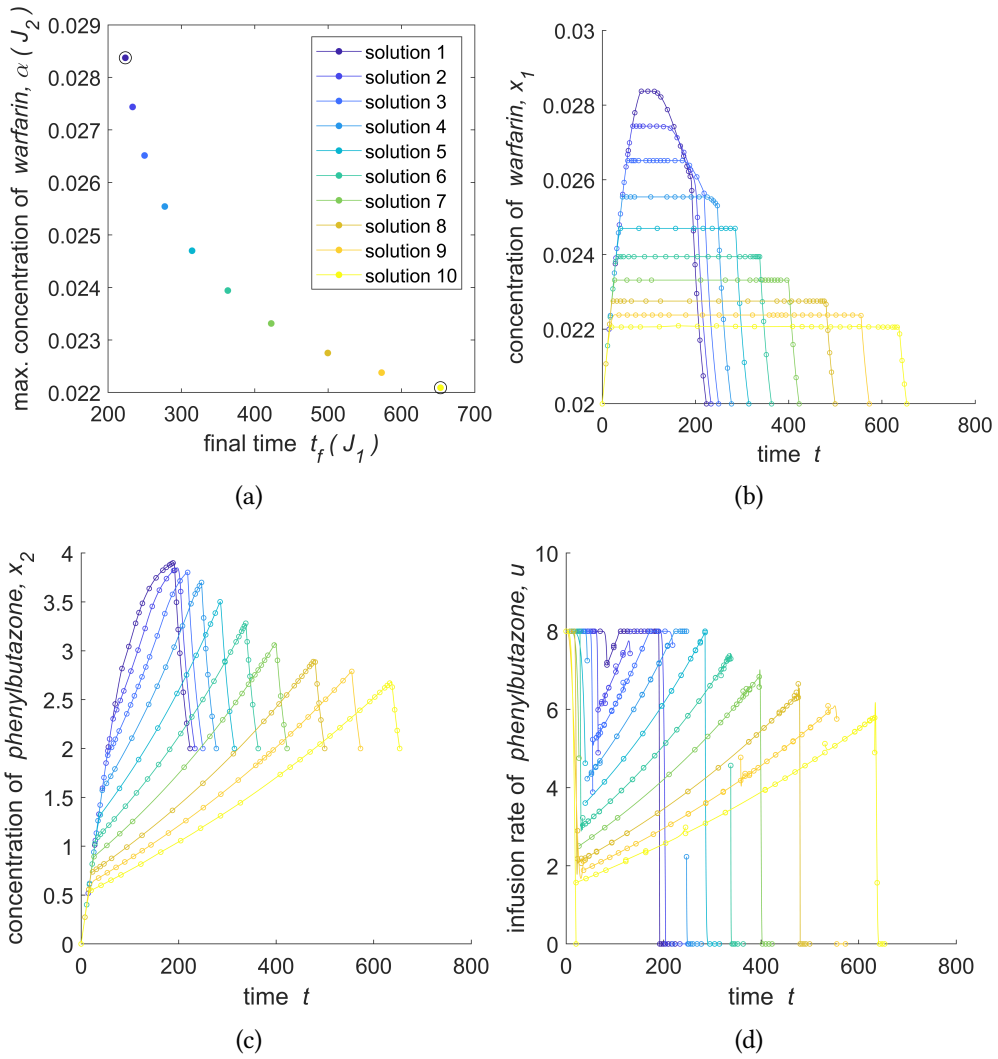
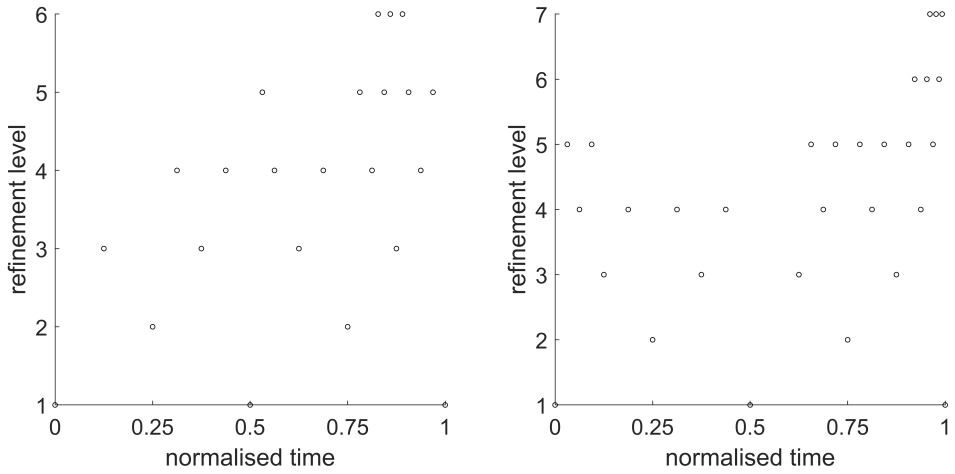


FIGURE 6.15: Multi-objective solutions to Problem 6.7.1 (Drug Displacement), including (a) Pareto front and corresponding (b-c) state and (d) control profiles.

Table 6.6: Results for Problem 6.7.1 (Drug Displacement), recorded at $n_{iter} = 200$. Reported values are 10-run averages, with standard deviations included in brackets.

Resolution level		Mesh error		No. func. evals		Run time		Objectives
J_{min}	J_{max}	ϵ_{max}		n_{feval}		[s]		$\mathbf{J}_{best}$
2	2	5.50e-4	(1.79e-4)	**15698**	**(1.17e+3)**	**1081**	**(4.73e+1)**	[226.63 (2.98e+0), 0.0221 (1.73e-4)]
3	3	3.80e-4	(9.36e-5)	29308	(3.62e+3)	2229	(2.30e+2)	[227.19 (9.20e+0), **0.0220** **(4.50e-5)**]
4	4	1.28e-4	(2.76e-5)	52643	(7.04e+3)	6541	(7.12e+2)	[223.08 (9.00e-1), 0.0221 (1.69e-4)]
5	5	4.77e-5	(1.39e-5)	92637	(7.25e+3)	40965	(4.09e+3)	[**222.35** **(5.74e-1)**, 0.0221 (1.41e-4)]
0	8	**2.01e-5**	**(8.93e-6)**	78239	(9.16e+3)	18209	(4.60e+3)	[224.16 (3.39e+0), 0.0221 (1.15e-4)]

and the lowest uniform mesh resolution displaying the strongest value of AHD (though the value range is much less significant). Of note in this case, the values of standard deviation for both metrics are superior for the multi-resolution approach to a greater degree than the mean values themselves, suggesting greater repeatability.



(a) (b)

FIGURE 6.16: Dyadic distribution of collocation points for the orthogonal solutions to (a) objective 1 and (b) objective 2 of Problem 6.7.1 (Drug Displacement).

Table 6.7: Mean values (standard deviation in brackets) for the IGD and AHD for multi/uniform-resolution solutions to Problem 6.7.1 (Drug Displacement).

Resolution level		Inverted Generational Distance (IGD)		Averaged Hausdorff Distance (AHD)	
J_{min}	J_{max}				
2	2	5.02	(4.41)	**43.93**	(6.98)
3	3	4.55	(1.43)	44.86	(4.11)
4	4	5.84	(4.68)	44.99	(6.74)
5	5	5.15	(3.68)	44.49	(4.03)
0	8	**4.04**	(**0.81**)	44.68	(**2.61**)

Fig. 6.17 shows the values of maximum relative mesh integration error and AHD taken at each iteration for each process. The multi-resolution approach is seen to traverse the hierarchy of mesh error values observed for the uniform mesh approaches, whilst simultaneously achieving lower AHD values in less time. This combined behaviour may be more explicitly depicted via the combined metric shown in Fig. 6.17 (c).

Fig. 6.18 shows the accumulation of total CPU time and objective function evaluations relative to the number of completed iterations. As in the previous example, the progression of average agent complexity is clearly reflected in the accumulation of total execution time, beginning at very low cost per iteration for coarser grids, before traversing upwards to reach levels more comparable with the dense uniform grids. This is additionally reflected in the greater accumulation of objective function values in the latter part of the optimization process. The described open-ended evolution of the population is shown in Fig. 6.19. With reference to Table 6.6, the multi-resolution approach is again seen to represent higher levels of detail within

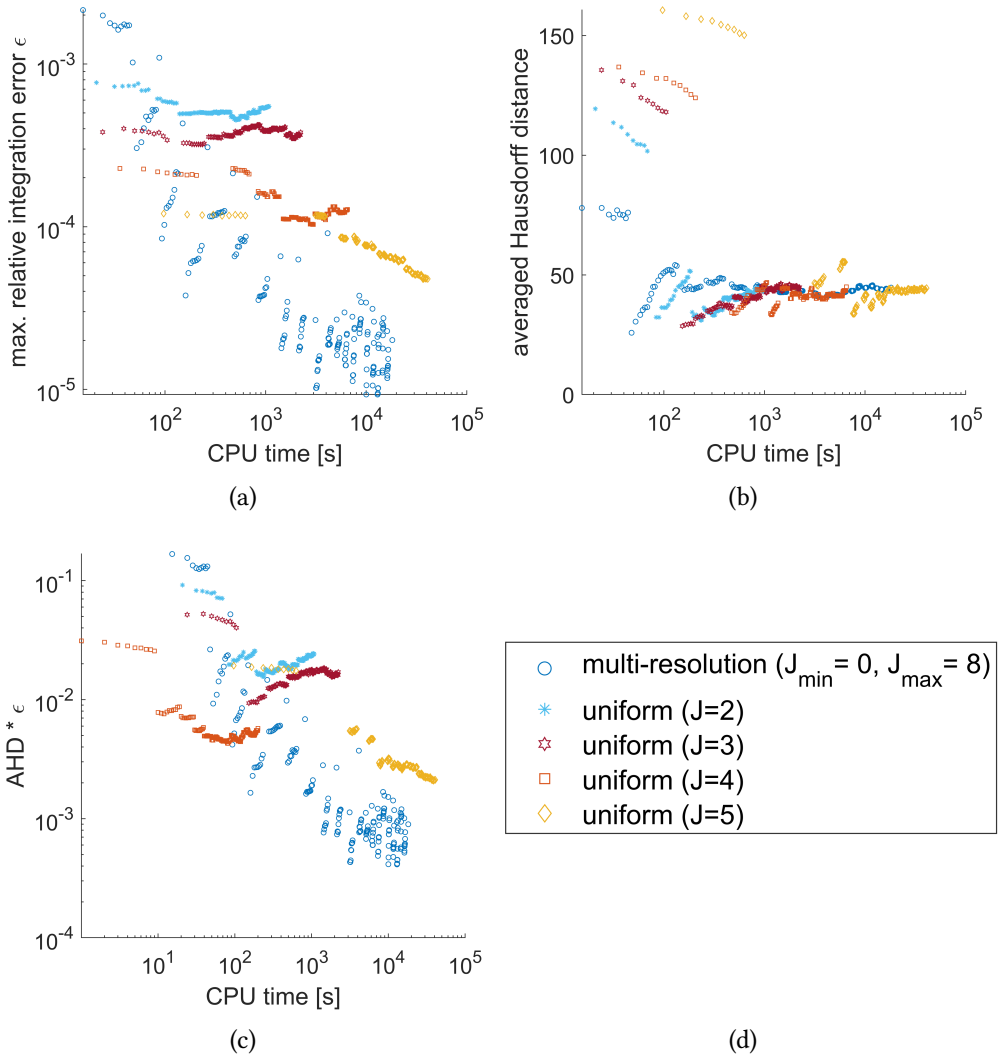


FIGURE 6.17: Problem 6.7.1 (Drug Displacement) mesh integration error and AHD relative to the total CPU time.

the solution vector whilst maintaining a relatively low overall length.

6.7.2 Apollo Re-entry

This section presents a demonstration of the proposed multi-resolution method applied to a MOCP representing the controlled re-entry of an Apollo-type space vehicle (Pesch, 1989). This is an established benchmark problem in trajectory optimisation that is known to be very challenging, owing to its sensitivity in terms of the initial guesses (Jain and Tsiotras, 2008). This problem includes an inequality constraint on the control variable and a variety of alternative sharp/switching structures.

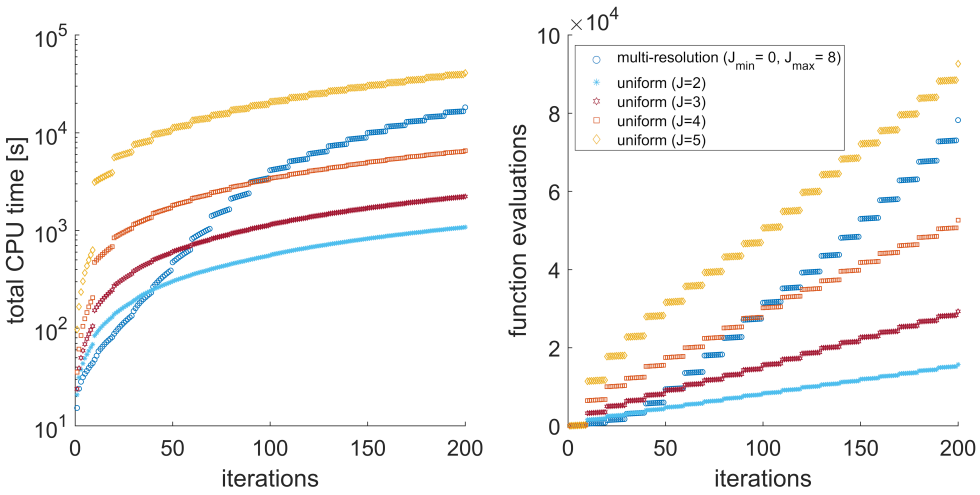


FIGURE 6.18: (a) Total execution time and (b) number of function evaluations of Problem 6.7.1 (Drug Displacement) relative to the number of completed iterations.

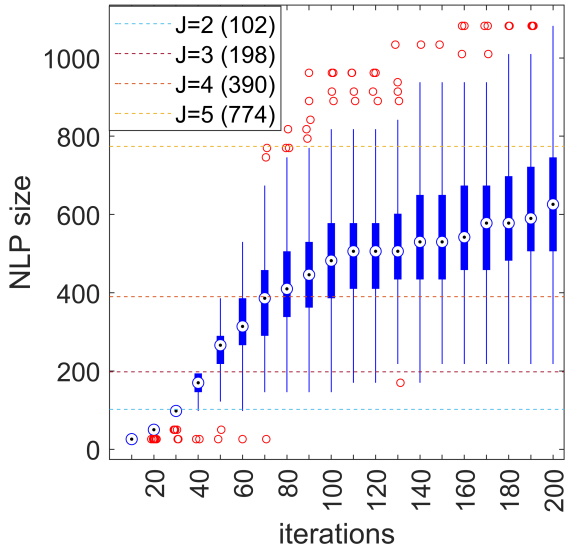


FIGURE 6.19: NLP problem size across the included population relative to the number of completed iterations of Problem 6.7.1 (Drug Displacement). For simplicity, only iterations immediately following a mesh refinement step are shown.

The equations of motion during atmospheric flight are given as:

$$\dot{v} = -\frac{S}{2m}\rho v^2 C_D(u) - \frac{g\sin\gamma}{(1+\xi)^2} \tag{6.73}$$

$$\dot{\gamma} = \frac{S}{2m}\rho v C_L(u) + \frac{v\cos\gamma}{R(1+\xi)} - \frac{g\cos\gamma}{v(1+\xi)^2} \tag{6.74}$$

$$\dot{\xi} = \frac{v}{R}\sin\gamma \tag{6.75}$$

$$\dot{\zeta} = \frac{v}{1+\xi}\cos\gamma \tag{6.76}$$

where v is the vehicle velocity, γ is the flight-path angle, $\xi = h/R$ is the normalised altitude, h is the altitude above the Earth's surface, R is the Earth's radius, ζ is the distance travelled across the Earth's surface, S is the reference surface area and m is the vehicle mass. The control variable u is the angle of attack. Lift and drag forces are defined according to the following relations:

$$C_D = C_{D_0} + C_{DL} \cos u \tag{6.77}$$

$$C_L = C_{L_0} \sin u \tag{6.78}$$

Where $C_{D_0} = 0.88$, $C_{DL} = 0.52$ and $C_{L_0} = -0.505$. The air density is assumed to satisfy the relationship $\rho = \rho_0 e^{-\beta R \xi}$. The values of the included constants are specified as:

$$\begin{aligned} R &= 209.0352(10^5 \text{ft}) \\ S/m &= 50,000(10^{-5} \text{ft}^2 \text{slug}^{-1}) \\ \rho_0 &= 2.3769 \times 10^{-3} (\text{slug ft}^{-3}) \\ g &= 3.2172 \times 10^{-4} (10^5 \text{ft s}^{-2}) \\ \beta &= 1/0.235 (10^{-5} \text{ft}^{-1}) \end{aligned}$$

The first objective function to be minimised describes the total stagnation point convective heating per unit area, given by the integral:

$$J_1 = \int_0^{t_f} 10 v^3 \sqrt{\rho} \, \mathrm{d}t \tag{6.79}$$

A multi-objective problem is formulated by including the minimisation of total time of flight as an additional objective, i.e., $J_2 = t_f$. The complete MOCP is therefore:

$$\min_{t_f, \mathbf{u}} [J_1, J_2]^T \tag{6.80}$$

The initial conditions at the moment of atmospheric entry are given as:

$$\begin{aligned} v(0) &= 0.35(10^5 \text{ft s}^{-1}) \\ \gamma(0) &= -5.75 \, \text{deg} \\ \xi(0) &= 4/R \quad (h(0) = 400,000 \text{ft}) \\ \zeta(0) &= 0(10^5 \text{ft}) \end{aligned}$$

with corresponding final-time conditions given as:

$$\begin{aligned} v(t_f) &= 0.00165(10^5 \text{ft s}^{-1}) \\ \gamma(t_f) &= \text{ free} \\ \xi(t_f) &= 0.75530/R \quad (h(t_f) = 75,530 \text{ft}) \\ \zeta(t_f) &= 51.6912(10^5 \text{ft}) \end{aligned}$$

Finally, a control inequality constraint that limits the deceleration of the vehicle is imposed:

$$|u| \leq u_{max}, \quad u_{max} > 0 \tag{6.81}$$

where $u_{max} = 180\text{deg}$.

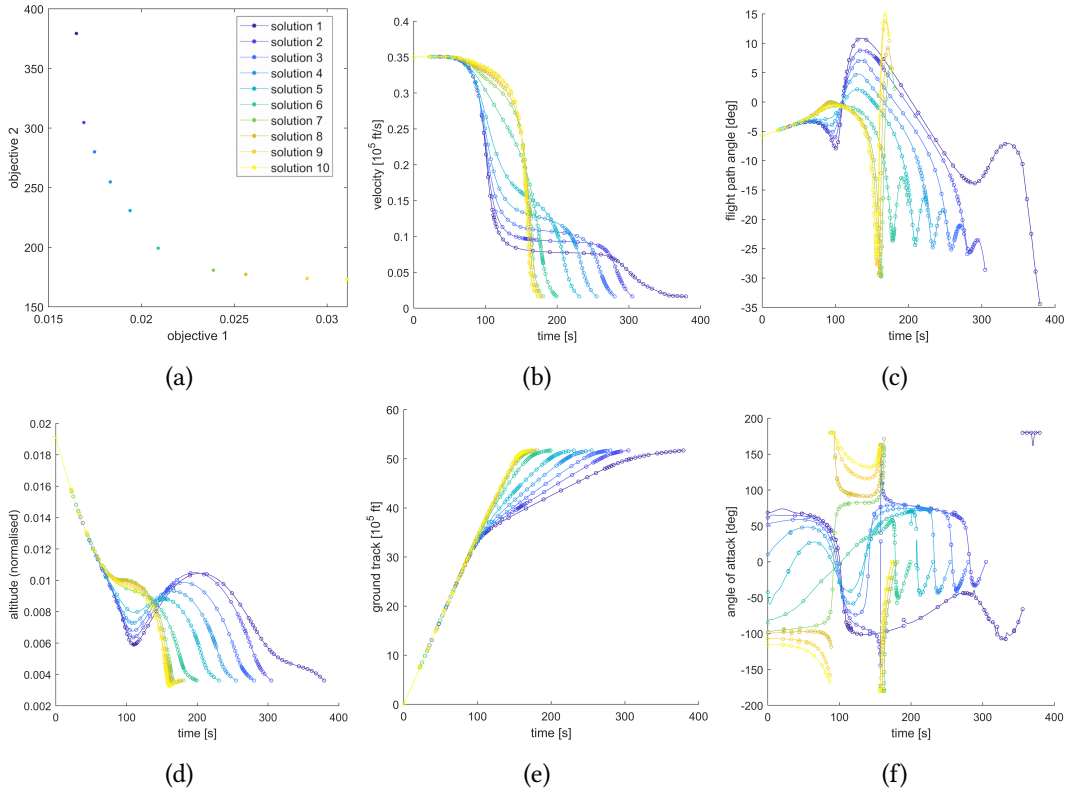
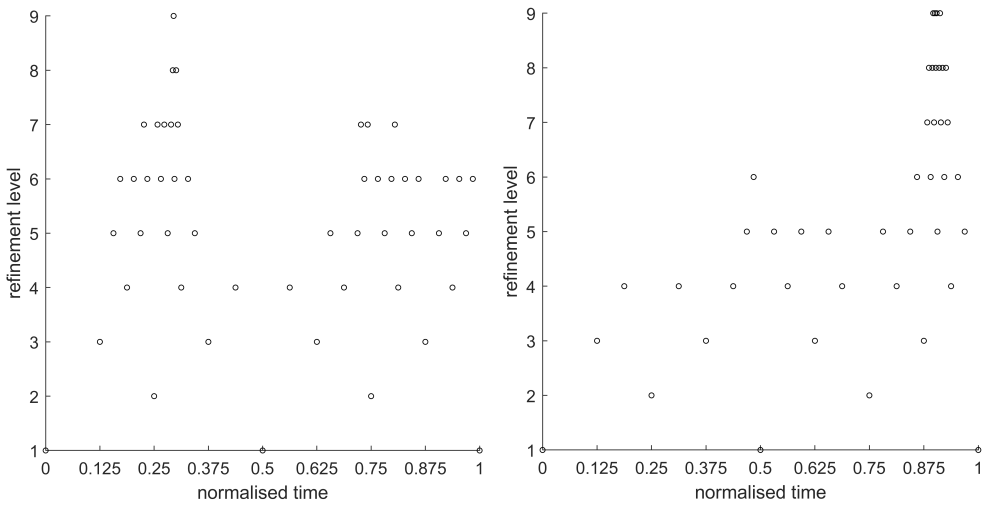


FIGURE 6.20: Multi-objective solutions to Problem 6.7.2 (Apollo Re-entry), including (a) Pareto front and corresponding (b-e) state and (f) control profiles.

Fig. 6.20 shows the time histories of the state and control profiles from an example solution, including the corresponding grid points (represented by circle markers) for each. Additionally, Fig. 6.21 shows the dyadic distribution of grid points for both orthogonal (extremal) solutions with respect to resolution level J across the normalised time domain. The solution representing the minimum value of objective J_1 closely matches the single-objective solution reported by Jain and Tsiotras (2008), including the final distribution of grid points following the author's own mesh refinement procedure.

Table 6.8 presents the algorithm performance comparisons between solving Problem 6.7.2 via the proposed multi-resolution approach and via a globally dense uniform-grid approach, where the latter is performed at successive levels of resolution. Reported are the 10-run average values, with corresponding standard deviations included in brackets. This problem was found to be relatively demanding computationally, with only a small number of uniform-grid levels analysed before the expense of doing so became prohibitive. Nonetheless, an



(a) (b)

FIGURE 6.21: Dyadic distribution of collocation points for the orthogonal solutions to (a) objective 1 and (b) objective 2.

Table 6.8: Mean statistics for Problem 6.7.2 (Apollo Re-entry), recorded at $n_{iter} = 200$ (standard deviation included in brackets).

Resolution		Mesh error		No. func. evals		Run time		Objectives
J_{min}	J_{max}	ϵ_{max}		n_{feval}		[s]		$\mathbf{J}_{best}$
2	2	9.178e-3	(1.2709e-3)	**62065**	**(3327)**	**3593**	**(434)**	[1.694e-2 (3.80e-5), 181.89 (1.76)]
3	3	4.186e-3	(1.226e-3)	108070	(3871)	10333	(2371)	[1.652e-2 (1.332e-5), 176.08 (0.55)]
4	4	3.917e-3	(1.042e-3)	140938	(6301)	40952	(6969)	[1.651e-2 (1.544e-5), 175.39 (0.34)]
5	5	3.131e-3	(1.003e-3)	159279	(18239)	383558	(113465)	[1.651e-2 (1.012e-5), 174.56 (0.73)]
0	16	**2.400e-5**	**(6.105e-5)**	114064	(7173)	215224	(70065)	[**1.650e-2 (6.599e-6)**, **173.21 (0.02)**]

inverse correlation between mesh resolution and relative integration error is observed across the uniform-grid cases. Furthermore, higher mesh densities incur an exponential increase in computational costs given the associated evolution of the NLP problem size. On the other hand, the multi-resolution approach is seen to temper this trade-off, achieving competitive mesh error values for reduced computational effort. In this example, the uniform-grid approach was unable to meet the prescribed mesh integration error tolerance, even at the highest practical resolution level. This is despite the rapid exponential accumulation of computational effort beyond that of the multi-resolution approach.

Fig. 6.22 shows the accumulation of total CPU time for each method relative to the number of completed iterations.While the multi-resolution approach demonstrates efficiency savings in CPU time relative to the hierarchy of uniform-grid examples, a significant accumulation of objective function evaluations is observed during later iterations. This rise can be seen to occur primarily during the local search step, suggesting an increasing difficulty for the gradient-based optimiser to converge on a locally optimal solution following successive grid refinement.

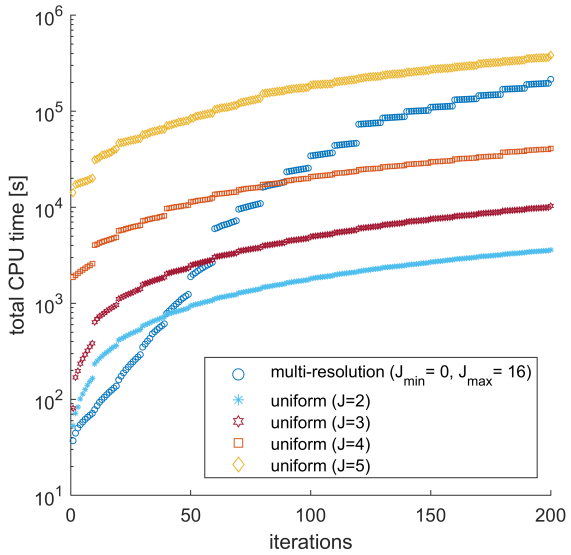


FIGURE 6.22: Total execution time relative to the number of completed iterations.

The usability of the proposed approach is fundamentally defined by the trade-off between achievable solution accuracy and computational expense. Fig. 6.23 shows the values of maximum relative mesh integration error recorded at each iteration for each multi/uniform-resolution process. A clear hierarchy can be seen in the values of mesh integration error achievable via each level of mesh density. Corresponding to the progressive refinement process, the multi-resolution method is seen to move through each level rather than stagnate at any particular limit.

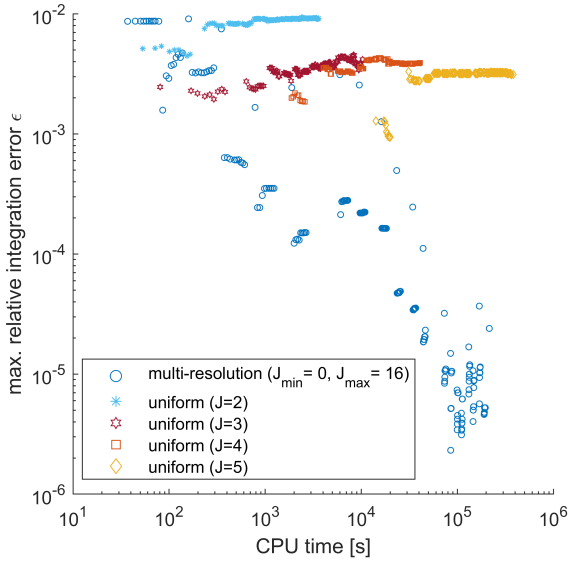


FIGURE 6.23: Relative mesh integration error and population energy metrics relative to the total execution time.

Fig. 6.24 depicts the open-ended evolution of the NLP problem size.

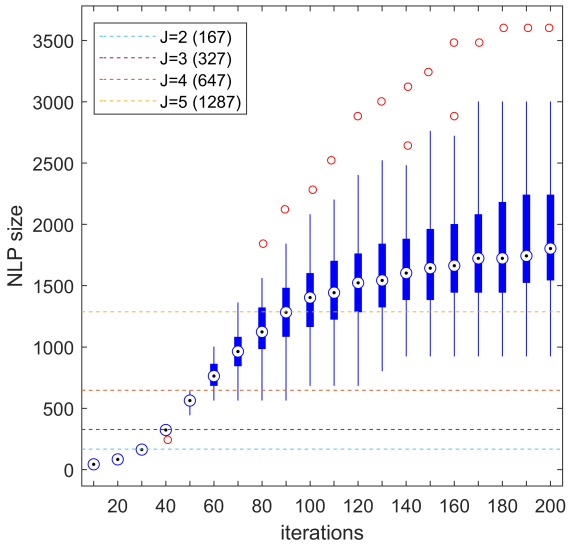


FIGURE 6.24: NLP problem size across the included population relative to the number of completed iterations. For simplicity, only iterations immediately following a refinement step are shown.

6.7.3 Betts Shuttle Re-entry

This problem, adapted from the classic test case by Betts (2010), considers maximising the cross-range distance $\phi(t_f)$ of a space shuttle-like vehicle during atmospheric re-entry whilst minimising the peak convective heat flux q_U experienced during flight. The MOCP may be expressed as:

$$\min_{t_f,\mathbf{u}}[J_1,\ J_2]^T = [-\phi(t_f),\ q_U]^T \tag{6.82}$$

subject to the following dynamics on position vector $\mathbf{r} = [r, \theta, \phi]$ and the velocity vector $\mathbf{v} = [v, \gamma, \psi]$:

$$\dot{r} = \dot{h} = v \sin\gamma \tag{6.83}$$

$$\dot{\theta} = \frac{v \cos\gamma \sin\psi}{r \cos\phi} \tag{6.84}$$

$$\dot{\phi} = \frac{v \cos\gamma \cos\psi}{r} \tag{6.85}$$

$$\dot{v} = -\frac{D}{m} - g \sin\gamma \tag{6.86}$$

$$\dot{\gamma} = \frac{L \cos\beta}{mv} - \left(\frac{g}{v} - \frac{v}{r}\right) \cos\gamma \tag{6.87}$$

$$\dot{\psi} = \frac{L \sin\beta}{mv \cos\gamma} + \frac{v \cos\gamma \sin\psi \tan\phi}{r} \tag{6.88}$$

with altitude h, longitude θ, latitude ϕ, velocity v, flight path angle γ and heading angle ψ. Consistent with the problem definition in Betts and Huffman (1998), the boundary conditions at initial t_0 and final t_f times are specified in Table 6.9.

Table 6.9: Initial and final conditions for Problem 6.7.3 (Shuttle Re-entry).

Variable	Initial ($t = 0$)		Final ($t = t_f$)	
Altitude h [ft]	$h(0) = h_0$	260000	$h(t_f) = h_f$	80000
Longitude θ [deg]	$\theta(0) = \theta_0$	0	$\theta(t_f) = \theta_f$	free
Latitude ϕ [deg]	$\phi(0) = \phi_0$	0	$\phi(t_f) = \phi_f$	free
Velocity v [ft/s]	$v(0) = v_0$	25600	$v(t_f) = v_f$	2500
Flight path angle γ [deg]	$\gamma(0) = \gamma_0$	-1	$\gamma(t_f) = \gamma_f$	-5
Heading angle ψ [deg]	$\psi(0) = \psi_0$	90	$\psi(t_f) = \psi_f$	free

Path constraints are given as:

$$\begin{aligned} -90 \text{ deg} \leq \alpha \leq 90 \text{ deg} \\ -89 \text{ deg} \leq \beta \leq 1 \text{ deg} \\ |\dot{\gamma}(t)| \leq 2 \text{ deg/s} \\ |\dot{\psi}(t)| \leq 2 \text{ deg/s} \\ q \leq q_U \end{aligned} \tag{6.89}$$

where the control variables α and β are the angle of attack and bank angle respectively. The force of gravity is taken as:

$$g = \frac{\mu}{r^2} \tag{6.90}$$

where $r = R_e + h$ and the Earth radius $R_e = 20902900$ ft. The aerodynamic forces acting on the vehicle are given by:

$$L = \frac{1}{2} C_L S \rho v^2 \tag{6.91}$$

$$D = \frac{1}{2} C_D S \rho v^2 \tag{6.92}$$

where the coefficients of lift and drag (C_L and C_D respectively) are defined by the following fitted relationships:

$$C_L = a_0 + a_1 \hat{\alpha} \tag{6.93}$$

$$C_D = b_0 + b_1 \hat{\alpha} + b_1 \hat{\alpha}^2 \tag{6.94}$$

where $\hat{\alpha}$ is the angle of attack in radians and the constant values $a_0 = -0.20704$, $a_1 = 0.029244$, $b_0 = 0.07854$, $b_1 = -0.61592 \times 10^{-2}$ and $b_2 = 0.621408 \times 10^{-3}$. Atmospheric density ρ is defined by:

$$\rho = \rho_0 \exp[-h/h_r] \tag{6.95}$$

where the sea-level density is $\rho_0 = 0.002378$ lb/ft^2 and reference altitude $h_r = 23800$ ft. The convective heat flux at the vehicle nose, q is calculated using:

$$q = q_a q_r \tag{6.96}$$

where

$$q_a = c_0 + c_1\hat{\alpha} + c_2\hat{\alpha}^2 + c_3\hat{\alpha}^3 \tag{6.97}$$

$$q_r = 17700\sqrt{\rho}(0.0001v)^{3.07} \tag{6.98}$$

and is constrained by a maximum peak value $q_U \in [0, 70]$, itself a static parameter. The constant values are $c_0 = 0.0672181$, $c_1 = -0.19213774 \times 10^{-1}$, $c_2 = 0.21286289 \times 10^{-3}$ and $c_3 = -0.10117249 \times 10^{-5}$. The reentry trajectory begins at an altitude where the aerodynamic forces are small, with a vehicle weight of $m = 203000$ lb.

This problem is solved in the same manner as for the previous problems, with largely the same numerical settings. In this case however, 7th order Bernstein polynomials are used to represent the state and control profiles and the function, constraint violation and step tolerances of the inner-level optimisation step were set to 1e-6, 1e-12 and 1e-9 respectively. Lastly, the maximum number of allowable iterations for this case was set to $n_{iter,max} = 120$.

Figure 6.25 shows the spacing of an example set of solutions in objective space, with Figs. 6.26 and 6.27 displaying the corresponding colour-coded time histories of selected states and controls. Additionally, Fig. 6.28 shows the dyadic distribution of grid points for both orthogonal objective solutions. This example features complex state and control profiles, with no apparent discontinuity (though including distinct areas of rapid change)However, the similarity in form between solutions across the population can be clearly seen. Importantly, this demonstrates that important areas are being discovered, preserved and propagated throughout the optimisation process. Objective function values correspond to those reported by Betts and Huffman (1998); Ricciardi and Vasile (2018).

Table 6.10 presents the performance comparisons between the multi-resolution approach and the series of globally dense uniform grids. In a similar manner to the previous cases, the multi-resolution approach achieves a favourable balance between solution accuracy and computational cost relative to the the uniform grid examples. For example, the highest level of dense grid considered here, $J = 5$, was unable to reach the same level of maximum relative mesh integration error as the multi-resolution approach within the set number of iterations, despite the latter terminating with 12.7% of the computational cost. Furthermore, as in the previous examples, there seems to be no significant benefit to the dense grid approach when considering the orthogonal objective function solutions, with a 1.02% improvement and a 17.7% deterioration in J_1 and J_2 respectively relative to the multi-resolution approach.

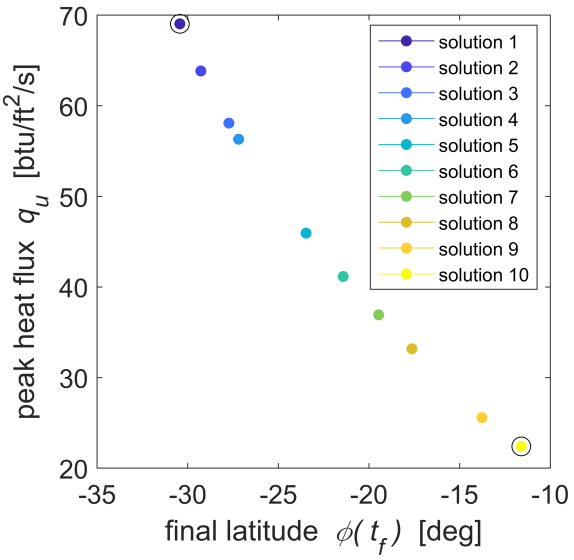


FIGURE 6.25: Multi-objective Pareto front for Problem 6.7.3 (Shuttle Re-entry).

Table 6.10: Mean statistics for Problem 6.7.3 (Shuttle Re-entry), recorded at $n_{iter} = 120$ (standard deviation included in brackets).

Resolution level J_{min}	J_{max}	Mesh error ϵ_{max}		No. func. evals n_{feval}		Run time [s]		Objectives $\mathbf{J}_{best}$
2	2	5.15e-3	(1.71e-3)	**27578**	(**3.75e+3**)	**10330**	(**4.28e+3**)	[-0.506 (5.65e-2), 22.16 (1.32e+0)]
3	3	3.66e-4	(1.78e-5)	39967	(6.70e+3)	43108	(1.98e+4)	[-0.530 (9.12e-3), **21.95** (4.16e-2)]
4	4	1.43e-4	(9.85e-6)	50818	(1.14e+4)	116624	(1.91e+4)	[-0.529 (8.00e-3), 27.95 (1.09e+1)]
5	5	5.26e-5	(**5.33e-6**)	63374	(1.02e+4)	640130	(8.63e+4)	[**-0.534** (**5.10e-4**), 26.39 (7.50e+0)]
0	8	**4.11e-5**	(4.50e-5)	84834	(3.05e+4)	81155	(1.19e+4)	[-0.529 (8.56e-3), 22.42 (**2.36e-2**)]

The mean values and corresponding standard deviations of IGD and AHD for each method are reported in Table 6.11, where the reference Pareto-front is taken directly from Ricciardi and Vasile (2018) and Betts and Huffman (1998). In this case, the multi-resolution approach achieves superior values for mean and variance according to both metrics. This may be due to a number of reasons. At lower mesh resolutions, the optimiser may not be able to capture the finer features of the state/control profiles, thus while easily achieving a well spread population, the approximation accuracy of the Pareto-front and mesh errors remain relatively poor. On the other hand, higher mesh densities may more capably capture detailed behaviour but the resulting size of the NLP problem (e.g. 2057 total variables for the $J = 5$ grid) and the inherent existence of over-refined areas hinders the spreading of the population in objective space, indicated by the lower IGD values relative to AHD. The ability of the multi-resolution approach to efficiently manage the gradual refinement of the population so as to maintain global exploratory capability can also be seen in Fig. 6.29.

Fig. 6.30 shows the accumulation of total CPU time and objective function evaluations respectively relative to the number of completed iterations. Of interest in this case is the more rapid

accumulation of CPU time at earlier stages of the multi-resolution process. This suggests a difficulty for the optimiser to produce feasible solutions at lower levels of refinement. This is further supported by Fig. 6.31, showing that the evolution of NLP problem size progresses in almost a uniform manner until around 50 iterations (or above resolution level $J = 3$). As in the previous example, excessive numbers of objective function evaluations are prevalent during later (> 70) iterations, though Fig. 6.31 at least indicates that true NLP bloat is being tempered given the apparent convergence of average NLP size. This could possibly be attributed to the indirect nature of solution growth mentioned earlier, in that refinement will not occur indefinitely without the mesh error continuously failing to meet the required tolerance.

Table 6.11: Mean values (standard deviation in brackets) for IGD and AHD for all uniform and multi-resolution solutions to Problem 6.7.3 (Shuttle Re-entry).

Resolution level		Inverted Generational Distance (IGD)		Averaged Hausdorff Distance (AHD)	
J_{min}	J_{max}				
2	2	1.44	(1.27)	4.76	(3.87)
3	3	0.93	(0.29)	3.27	(0.81)
4	4	1.66	(1.00)	5.26	(3.17)
5	5	1.02	(0.54)	3.42	(1.57)
0	8	**0.59**	(**0.05**)	**2.51**	(**0.27**)

6.7.4 Discussion

The examples presented in Section 6.7 illustrate the efficiency of the proposed multi-resolution approach compared to a more traditional uniform-grid approach for MOCPs.

Test case 6.7.1 (Drug Displacement) includes two discontinuities at non-constant relative locations within the normalised time domain. Test cases 6.7.2 (Apollo Re-entry) and 6.7.3 (Shuttle Re-entry), include multiple control variables and complex state/control profiles with areas of high non-linearity, though with no clear discontinuities. In these cases, the multi-resolution approach is seen to preserve the form of the optimal trajectory throughout the population, with individualistic refinement applied where appropriate. This demonstrates cost savings over a uniform-grid approach for the same level of accuracy.

Common to each case is the reduction in computational cost required to reach specified levels of relative local error compared to selected uniform grid resolutions (themselves chosen in order to remain within practical computational bounds). However, a greater reliance on objective function evaluations is observed, particularly in later iterations when the multi-resolution populations begins to converge. While potentially merely a symptom of the periodic local refinement procedure requiring more intensive computation given the increased NLP problem size, this may also correspond to the inherent probability of retaining/propagating overrefined segments, thus introducing unnecessary computational burden through excessive NLP 'bloat'.

Such a behaviour may form as a result of particular areas of a solution undergoing refinement during earlier stages of the process, later becoming unnecessary as the population develops. An additional factor could be the propagation of overrefined segments purely by association with a favourable sub-sequence (i.e., one that improves the objective function values of a solution when inserted via crossover, despite containing segments of unnecessarily high resolution). This issue is in large introduced by considering that the removal of overrefined segments is only at this stage performed in an indirect manner. More specifically, an overrefined segment can only be removed if it fails to propagate across successive generations, either via substitution during social action or by the parent(s) being removed from the current archive. An issue here is that such occurrences are only possible if there exists a solution with superior fitness despite the absence of that particular segment. If an overrefined solution exists but is sufficiently dominant within the population, it shall remain undisturbed. Clearly then an additional step introduced to validate each individual, identifying and removing overrefined segments, would be a valid subject of future work. Of note is that this behaviour seems less pronounced for the Shuttle Re-entry test case, possibly suggesting an overly strict value of relative local error tolerance ϵ_{tol} for the Moon Lander example in Section 6.6.

Adapting the degree of the polynomial basis functions, thus enabling full *hp*-adaptivity, was initially determined outwith the scope of this work, though remains an interesting subject for future study. For example, the work by Patterson et al. (2015), in reference to the problem reproduced in Section 6.7.3, suggests a positive effect in solution accuracy and cost by increasing the polynomial degree (within user-defined applicability bounds) as opposed to the number of mesh elements.

6.8 Chapter Summary

This chapter has presented an adaptive mesh refinement approach for solving direct MOCPs within an evolutionary architecture. Each individual solution is discretised upon an independent non-uniform dyadic grid using *Direct Finite Elements in Time* (DFET). New grid points may be added at each iteration according to a criteria based on simple local errors evaluated at element boundaries. The transcribed problem is solved using a new version of the *Multi-Agent Collaborative Search* (MACS) algorithm, where the original formulation has been modified to incorporate adaptive, multi-resolution solution vectors. This includes both the proposed dyadic mesh-refinement procedure and new social operators to enable collaborative action between solutions of inconsistent grid resolution. Concerning the latter, common segments between unique individuals are identified using the nodal intersections of their associated discretisation grids. Next, a biased variable-length crossover technique that prioritises the exchange of higher resolution segments is applied to create entirely new candidate solutions.

The proposed approaches are applied to several demonstrative test cases, where in each the associated *Multi-Objective Optimal Control Problem* (MOCP) is treated as an open-ended evolutionary design problem. As the primary motivation of this work was a desire to increase the achievable fidelity of MOCP solutions as well as the numerical efficiency, the computational cost, achievable solution accuracy and population spread/diversity were measured for each case and compared against a traditional uniform-grid approach. An order of magnitude reduction in computational cost required to reach a desired level of mesh integration error was observed for both cases. Furthermore, the progressive nature of population refinement was shown to facilitate a more complete and diverse representation of the Pareto-front.

A primary limitation of the proposed approach was found to be related to unregulated and/or excessive NLP growth, or 'bloat'. This issue is in large due to the fact that the removal of overrefined segments is only at this stage performed in an indirect manner, without any specific/targeted detection and removal of overrefined solution segments. Whilst a common issue amongst other variable-length algorithms, resulting effects include a skewed reliance on increased objective function evaluations at later generations of the optimisation process, naturally becoming much more expensive given the increasing NLP size as the algorithm approaches convergence. Accordingly, an additional validation step applied to each individual agent to identify and remove overrefined segments would be a valid subject of future work.

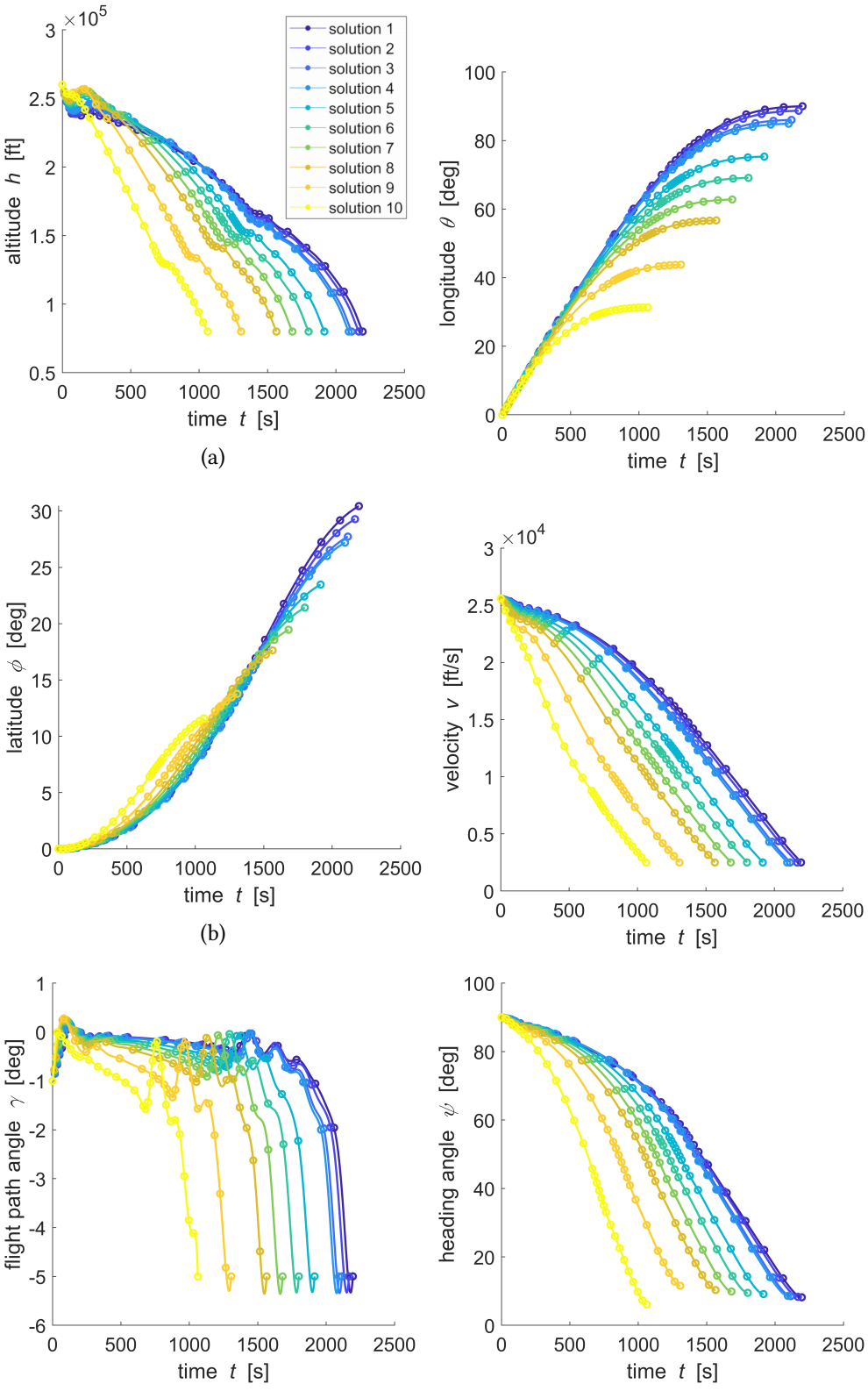
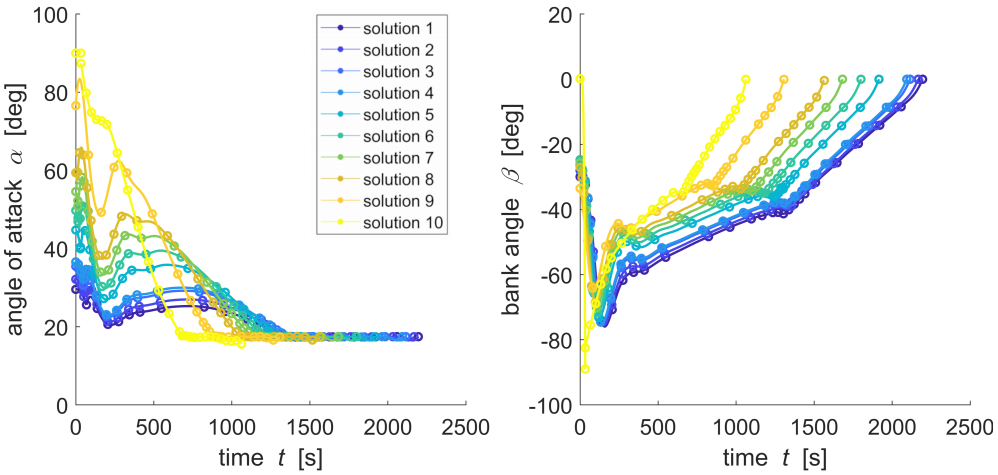
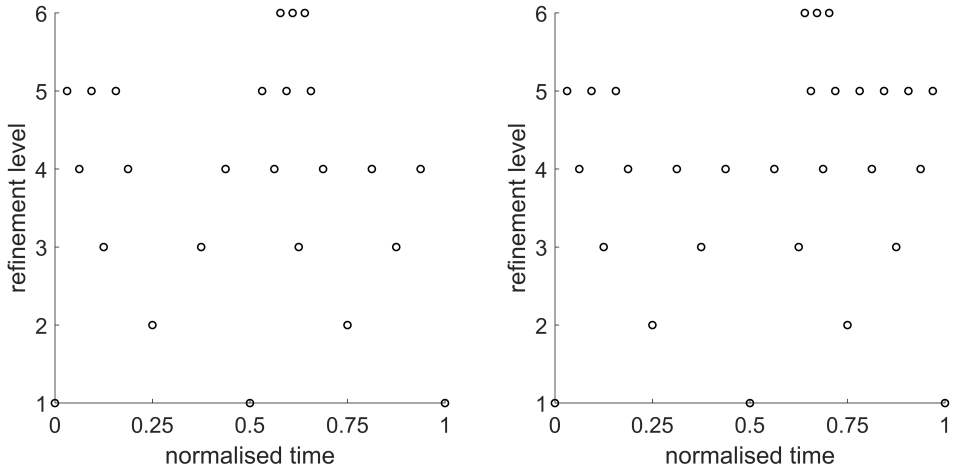


FIGURE 6.26: Multi-objective state profiles for Problem 6.7.3 (Shuttle Re-entry).



(a) (b)

FIGURE 6.27: Multi-objective control profiles for Problem 6.7.3 (Shuttle Re-entry).



(a) (b)

FIGURE 6.28: Dyadic distribution of collocation points for the orthogonal solutions to (a) objective 1 and (b) objective 2 for Problem 6.7.3 (Shuttle Re-entry).

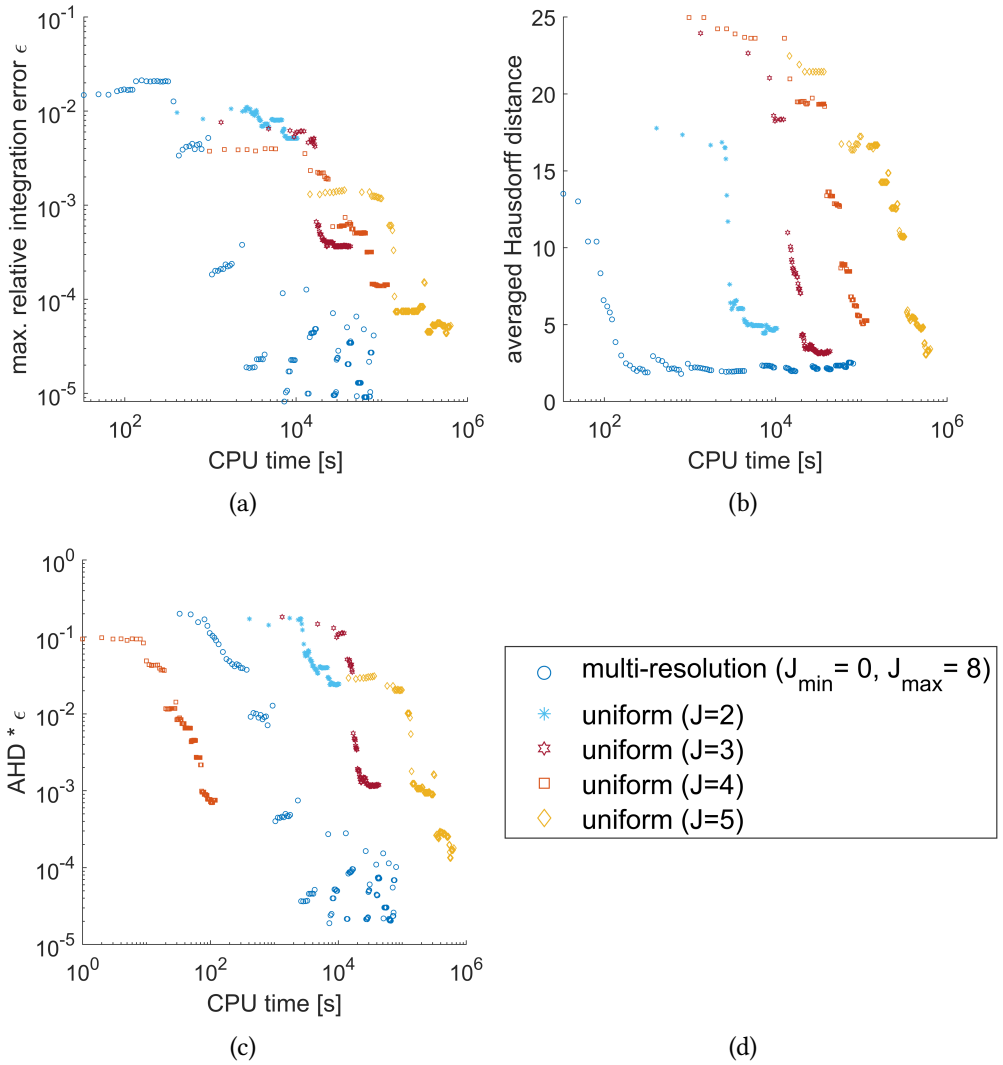
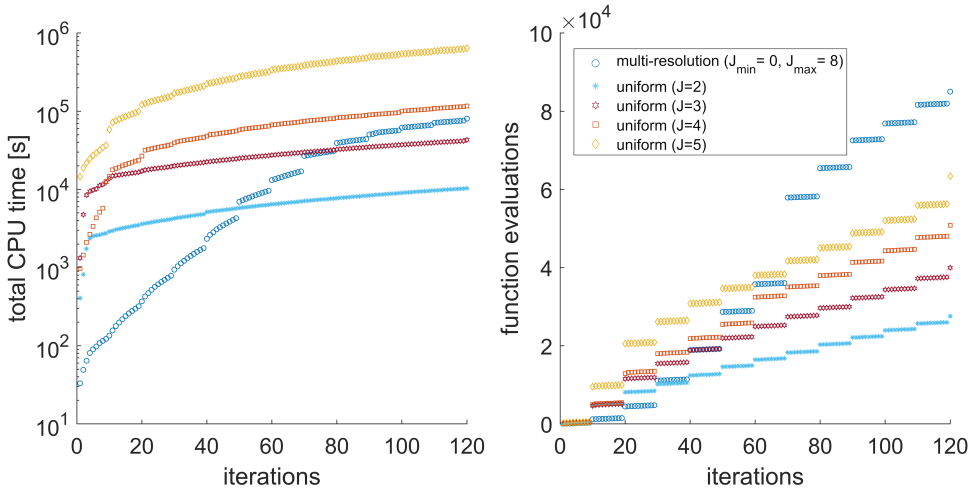


FIGURE 6.29: Problem 6.7.3 (Shuttle Re-entry) mesh error and AHD relative to the total CPU time.



(a) (b)

FIGURE 6.30: (a) Total execution time and (b) number of function evaluations of Problem 6.7.3 (Shuttle Re-entry) relative to the number of completed iterations.

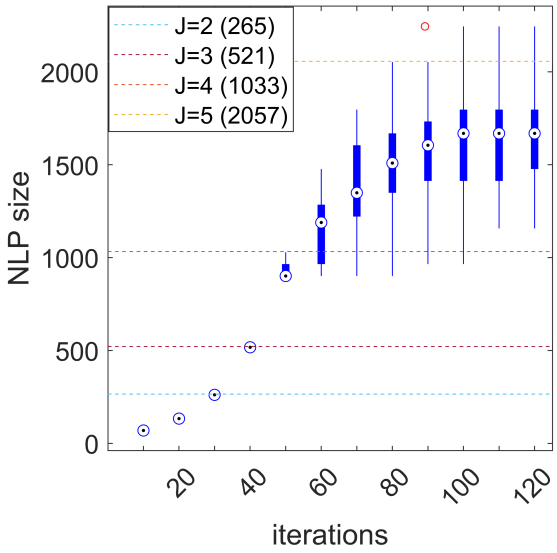


FIGURE 6.31: NLP problem size across the included population relative to the number of completed iterations of Problem 6.7.3 (Shuttle Re-entry).

6.9 Bibliography

O. Abdelkhalik. Hidden genes genetic optimization for variable-size design space problems. *Journal of Optimization Theory and Applications*, 156(2):450–468, 2013.

J. T. Betts. *Practical methods for optimal control and estimation using nonlinear programming*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2010. ISBN 978-0-89871-688-7. doi: https://doi.org/10.1137/1.9780898718577.

J. T. Betts and W. P. Huffman. Mesh refinement in direct transcription methods for optimal control. *Optimal Control Applications and Methods*, 19(1):1–21, 1998.

M. Borri, C. Bottasso, and P. Mantegazza. Basic features of the time finite element approach for dynamics. *Meccanica*, 27(2):119–130, 1992. ISSN 1572-9648.

C. L. Bottasso. A new look at finite elements in time: A variational interpretation of Runge-Kutta methods. *Applied Numerical Mathematics*, 25(4):355–368, 1997. ISSN 01689274. doi: 10.1016/S0168-9274(97)00072-X.

M. Brameier. On Linear Genetic Programming. *Thesis*, 2004.

D. S. Burke, K. A. De Jong, J. J. Grefenstette, C. L. Ramsey, and A. S. Wu. Putting more genetics into genetic algorithms. *Evolutionary computation*, 6(4):387–410, 1998. ISSN 10636560. doi: 10.1162/evco.1998.6.4.387.

S. J. Citron. *Elements of Optimal Control*. Holt, Rinehart and Winston, 1969. ISBN 9780030733703.

D. Goldberg. Messy Genetic Algorithms : Motivation, Analysis, and First Results. *Complex Systems*, 3(5):493–530, 1989. ISSN 0891-2513. URL http://www.complex-systems.com/abstracts/v03_i05_a05.html.

A. Harten. Multiresolution representation of data: A general framework. *SIAM Journal on Numerical Analysis*, 33(3):1205–1256, 1996. URL https://www.jstor.org/stable/2158503.

I. Harvey. The SAGA Cross: The Mechanics of Recombination for Species with Variable-Length Genotypes. *Parallel Problem Solving from Nature*, 2:269–278, 1992.

Y. Hochberg and A. C. Tamhane. *Multiple comparison procedures*. John Wiley & Sons, Inc., 1987.

D. H. Hodges and R. R. Bless. Weak hamiltonian finite element method for optimal control problems. *Journal of Guidance, Control, and Dynamics*, 14(1):148–156, 1991. ISSN 07315090. doi: 10.2514/3.20616.

J. H. Holland. Genetic Algorithms. *Scientific American*, 267(1):66–73, jun 1992. ISSN 00368733, 19467087. URL http://www.jstor.org/stable/24939139.

B. Hutt and K. Warwick. Synapsing variable-length crossover: Meaningful crossover for variable-length genomes. *IEEE Transactions on Evolutionary Computation*, 11(1):118–131, 2007. ISSN 1089778X. doi: 10.1109/TEVC.2006.878096.

S. Jain and P. Tsiotras. Trajectory optimization using multiresolution techniques. *Journal of Guidance, Control, and Dynamics*, 31(5):1424–1436, 2008. ISSN 15333884. doi: 10.2514/1.32220.

J. R. Koza. The genetic programming paradigm: Genetically breeding populations of computer programs to solve problems. *Dynamic, Genetic, and Chaotic Programming*, (June):203–321, 1992. doi: doi:10.1109/TAI.1990.130444.

W. B. Langdon and R. Poli. Fitness causes bloat: Mutation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 1391(July 2006):37–48, 1998. ISSN 16113349. doi: 10.1007/BFb0055926.

N. K. Madavan. Multiobjective Optimization Using a Pareto Differential Evolution Approach. *Proceedings of the 2002 Congress on Evolutionary Computation*, 2, 2002. doi: https://doi.org/10.1109/CEC.2002.1004404.

S. G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *Fundamental Papers in Wavelet Theory*, II(7), 1989. doi: https://doi.org/10.1109/34.192463.

H. Maurer and M. Wiegand. Numerical solution of a drug displacement problem with bounded state variables. *Optimal Control Applications and Methods*, 13(1):43–55, 1992. ISSN 10991514. doi: 10.1002/oca.4660130104.

J. Meditch. On the problem of optimal thrust programming for a soft lunar landing. 1964. doi: https://doi.org/10.1109/TAC.1964.1105758.

A. Merlevede, H. Åhl, and C. Troein. Homology and linkage in crossover for linear genomes of variable length. *PLoS ONE*, 14(1):1–16, 2019. ISSN 19326203. doi: 10.1371/journal.pone.0209712.

M. Munetomo, Y. Takai, and Y. Sato. Migration scheme for the genetic adaptive routing algorithm. *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, 3:2774–2779, 1998. ISSN 08843627. doi: 10.1109/icsmc.1998.725081.

B. Parsonage and C. Maddock. Biased Dyadic Crossover for Variable-Length Multi-Objective Optimal Control Problems. *2024 IEEE Congress on Evolutionary Computation, CEC 2024 - Proceedings*, pages 1–9, 2024a. doi: 10.1109/CEC60901.2024.10611888.

B. Parsonage and C. Maddock. Solving multi-objective optimal control problems using a multiresolution approach. *Journal of Guidance, Control, and Dynamics*, pages 1–14, 2024b. ISSN 1533-3884. doi: 10.2514/1.g008223.

A. Pascoletti and P. Serafini. Scalarizing vector optimization problems. *Journal of Optimization Theory and Applications*, 42(4):499–524, 1984. ISSN 00223239. doi: 10.1007/BF00934564.

M. A. Patterson, W. W. Hager, and A. V. Rao. A ph mesh refinement method for optimal control. *Optimal Control Applications and Methods*, 36(4):398–421, 2015. ISSN 10991514. doi: 10.1002/oca.2114.

W. R. Pearson. An introduction to sequence similarity ("homology") searching. *Current Protocols in Bioinformatics*, (SUPPL.42):1–8, 2013. ISSN 19343396. doi: 10.1002/0471250953.bi0301s42.

H. J. Pesch. Real-time computation of feedback controls for constrained optimal control problems. part 2: A correction method based on multiple shooting. *Optimal Control Applications and Methods*, 10(2):147–171, 1989. ISSN 10991514. doi: 10.1002/oca.4660100206.

Pomax. A Primer on Bezier Curves, 2023. URL https://pomax.github.io/bezierinfo/.

Z. Qiongbing and D. Lixin. A new crossover mechanism for genetic algorithms with variable-length chromosomes for path optimization problems. *Expert Systems with Applications*, 60: 183–189, 2016. ISSN 09574174. doi: 10.1016/j.eswa.2016.04.005.

G. R. Reeck, C. de Haen, D. C. Teller, R. F. Doollttle, W. M. Fitch, R. E. Dickerson, P. Chambon, A. D. McLachlan, E. Margoliash, T. H. Jukes, and E. Zuckerkandl. "Homology" in Proteins and Nucleic Acids: A Terminology Muddle and a Way out of It. *Accountability in Research*, 1987.

L. A. Ricciardi. *Multi-Objective Hybrid Optimal Control with Application to Space Systems*. Doctoral dissertation, University of Strathclyde, 2019.

L. A. Ricciardi and M. Vasile. MODHOC - Multi Objective Direct Hybrid Optimal Control. 2018. URL https://strathprints.strath.ac.uk/67085/.

L. A. Ricciardi and M. Vasile. Direct transcription of optimal control problems with finite elements on bernstein basis. *Journal of Guidance, Control, and Dynamics*, 42(2):229–243, 2019a. ISSN 07315090. doi: 10.2514/1.G003753.

L. A. Ricciardi and M. Vasile. Improved Archiving and Search Strategies for Multi Agent Collaborative Search. *Computational Methods in Applied Sciences*, 48(January):435–455, 2019b. ISSN 18713033. doi: 10.1007/978-3-319-89988-6_26.

L. A. Ricciardi and M. Vasile. A relaxation approach for hybrid multi-objective optimal control: Application to multiple debris removal missions. *Advances in the Astronautical Sciences*, 168: 681–700, 2019c. ISSN 00653438.

L. A. Ricciardi, M. Vasile, F. Toso, and C. A. Maddock. Multi-Objective Optimal Control of the Ascent Trajectories of Launch Vehicles. pages 1–14, 2016. doi: 10.2514/6.2016-5669.

L. A. Ricciardi, C. A. Maddock, and M. Vasile. Direct Solution of Multi-Objective Optimal Control Problems Applied to Spaceplane Mission Design. *Journal of Guidance, Control, and Dynamics*, 42(1):30–46, 2018a. ISSN 0731-5090. doi: 10.2514/1.g003839.

L. A. Ricciardi, C. A. Maddock, and M. Vasile. Multi-objective optimal control of re-entry and abort scenarios. In *Space Flight Mechanics Meeting, 2018*, number 210009, 2018b. ISBN 9781624105333. doi: 10.2514/6.2018-0218.

O. Schütze, X. Esquivel, A. Lara, and C. A. Coello. Using the averaged hausdorff distance as a performance measure in evolutionary multiobjective optimization. *IEEE Transactions on Evolutionary Computation*, 16(4):504–522, 2012. ISSN 1089778X. doi: 10.1109/TEVC.2011.2161872.

S. F. Smith. *A learning system based on genetic adaptive algorithms.* PhD thesis, University of Pittsburgh, 1980.

M. Vasile. Finite elements in time: A direct transcription method for optimal control problems. *AIAA/AAS Astrodynamics Specialist Conference 2010*, (August 2010), 2010a. doi: 10.2514/6.2010-8275.

M. Vasile. Finite elements in time: A direct transcription method for optimal control problems. *AIAA/AAS Astrodynamics Specialist Conference 2010*, pages 1–25, 2010b. doi: 10.2514/6.2010-8275.

M. Vasile and F. Bernelli-Zazzera. Trajectory design combining low-thrust and gravity assist manoeuvres. In *Operations Research in Space and Air*, pages 203–229. Springer, 2003.

M. Vasile and A. Finzi. Direct lunar descent optimisation by finite elements in time approach. *International Journal of Mechanics and Control*, 1(1):1–15, 2000. ISSN 1590-8844.

M. Vasile and F. Zuiani. Multi-agent collaborative search: An agent-based memetic multi-objective optimization algorithm applied to space trajectory design. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, 225(11):1211–1227, 2011. ISSN 20413025. doi: 10.1177/0954410011410274.

R. S. Zebulum, M. A. Pacheco, and M. Vellasco. Variable length representation in evolutionary electronics. *Evolutionary computation*, 8(1):93–120, 2000a. ISSN 10636560. doi: 10.1162/106365600568112.

R. S. Zebulum, M. Vellasco, and M. A. Pacheco. Variable length representation in evolutionary electronics. *Evolutionary Computation*, 8(1):93–120, 2000b.

J. Zhao and T. Shang. Dynamic optimization using local collocation methods and improved multiresolution technique. *Applied Sciences (Switzerland)*, 8(9), 2018. ISSN 20763417. doi: 10.3390/app8091680.

Engineering Test Case: Interplanetary Transportation 7

7.1 Background . . . 316
7.1.1 Propulsion . . . 316
7.1.2 Mission Architectures . . . 317
7.2 System Models . . . 318
7.2.1 Vehicle . . . 318
7.2.2 Dynamics . . . 324
7.3 Approach . . . 325
7.3.1 Optimisation Algorithm . . . 326
7.4 Results . . . 327
7.4.1 Earth-Mars Transfer . . . 327
7.4.2 Earth-Mars-Earth Transfer . . . 329
7.5 Conclusion . . . 332
7.6 Bibliography . . . 334

An established limiting factor for crewed interplanetary exploration missions to Mars is the health risk posed by increased exposure to solar and cosmic radiation (Bruno and Dujarric, 2013). Given the directly proportional relationship between total dose and radiation flux/time of exposure, logical mitigation strategies include either reducing the radiation flux via some form of protective shielding, or decreasing the interplanetary transfer time. Research into active/passive shielding is ongoing, but it is currently accepted that with the significant volume of shielding required for a Hohmann-type transfer to Mars, the costs to deliver this additional payload to LEO would be largely prohibitive to the overall mission. As such, it is prudent to examine the propulsion options available to reduce the time of flight. While conventional LO2/LH2 or LO2/LCH4 chemical propulsion systems may still remain practical

for Earth orbit and Lunar missions, their characteristic high thrust/low specific impulse results in largely infeasible mass fractions when applied to interplanetary trajectories. Given the resultant need for higher power density and specific impulse I_{sp}, a viable alternative is the use of nuclear reactions.

The following chapter examines the preliminary design trade-offs, given near-term future technologies, of a high powered transportation system to Mars. Mission performance and vehicle configuration are analysed, using numerical models for structural mass, radiation dosage, propulsion system and long-duration habitation module (including consumable requirements). The presented analysis focuses on a nominal crewed mission, the more limiting option versus cargo-only. The requirements and assumptions for the system are as follows:

- The vehicle must be capable of an Earth/Mars transfer trajectory of less than 90 days.
- The system must be capable of carrying, to the surface of Mars, at least 50 tons of cargo and/or a minimum of 3 passengers.
- Launch and landing segments of the mission are not considered.
- In-orbit manufacturing and re-fuelling facilities are assumed as operational around both Earth and Mars.

Vehicle design parameters include engine sizing and gross/dry vehicle mass. For a semi-cycler based mission architecture (Landau and Longuski, 2004), single and return legs are analysed independently and as part of a complete mission scenario using continuous control to ascertain the optimal engine switching structure. The stay time on Mars was included as a variable parameter to assess the impact on the optimal synchronisation of inbound and outbound trajectories. In order to generalise the analysis and avoid dependence on any particular set of launch dates/transfer windows, Earth and Mars are assumed to follow circular Sun-centred orbits. The optimal phasing between Earth and Mars and the overall mission architecture may therefore be more independently examined. An additional design parameter representing the phasing between Earth and Mars at launch is therefore additionally included.

The problem is formulated as a *Multi-Objective Optimal Control Problem* (MOCP) and transcribed into a finite-dimensional NLP problem using a direct approach. Initial investigations identified this case as being numerically difficult and computationally very expensive, particularly when using a fixed resolution discretisation mesh. This is in part due to the number of optimisation variables (in the order of 10^3), the complex state profile and the existence of numerous discontinuities in the optimal control profile (representing *bang-bang* control of the engine thrust). It was found that the level of discretisation mesh resolution required to adequately represent the problem trade-off surface incurred a highly prohibitive computational

cost. Reducing the mesh resolution to practical levels rendered the solver unable to represent the optimal solutions to an acceptable degree. As such, this problem was especially suited to test the practical applicability of the multi-resolution approach presented in Chapter 6 of this thesis.

Parts of the work presented in the following chapter can be found in the following publications:

- C. Maddock, B. Parsonage, L. Ricciardi, M. Vasile, and O. Cohen. Design optimisation and analysis of very high power transportation system to Mars. *Proceedings of the International Astronautical Congress, IAC*, A5(October):25–29, 2021. ISSN 00741795
- B. Parsonage and C. Maddock. Solving multi-objective optimal control problems using a multiresolution approach. *Journal of Guidance, Control, and Dynamics*, pages 1–14, 2024b. ISSN 1533-3884. doi: 10.2514/1.g008223
- B. Parsonage and C. Maddock. Multi-Objective Design Optimisation and Analysis of a Crewed Earth-Mars Transportation System using Nuclear Thermal Propulsion. *Proceedings of the International Astronautical Congress, IAC*, (October), 2024c

7.1 Background

7.1.1 Propulsion

Conventional LO2/LH2 or LO2/LCH4 chemical propulsion systems typically exhibit a characteristic high thrust/low specific impulse operational performance. Whilst currently remaining practical for Earth orbit and Lunar missions, this behaviour offers largely infeasible mass fractions when applied to interplanetary trajectories. Given the resultant need for higher power density and specific impulse I_{sp}, a viable alternative is the use of nuclear reactions.

As early as the 1940s, nuclear derived technology was recognised as a fundamental enabler for future space travel. Landmark studies, such as by Shepherd and Cleaver (1948) concerning *Nuclear Thermal Rockets* (NTRs) and *Nuclear Electric Propulsion* (NEP) for interplanetary trajectories, helped prompt initial practical investigations into nuclear propulsion systems. Subsequent investigations included the Soviet NTR engine development program (Zakirov and Pavshook, 2007), the Atomic Energy Commission 'Project Rover' (Koenig, 1986) and the NASA *Nuclear Engine for Rocket Vehicle Application* (NERVA) program (Finseth and NASA, 1991), the latter incorporating a promising series of successful terrestrial test firings, prompting inclusion into several proposals for crewed missions to Mars (Braun, 1969; Willis, 1971). Since then, advances in materials, magnetic fields and plasma dynamics have supported the development of numerous appealing concepts (National Academies of Sciences and Medicine, 2021). Currently, many of the major space agencies and governments have programmes examining nuclear

propulsion for space exploration, in particular targeting Mars, such as NASA and DARPA, ESA, UKSA, Roscomos, China, and India.

Nuclear Propulsion (NP) systems proposed for space travel primarily include fission-based *Nuclear Thermal Propulsion* (NTP) and NEP, in which an electric propulsion system is powered by passing the working fluid (heated via nuclear reaction) through a generator. A particular advantage of NEP is the independence from solar power, which decreases proportionally to r^{-2} (a recognised hindrance to solar electric propulsion methods). NEP is thus attractive for deep space applications, despite a cost in efficiency due to thermal-electric power conversion. Multi-modal options exist, where a portion of propellant is passed through a power generation subsystem, thus relegating the need for inefficient solar power systems. A notable example is the bi-modal *Nuclear Thermal Electric Rocket* (NTER) under development at ESA, theoretically offering an increased specific impulse over solid fuel NTR by adding a thermal induction heater downstream of the reactor core (Dujarric, 2001; Dujarric et al., 2013).

7.1.2 Mission Architectures

The chosen propulsion system largely influences the range of possible mission architectures available for a given mission. Considering crewed expeditions to Mars, transfer trajectories between Earth and Mars have been well studied in the literature (Willis, 1971; Landau and Longuski, 2004; Wooster et al., 2007). Coherent mission architectures are defined by the synchronisation of inbound and outbound trajectories, most commonly categorised as either *conjunction* class (long stay) or *opposition* class (short stay). Despite short total mission duration, opposition class transfers are typically not considered feasible for incorporation into a crewed missions to Mars. This is due to the significantly higher proportion of time spent in transfer relative to time spent at Mars and the projected higher radiation dosage in accordance with both the extended flight time and the close passage (< 0.7 AU) to the sun. Various conjunction class mission architectures for Earth/Mars round trips with high-thrust propulsion have been examined (Landau and Longuski, 2004, 2006; Zubrin et al., 1991). These include *direct*, *semi-direct* and *stop-over* missions. Each of these may be modified to include a free-return abort trajectory option, although the resulting time of flight in this case would be significantly increased. Additional architectures utilising multiple reusable vehicles in sustainable transportation paradigms are Earth-to-Mars and Mars-to-Earth *semi-cyclers* and *cyclers*.

Some modern architectures have been proposed to accommodate more novel propulsion methods. For example, Loeb et al. (2015) propose avoiding the complex mission architecture and high launch mass of the NASA *Design Reference Architecture* (DRA) (Drake, 2009) by following a continuous low-thrust trajectory using 64 RFIT-45 ion thrusters. The effects of varying the mass of the spacecraft in LEO prior to mission start, the mass of the crew

transportation vehicle, the mass of the Mars lander as well as the *High Earth Orbit* (HEO) altitude, on the overall mission profile/feasibility are examined. Loeb et al. (2015) found that higher masses (with thrust held constant) yield elongated transfer times, while a higher HEO instead shortens them.

The advantages of low-thrust electric propulsion have since prompted investigations into using such systems as a means to augment high-thrust chemical/nuclear propulsion, enabling a more energy-efficient operation (Chai et al., 2021; Owens et al., 2021; Oleson et al., 2020). Oleson et al. (2020) examine the case where a small NEP system is used in conjunction with high-thrust LO2/LCH4 rocket engine to enable an opposition type mission architecture. In this manner, the required mass of radiator panels (acknowledged as a significant limiting factor for interplanetary NEP) can be reduced. While the authors find the incorporation of NEP gives a 5-10x reduction in δV requirements for capture/escape, the time spent in space during transfer is still very large relative to the time spent at Mars, with an imposed upper limit of 700 days.

7.2 System Models

A study of various spacecraft configurations and designs previously proposed in the literature was undertaken. Key drivers were identified, including the type and design of the nuclear propulsion system, radiation mitigation measures necessary for a crewed mission and mass modelling of the crew habitat module. Numerical models for the vehicle were developed for integration into a *Multi-Objective Optimisation* (MOO) problem as part of the mission design trade-off, with a particular focus on component mass scaling laws.

7.2.1 Vehicle

Crewed missions are driven by the requirement to minimise radiation exposure, thus also minimising flight time (Durante, 2014). As reduced transfers translate, at present, into a need for nuclear propulsion, this then strongly drives the design of the spacecraft as a whole. For example, the need for shielding for both the crew and spacecraft systems from both environmental (*Galactic Cosmic Radiation* (GCR) and *Solar Particle Events* (SPEs)) and artificial (NP) sources (which includes maximising reactor displacement) is dependant on the size and operational cycle of the nuclear reactor and the solar environment. The volume of the habitation module, and consequentially the internal component mass breakdown (including consumable requirements/rates), is directly driven by the number of crew, the transfer duration and overall mission architecture (e.g., the transportation habitat may double as part of the surface habitation module(s) on Mars). Finally, the system power requirements (which, assuming no bi-modal functionality, translates into the required mass/configuration of *Photovoltaic Arrays* (PVAs)) are defined by the size of the habitation module and the overall mission duration, given the exponentially decreasing availability of solar power with radial distance from the Sun.

The nominal spacecraft configuration used as a baseline for this study is shown in Figure 7.1. Included is a truss connecting the NTP system (engine, reactor and shield), with the rest of the spacecraft, consequently housed within the radiation shadow cone.

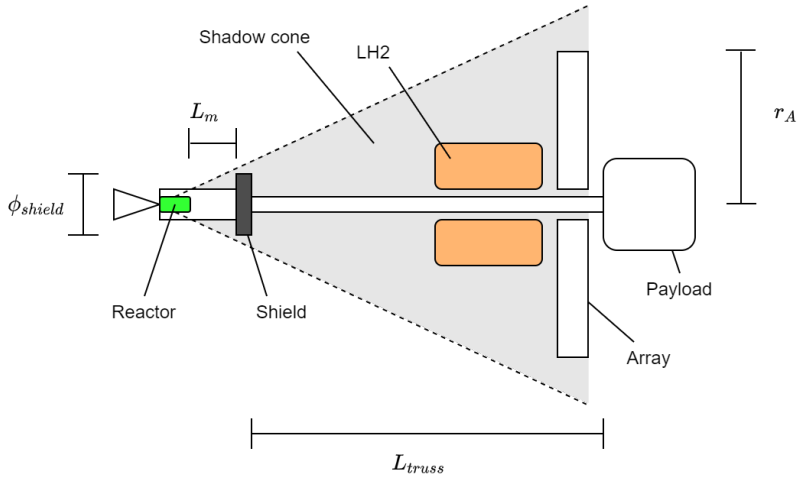


FIGURE 7.1: Nominal spacecraft configuration.

7.2.1.1 Propulsion

Propulsion system sizing relationships were extracted from published data concerning (relatively) high-TRL concepts. Those selected, primarily solid-core fission reactors (see Table 7.1) have either been terrestrially demonstrated as completed test engines (e.g., Pewee 1, NERVA NRX and XE Prime, RD-0410 and NTE 68 kN) or are conceptually based upon successful tests of constituent subsystems/components (e.g., NERVA-1, and AR-NTP). Continuous functions of engine thrust, mass, propellant mass flow rate $\dot{m}$ and specific impulse I_{sp} are defined relative to the thermal power of the reactor.

Engine	Power (MW)	Thrust (kN)	I_{sp} (s)	Mass (kg)	$\dot{m}_{prop}$ (kg/s)
NERVA NRX	1072	237.63	803	8080 (reactor)	32.7
Pewee-1	507	111.2	901	3300	12.6
RD-0410	196	35.28	910	2000	4.0
NERVA XE Prime	1140	246.66	841	7700	35.8
NERVA 1	1570	333.6	825	12577	41.9
NRE 68kN	340	68	900	2890	7.1
AR-NTP	542.4	111.6	875	3793	12.7

Table 7.1: Solid-core engine specifications sourced from: (Finseth and NASA, 1991; Walton, 1991; Koenig, 1986; Drake, 2009; Koroteev et al., 2018; Zakirov and Pavshook, 2007; Frischauf and Hamilton, 2004; Joyner et al., 2020).

In this study, the conceptual NERVA-1 engine is considered the upper limit to propulsion system size. Though never constructed, the NERVA-1 flight engine concept is considered in essence the final recommendation of the project engineers as to the system specifications of a feasible space-based nuclear engine that could be developed based on the findings and experience gained during the NERVA program (Finseth and NASA, 1991; Walton, 1991).

Furthermore, integer variables (e.g., number of engines) are not considered, therefore it is of interest that the incorporated scaling relationships may also represent clustered propulsion systems. Example specifications of such, including the NASA DRA (Drake, 2009) (3x Pewee-1 engines), Russian MEC (Koroteev et al., 2018) (4x NRE 68kN engines) and the latest in a series of mission trade-off analyses conducted by Aerojet Rocketdyne in conjunction with NASA (Joyner et al., 2020) (3x AR-NTP engines) are included and found to coincide with the defined scaling relationships, with a tendency to slightly overestimate performance requirements.

Mission	Engine	Cluster	Power (MW)	Thrust (kN)	$\dot{m}$ (kg/s)	Mass (kg)
DRA 5.0	Pewee-1	3	1521	333.6	37.8	9900
MEC	NRE 68 kN	4	1360	272	28.4	11560
Aerojet	AR-NTP	3	1627.2	334.8	38.1	11380

Table 7.2: Clustered NTP systems in existing crewed Mars mission concepts, sourced from Drake (2009), Koroteev et al. (2018) and Joyner et al. (2020).

This study initially identified the bi-modal NTER engine as a promising hybrid thermal/electric solution for future missions to Mars. Unfortunately, while with base reactor specifications derived from the Pewee-1 engine, the additional mass of turbomachinery required to provide bi-modal functionality and thus increased I_{sp}, results in the NTER engine specifications being largely incompatible with the derived propulsion system scaling relationships. As such, incorporation of this system is reserved for separate analyses.

7.2.1.2 Habitat

Salotti et al. (2014) present a method to estimate habitat mass breakdown based on the NASA TransHab specification. Similarly, this study applies subsystem specific scaling coefficients to estimate the reduced component mass/volume for 4 people relative to the established NASA 6-person habitat concept, as detailed in Table 7.3.

Consumables are divided in the following groups: 1) food 2) fecal canisters, urine pre-filters, and trash bags 3) personal hygiene kit, hygiene consumables, wipes, towels, and health care consumables 4) clothing and 5) personal stowage and operational supplies. Consumable requirements are largely taken from Lopez et al. (2015). Consumption rates for all logistics goods are defined using historical usage/resupply data from the ISS in combination with

Component	Crew = 6 Mass [kg]	Vol. [m^3]	% red.	Crew = 4 Mass [kg]	Vol. [m^3]
Power system	5840	–	22.5	4526	–
Avionics	290	0.1	0	290	0.1
ECLSS	3950	19.1	15	3357.5	16.235
TMS	1260	5.3	15	1071	4.505
Crew accommodations	4210	29.7	22.5	3262.75	23.0175
EVA systems	870	2.9	30	609	2.03
Structure	2020	–	22.5	1565.5	
(inc. 30% margin)	4920			3934.875	
Spares	4180	1.4	15	3553	1.19
Crew	6×80	–	–	4×80	–
Radiation shelter	6×739.36			4×739.36	
Total (no consumables)	32536.16			25447.06	

Table 7.3: Mass breakdown, adapted from Drake (2009); Salotti et al. (2014) with radiation 'storm' shelter estimate from Simon et al. (2013).

data from the Advanced Life Support Baseline Values and Assumptions Document, Human Integration Design Handbook, and Orion Commercial Crew Development design values. Mass scaling models were developed for each category, dependent on the number of crew members, flight duration, and stay time on/around Mars.

7.2.1.3 Power

Total power requirement is estimated at 60 kW, with 50 kW reserved for the crew habitation module (including 30 kW for life support, 15 kW for Zero-Boil-Off cryo-coolers and 5 kW for high data rate communications with Earth) (Drake, 2009).

The availability of solar power decreases proportional to the squared distance from the sun according to Equation 7.1, where the sun is assumed a perfect emitter (i.e., $\epsilon = 1$), the Stefan Boltzmann constant $\sigma = 5.67 \times 10^{-8}$ W/m^{2}K^4, the temperature and radius of the sun are $T = 5760$ K and $R_{sun} = 6.97 \times 10^5$ km respectively and R is the distance from the sun in km. Figure 7.2 shows the value of solar constant relative to radial distance (in astronomical units), including values at the aphelion and perihelion of Earth (1311 kW and 1401 kW) and Mars (488 kW and 711 kW) orbits for reference.

$$G_{SC} = \epsilon \sigma T^4 \frac{4\pi R_{sun}^2}{4\pi R^2} \tag{7.1}$$

Assuming PVA specifications in line with the NASA DRA (i.e., specific mass of 2 kg/m^2, power generation 0.12 kW/m^2 and 25% efficiency), the minimum required area of PVA can be taken as

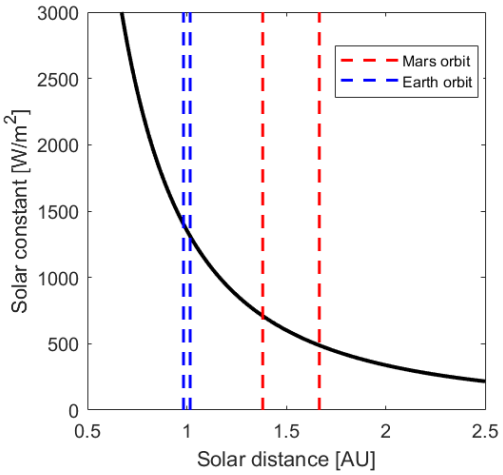


FIGURE 7.2: Solar constant at Earth and Mars.

500m^2, with an equivalent mass of 1000kg. However, the lateral diameter (and thus available area) of the panels will be constrained by the radiation shield shadow, given the detrimental effects of radiation on sensitive electronics.

7.2.1.4 Thermal Management

NTP systems utilise open-cycle cooling, where waste heat is rejected both by thermal radiation and exhaust convection. As such, NTP systems do not typically require a significantly large thermal management system, with the highest system demands likely present during engine cooldown. For this reason, thermal radiators (and any associated mass penalty) are not considered in detail within this study.

7.2.1.5 Radiation

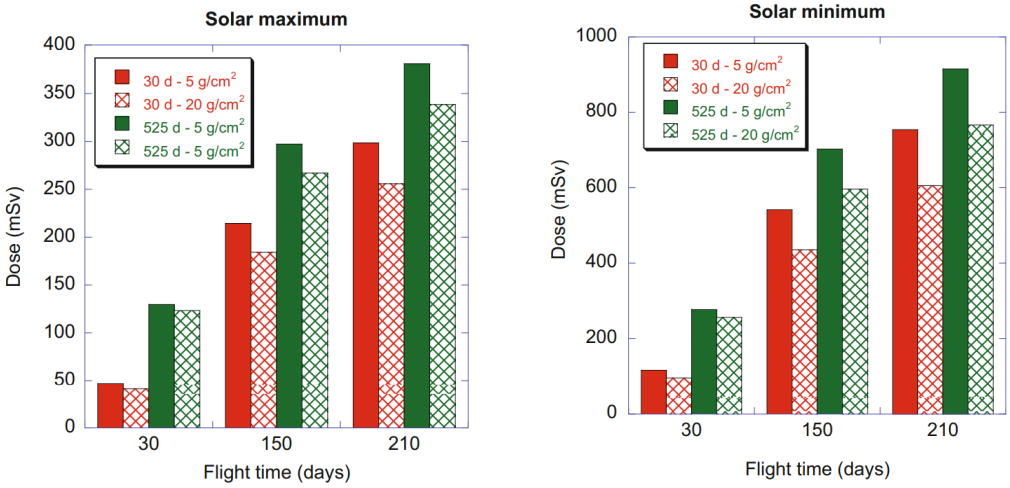
Concerning modern understanding of the undesirable biological effects associated with long-duration space habitation, human health considerations potentially represent the most demanding performance constraints for future interplanetary expeditions. Translated mission/vehicle design requirements relevant to the mitigation of acute and late risks caused by radiation exposure may be defined as follows:

- **Total accumulated dose for a single crew member must not exceed 500mSv.** This corresponds to a mission incorporating a one-way transfer time of 90 days, a Mars surface stay of 525 days (conjunction class), solar minimum conditions and spacecraft shielding equivalent to 5g/cm^2 Al (Durante and Bruno, 2010).
- **The CTV must include a dedicated means of protection against intense SPE radiation.** Such shelter, either internal or external, must provide radiation shielding

equivalent to 20-35g/cm^2 Al (Durante and Cucinotta, 2011).

- **Additional radiation due to nuclear propulsion must be minimised.** This will include considerations of component layout (for habitat occlusion), spacecraft length, reactor size, impulse duration/frequency and multi-layer external radiation shield sizing.

To address these requirements, approximate relationships are defined for expected radiation dose during transit and surface operations at Mars (see Figure 7.3) (Durante and Bruno, 2010). The latter is relevant given the availability of interplanetary transfer windows for a desired flight duration. As such, should the surface stay be of a necessarily long duration in comparison to the transit, additional environmental radiation dosage is accounted for.



(a) (b)

FIGURE 7.3: Estimated total equivalent doses for a Mars mission. Green bars are relative to a long stay on the Mars base (over 1 year), red bars for a Mars sortie (30 days on the planet). Solid bars are calculations for a spacecraft with a minimum shield (5 g/cm^2 Al); dashed bars for a thick shield in the spacecraft (20 g/cm^2 Al). All doses are calculated assuming the dose rate of the 1970 solar maximum (Durante and Bruno, 2010).

During transit, GCR shielding equivalent to 5g/cm^2 Al is assumed for the crew habitation module (Durante, 2014). Solar minimum conditions are considered for two primary reasons. First, missions during solar minimum are less likely to be affected by extreme SPEs, despite the larger GCR dosage. Furthermore, there is currently no reliable method of accurately predicting extreme SPEs, though they are know to occur at least once annually during 9 out of the 11 years in a solar cycle. Secondly, as this study does not assume any specific launch date, taking solar minimum conditions therefore gives a conservative estimate of radiation dosage[1]. In the case of SPEs, the crew may take shelter in a small shelter housed either within the crew

[1]The greater GCR levels present during solar minimum conditions (more difficult to protect against compared to extreme SPEs) represents the greatest risk to mission feasibility from a human health standpoint.

habitation module or within components of the external structure. For example, a radiation shelter housed between the hydrogen propellant tanks, as in Loeb et al. (2015), or a docked orbital return vehicle (e.g. an Orion capsule), as in Drake (2009). For this study, internal crew quarters are assumed to represent option (3) as presented by Simon et al. (2013), providing heavy radiation shielding equivalent to at least 20 g/cm^2 Al at a mass penalty of 739.36 kg/crew. Shelters may be utilised both to protect from SPEs and during nominal crew sleep cycles.

Based on the data presented by Durante and Bruno (2010), the rates of expected radiation dosage relative to both transit time and surface stay time (assuming solar minimum conditions and radiation shielding equivalent to either 5 g/cm^2 or 20 g/cm^2) are:

$$r_{transit(5)} = 1.7309 \text{ mSv/day} \tag{7.2}$$

$$r_{transit(20)} = 1.4017 \text{ mSv/day} \tag{7.3}$$

$$r_{stay} = 0.32423 \text{ mSv/day} \tag{7.4}$$

Of note, the reported dosage rate for the surface portion of the mission may be significantly reduced by constructing a heavily shielded surface habitat, e.g., by exploiting Martian regolith, as in Zubrin et al. (1991) (Durante and Bruno, 2010).

7.2.1.6 Structure

To account for the mass of components such as propellant tanks, external subsystems (e.g., communications arrays) and external structure (e.g., truss), the overall structural mass is assumed 10% of the initial spacecraft mass in LEO:

$$m_{structure} = 0.1 * (m_{engine} + m_{passenger} + m_{hab} + m_{consmables} + m_{structure})|_{t=t_0} \tag{7.5}$$

where the total mass of the spacecraft is assumed comprised of the mass of the engine m_{engine}, the passengers $m_{passenger}$, the habitat m_{hab}, the consumables $m_{consumables}$ and the external structure (including tanks, arrays, truss etc.) $m_{structure}$. This condition is formulated as an equality constraint evaluated at the initial time $t = t_0$.

7.2.2 Dynamics

The vehicle dynamics are defined via the following expressions for the rates of change of state vector $\mathbf{x} = [r, \theta, v_r, v_\theta, m_{prop}, m_{cons}]$ in an inertia reference frame, where r is the radial distance from the centre of the current *Sphere of Influence* (SOI), θ is the angular position, v_r

is the radial velocity, v_θ is the angular velocity and m_{prop} is the mass of propellant.

$$\dot{r} = v_r \tag{7.6}$$

$$\dot{\theta} = \frac{v_\theta}{r} \tag{7.7}$$

$$\dot{v}_r = \frac{v_\theta^2}{r} - \frac{\mu}{r^2} + \frac{\tau T_{max}}{m} \sin \alpha_{eng} \tag{7.8}$$

$$\dot{v}_\theta = -\frac{v_r v_\theta}{r} + \frac{\tau T_{max}}{m} \cos \alpha_{eng} \tag{7.9}$$

$$\dot{m}_{prop} = -\tau \dot{m}_{prop} \tag{7.10}$$

The scalar variable $\tau = [0,\ 1]$ is the throttle controlling the fraction of thrust applied, α_{eng} is the angle of the thrust vector and T_{max} is the maximum available engine thrust. The SOI for both Earth and Mars is given by,

$$SOI = a_{\text{E/M}} \left(\frac{m_{\text{E/M}}}{m_{sun}} \right)^{2/5} \tag{7.11}$$

where a is the semi-major axis and m is the planetary mass. A further assumption is that of perfectly circular orbits of Earth and Mars. This is to disentangle the initial mission analysis from dependence on considering specific launch dates/windows, instead allowing a more general approach that may be applicable across many future scenarios. The reference frames are centered on the primary gravitational body, either Earth, the Sun, or Mars depending on the mission phase.

7.3 Approach

The complete mission profile is divided into three distinct segments: the outbound Earth to Mars (EM) transfer, the stay time at Mars, and the inbound Mars to Earth (ME) transfer. Each interplanetary segment is composed of three phases: low-altitude planetary departure, interplanetary transfer and planetary capture to low altitude orbit.

The relationship between the primary mission segments is examined with respect to spacecraft design parameters by formulating a sequence of MOCPs. Of particular interest are the relative durations between each segment and the phasing between the outbound and inbound transfers.

Continuous control profiles for engine thrust were included to determine the optimal switching structure for each segment. More specifically, the engine throttle is defined as a continuous parameter evaluated at the control nodes, where the adaptive mesh refinement methodology of Chapter 6 is employed to determine the optimal locations of engine on/off switching whilst simultaneously minimising the size of the associated NLP problem.

The MOCP objectives were taken as the minimisation of total interplanetary transfer time (i.e., not including the stay time at Mars), and the gross vehicle mass (including any in-orbit

refuelling whilst in *Low Mars Orbit* (LMO)).

$$\min_{\mathbf{u}\in\mathbf{U},\mathbf{d}\in\mathbf{D}} \sum_{i=1}^{N} (t_{f,i} - t_{0,i}) \tag{7.12a}$$

$$\min_{\mathbf{u}\in\mathbf{U},\mathbf{d}\in\mathbf{D}} m(t_0) \tag{7.12b}$$

where N is the number of flight trajectory phases defined, t_0 and t_f are the initial and final times respectively, and

$$\mathbf{u} = \begin{bmatrix} \boldsymbol{\alpha}_i & \boldsymbol{\tau}_i & (\Delta t)_i \end{bmatrix} \tag{7.13}$$

is the trajectory control vector where $(\Delta t)_i = t_{f,i} - t_{0,i}$ is the time of flight for the i^{th} phase, $\boldsymbol{\alpha}$ is the set of thrust direction angles and $\boldsymbol{\tau}$ is the set of throttle values at every control node in the the i^{th} phase. The optimisation vector of static design parameters is given by,

$$\mathbf{d} = \begin{bmatrix} \phi_p(t_0) & P_{eng} & m_{str} \end{bmatrix} \tag{7.14}$$

where ϕ_p is the angle between Earth and Mars at the initial time, P_{eng} is the reactor power, m_{str} is the structural mass of the spacecraft. The gross mass $m_0 = m(t_0)$ is a sum of the calculated masses for the structure, engine system, propellant and tanks, crew, habitat and consumables.

The following additional assumptions were included:

- Prior to each transfer, the spacecraft is assumed positioned in a circular low-altitude planetary parking orbit (LEO, or LMO).
- A SOI trajectory approach is adopted, where the spacecraft is assumed affected by only the dominant gravitational force of the current phase.
- No advanced mid-flight events, e.g., mass jettisoning, refuelling, aero-braking or gravity assist fly-bys are considered, though the optimiser may include deep space manoeuvres whilst in interplanetary space.

7.3.1 Optimisation Algorithm

This work utilises the mesh-adaptive algorithm as presented in Chapter 6 of this thesis (Parsonage and Maddock, 2024b,a). Recall, this algorithm solves MOCPs using a direct approach, with individualistic grid adaptation facilitated by a local error analysis at element boundaries. Multiple objectives are considered using a dominance-based memetic framework applying both local and global search methods to a collaborative population of unique solutions. Together, this avoids the need for *a priori* specification of the quantity and temporal location of element boundaries, and the set of scalarisation weights defining the multi-objective descent

directions. Solution fidelity can thus increase concurrently with the exploration of the design space, which leads to increased numerical efficiency, particularly for discontinuous or highly non-linear problems, whilst propagating and maintaining population diversity.

In the following, system state and control profiles are represented using element-specific 9^{th} order Bernstein polynomials. A population of 10 solution vectors was initially discretised upon a uniform grid. Solution-specific grid adaptation was performed every 20 iterations, where the relative local error tolerance was set to $\epsilon_{tol} \leq 10^{-6}$. Inner-level optimisation was completed using the MATLAB fmincon solver with a SQP algorithm where function, constraint violation and step tolerances are set to 1e-6, 1e-12 and 1e-9 respectively. Each process was terminated upon reaching a maximum number of iterations, $n_{iter,max} = 200$.

7.4 Results

Results are presented for a full mission, including an outbound transfer, a period of time spent in the vicinity of Mars, and a return transfer. First, the initial outbound transfer was analysed separately, with results used to construct a set of interpolating surrogates representing the corresponding design parameters. These relationships were then used to analyse the complete mission with a reduced NLP size. Both analyses were formulated to minimise the time spent in transit (thus minimising the radiation risk to the crew), whilst simultaneously minimising the total propellant requirement (and thus cost).

7.4.1 Earth-Mars Transfer

The outbound EM transfer consists of three phases. The first assumes the spacecraft is initially positioned in LEO and bound to the Earth's gravitational field within the boundaries of the Earth's SOI. The second phase assumes a heliocentric orbit, influenced only by the Sun's gravity. Finally, the third phase assumes the spacecraft is influenced only by Mars, again within the boundaries of the corresponding SOI, with a fixed final state in LMO.

The vehicle is controlled by continuously varying both the magnitude and direction of the engine thrust, the former given the documented throttling capabilities of NTP systems (Finseth and NASA, 1991). The maximum thrust and mass flow rate of the engine, as well as it's overall mass, are dependant on a static optimisable parameter representing the thermal power of the reactor (see Section 7.2.1.1).

Figure 7.4 shows the mass breakdown of Pareto-optimal solutions obtained for the initial outbound mission segment. In line with traditional analyses of NTP systems (and conforming to the expected norms established by *Chemical Propulsion* (CP) systems), the primary driver is seen to be the propellant mass, directly related to the ΔV requirement for a particular

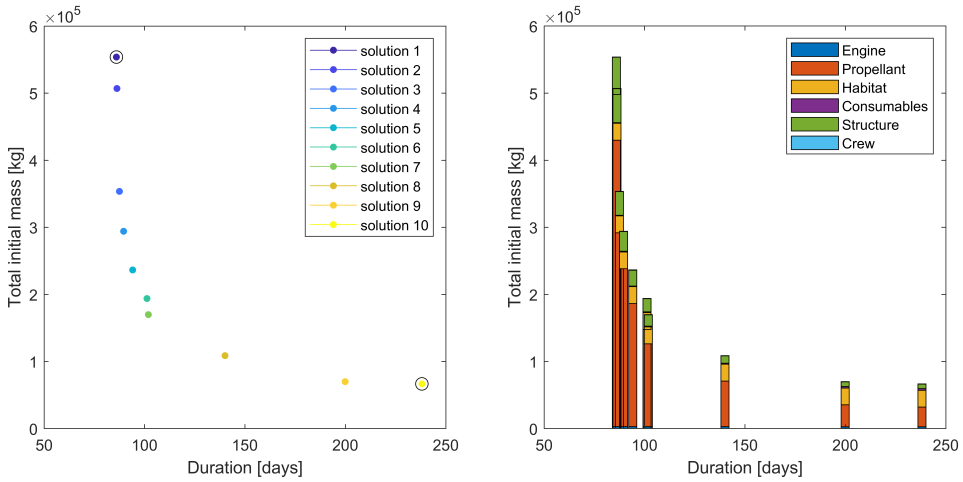


FIGURE 7.4: Outbound EM transfer: (a) Pareto-front (b) total initial mass breakdown against transfer duration.

#	t_{EM} [days]	m_{EM} [kg]	θ_{init} [deg]	P_{eng} [MW]	T_{eng} [kN]
1	86.0	5.54e5	17.2	406.5	83.9
2	86.3	5.07e5	17.2	381.2	78.1
3	87.5	3.54e5	17.8	377.8	77.3
4	89.6	2.94e5	17.8	381.1	78.0
5	94.1	2.36e5	18.3	391.5	80.5
6	101.2	1.94e5	18.9	308.8	61.3
7	101.9	1.70e5	20.1	389.0	79.9
8	140.1	1.09e5	29.8	372.4	76.0
9	199.9	0.70e5	32.7	321.4	64.2
10	238.1	0.67e5	38.4	319.7	63.8

Table 7.4: Outbound EM transfer: design parameters of Pareto-optimal solutions. Solution numbering corresponds to the legend in Figure 7.4.

transfer trajectory. This can be seen most clearly via comparison between the minimum-mass and minimum-time solutions, shown respectively in Figures 7.5 and 7.6. The minimum-mass solution here corresponds to the low-energy *conjunction* class transfer common in Mars mission analyses.

Table 7.4 summarises the design parameters of each Pareto-optimal solution. The optimiser is seen to favour NTP engines at the lower end of the performance spectrum represented by the engine sizing model (Section 7.2.1.1), generally preferring longer, sustained manoeuvres rather than short and intensive (particularly as transfer duration decreases). Importantly, the system is seen to be capable of a EM transfer duration of less than 90 days, corresponding to the original mission requirements.

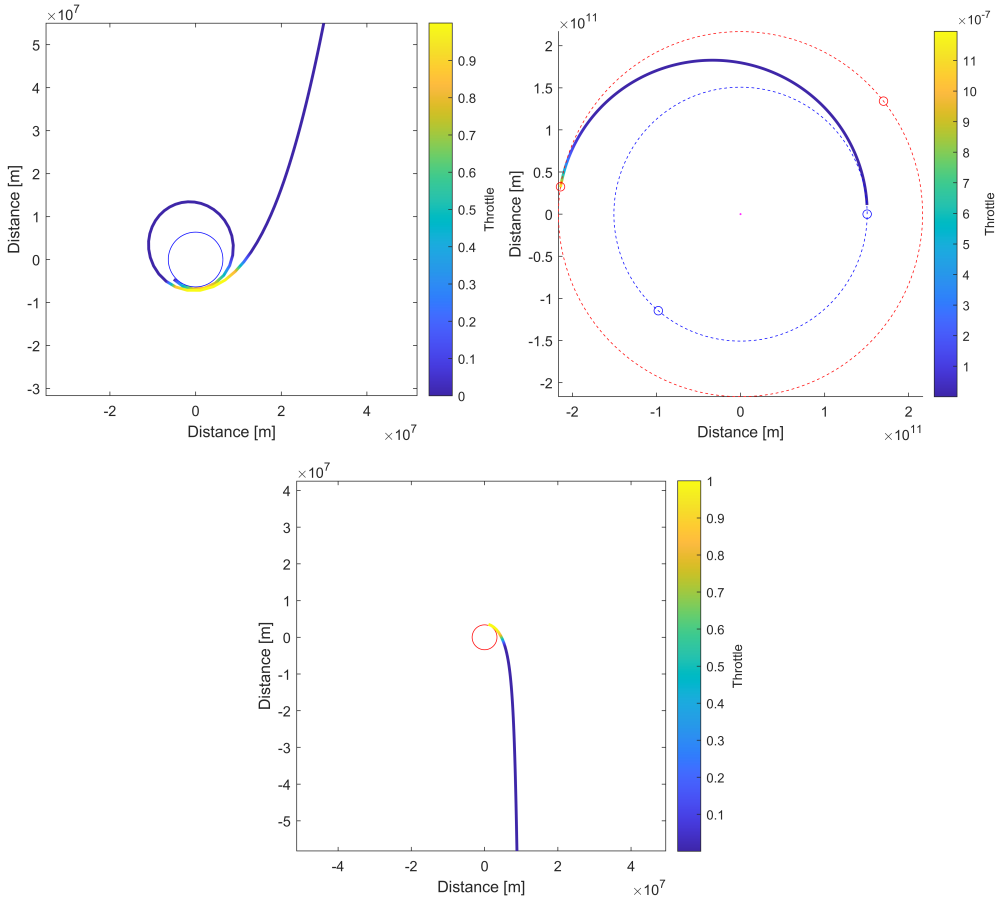


FIGURE 7.5: Outbound EM transfer: Minimum-mass solution.

7.4.2 Earth-Mars-Earth Transfer

The inbound transfer follows a period of time spent in the vicinity of Mars. To relate the planetary phasing at the commencement of the return leg to the overall mission timeframe, the initial outbound transfer duration and the Mars stay time are both included as optimisable static variables. To include the penalty associated to the outbound transfer duration and propellant expenditure, the Pareto-set presented in Section 7.4.1 are used to generate a set of continuous interpolating surrogate models representing relevant parameters of the EM transfer segment. These relationships, depicted in Figure 7.7, are made available to the optimiser such that the complete mission profile may be examined with continuity, but also with a significantly reduced NLP size. With regards to the depletable vehicle states, i.e. propellant and consumable masses, the availability of in-orbit manufacturing/refuelling whilst in LMO is assumed (and included in the calculation of total system mass).

Figure 7.8 shows the Pareto-set relating to the (a) outbound transfer and (b) inbound transfer, colour coded to represent the total system mass. Table 7.5 lists the design parameters cor-

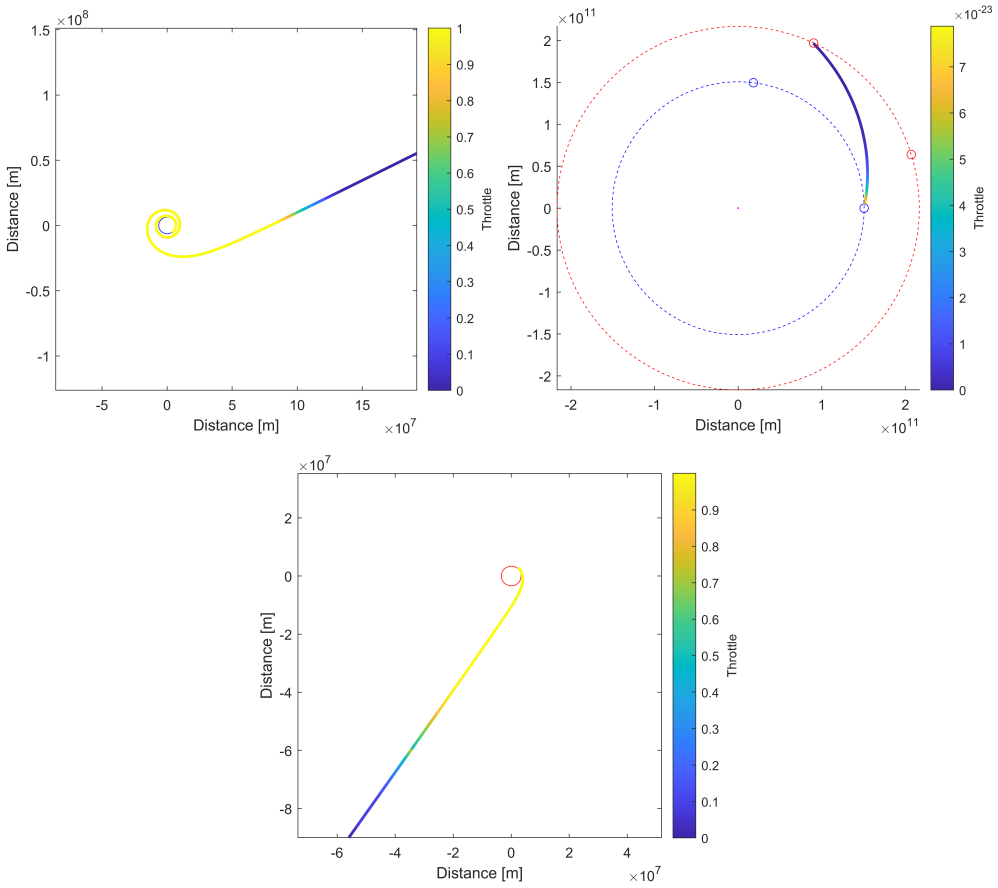


FIGURE 7.6: Outbound EM transfer: Minimum-time solution.

responding to each solution of the inbound MOCP. With regards to the inbound trajectories specifically, two distinct clusters may be seen. One contains the minimum-mass profiles. Each of these first performs a conjunction class outbound transfer. The stay time at Mars is then seen to be either very long (250-400 days) or very short ($\tilde{3}$0 days). The short stay solutions correspond to the two longest inbound transfers, thus it may be suggested that the switch back to short-stay from long-stay is triggered by the relative phasing between Earth and Mars. The minimum-mass solutions naturally exhibit the longest transfer durations, even passing outside the orbit of Mars during the inbound segment, thus may only be suitable for crewed expeditions provided an appropriate advancement in radiation shielding technology. The other contains the minimum-time profiles. These solutions all require the minimum allowable stay time at Mars (30 days), before immediately embarking on the return leg, itself corresponding to the traditional *opposition* class transfer. This class of transfer is usually identified as being unsuitable for crewed expeditions given the closer proximity to the Sun, thus inducing an increased dose of radiation to the crew. Furthermore, whilst the outbound transfer is capable of lying under the 90 day cutoff, the shortest return leg is seen to be almost double this value, despite the vastly increased propellant requirement.

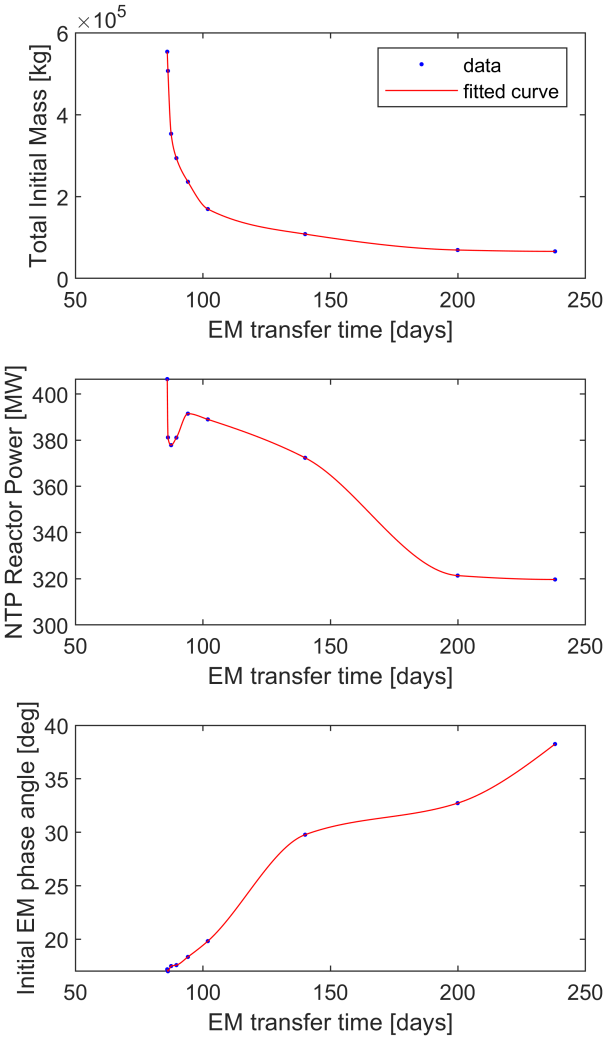


FIGURE 7.7: Outbound EM transfer parameter surrogates.

The propulsion system design parameters for the complete outbound/inbound mission are seen to follow the same trend as when examining the outbound transfer on it's own. That is, a tendency to lie within the lower end of the performance spectrum summarised in Table 7.1. This itself suggests that future development of NTP technology should be more focused on improving system reliability, particularly with respect to extended operational cycles and restartability, rather than base performance statistics.

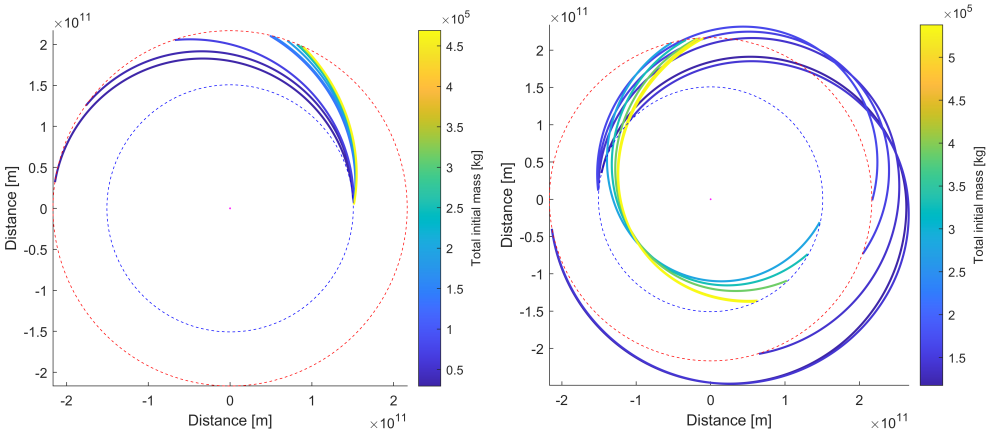


FIGURE 7.8: Multi-Objective solutions for (a) the outbound EM transfer and (b) the inbound ME transfer. Solution colours correspond to the numbering given in Figure 7.4.

#	t_{EM} [days]	t_{stay} [days]	t_{ME} [days]	m_{EM} [kg]	m_{ME} [kg]	P_{eng} [MW]
1	236.0	30.0	646.7	0.67e5	1.84e5	319.7
2	236.2	30.1	614.6	0.67e5	2.06e5	319.7
3	198.3	250.1	472.2	0.70e5	2.08e5	321.6
4	187.4	356.9	376.4	0.74e5	2.26e5	326.6
5	179.5	399.6	341.7	0.77e5	2.38e5	333.3
6	112.9	30.0	215.5	1.42e5	4.18e5	384.9
7	106.8	30.0	203.5	1.56e5	4.84e5	387.2
8	101.8	30.0	190.5	1.70e5	5.59e5	389.0
9	101.3	30.0	172.2	1.72e5	7.01e5	389.2
10	98.9	30.0	173.4	1.89e5	7.26e5	390.2

Table 7.5: Complete EME mission parameters.

7.5 Conclusion

A set of continuous numerical models representing the major subsystems of a crewed interplanetary transport vehicle were developed. Vehicle and mission performance design trade-offs were examined within a multi-disciplinary design optimisation framework.

The presented problem focused on optimising the vehicle design and mission phasing. More specifically, the engine sizing, propellant mass, dry mass (including the required mass of consumables) and the outbound/inbound trajectories (including a period of time spent at Mars). The transfer trajectory from Earth to Mars was first evaluated separately. The results of this analysis were used to generate simple numerical relationships representative of the outbound transfer design parameters. These relationships were then used to examine the complete mission architecture with respect to multiple performance objectives. These objectives were defined as the minimisation of the total transfer duration and simultaneously the gross vehicle

mass in LEO/LMO. Each analysis was completed using a multi-objective optimal control solver for non-linear constrained problems with multiple phases. This solver included an adaptive mesh refinement strategy with which to identify and refine optimal engine switching structures with respect to each objective.

Of particular interest were the relative durations between the three primary mission segments (outbound, stay, inbound). More specifically, the effect on the allowable stay time at Mars with respect to the minimisation of total time spent in transfer. It was found that mission profiles seemed to conform to distinct groups of solutions within objective space. This suggests a quantifiable set of transfer opportunities/mission profiles differentiated by total propellant requirement and desired duration of surface activities at Mars.

An additional finding was the universal agreement to reduce the size of the propulsion system to the lower end of the performance spectrum represented by the developed NTP sizing relationships. Whilst these performance requirements still represent a fundamental shift from traditional CP systems (particularly in terms of Specific Impulse), the implication is that current and/or future development of NTP technology would benefit from a focus on improving system reliability rather than base performance statistics. This could include investigations into, for example, extended operational cycles and restartability.

Finally, it was observed that the adaptive mesh refinement strategy was able to successfully define optimal engine switching structures unique to solutions at opposite ends of the Pareto front. For example, minimum-time solutions tended to include one prolonged burn of higher intensity, whereas minimum energy solutions included multiple, lower intensity burns timed to coincide with the periapses of the current orbit. All solutions represented their throttle control structure with *bang-bang* profiles, notoriously difficult for uniform-grid polynomials to approximate. Using the adaptive algorithm, grid points were gradually concentrated around the temporal location of each switch unique to each individual solution, ultimately minimising the total number of optimisation parameters required to represent the complete population of alternative trajectories.

Acknowledgements

Some of the work presented as part of this chapter was partially funded through a grant by the European Space Agency General Support Technology Programme (ESA 3-16642/20/NL/MG/rk).

7.6 Bibliography

W. V. Braun. Manned Mars Landing: Presentation to the space task group. 1969.

C. Bruno and C. Dujarric. In-space nuclear propulsion. *Acta Astronautica*, 82(2):159–165, 2013. ISSN 00945765. doi: 10.1016/j.actaastro.2012.08.022. URL http://dx.doi.org/10.1016/j.actaastro.2012.08.022.

P. R. Chai, K. T. McBrayer, A. C. McCrea, R. G. Merrill, A. Prasad, and M. Qu. Crewed mars mission mode options for nuclear electric/chemical hybrid transportation systems. *Accelerating Space Commerce, Exploration, and New Discovery conference, ASCEND 2021*, 2021. doi: 10.2514/6.2021-4136.

B. G. Drake. *Human Exploration of Mars Design Reference Architecture 5 . 0 Addendum*. Number July. 2009. ISBN 3016210134.

C. Dujarric. A new propulsion concept for interplanetary missions. *ESA Bulletin*, 108 (november):66–71, 2001.

C. Dujarric, A. Santovincenzo, and L. Summerer. The Nuclear Thermal Electric Rocket: A Proposed Innovative Propulsion Concept for Manned Interplanetary Missions. *Progress in Prpulsion Physics*, 4:293–312, 2013. doi: 10.1051/eucass/201304.

M. Durante. Space radiation protection: Destination Mars. *Life Sciences in Space Research*, 1(1): 2–9, 2014. ISSN 22145532. doi: 10.1016/j.lssr.2014.01.002. URL http://dx.doi.org/10.1016/j.lssr.2014.01.002.

M. Durante and C. Bruno. Impact of rocket propulsion technology on the radiation risk in missions to Mars. *European Physical Journal D*, 60(1):215–218, 2010. ISSN 14346060. doi: 10.1140/epjd/e2010-00035-6.

M. Durante and F. A. Cucinotta. Physical basis of radiation protection in space travel. *Reviews of Modern Physics*, 83(4), 2011. ISSN 00346861. doi: 10.1103/RevModPhys.83.1245.

J. Finseth and NASA. Overview of ROVER Engine Tests: Final Report. *Discovery*, (3), 1991.

N. Frischauf and B. A. Hamilton. Space nuclear power and propulsion - A basic tool for the manned exploration of the solar system. *Proceedings of the 2004 International Congress on Advances in Nuclear Power Plants, ICAPP'04*, (July 2004):2210–2216, 2004.

C. R. Joyner, M. Eades, J. Horton, T. Jennings, T. Kokan, D. J. Levack, B. J. Muzek, and C. B. Reynolds. LEU NTP Engine System Trades and Mission Options. *Nuclear Technology*, 206 (8):1140–1154, 2020. ISSN 19437471. doi: 10.1080/00295450.2019.1706982. URL https://doi.org/10.1080/00295450.2019.1706982.

D. Koenig. Experience Gained from the Space Nuclear Rocket Program (Rover). Technical report, 1986.

A. S. Koroteev, V. N. Akimov, L. E. Zakharenkov, A. V. Karevskiy, E. Y. Kuvshinova, A. V. Semenkin, and A. E. Solodukhin. Architecture Features and Application Aspects of High-Power Spacecrafts With Electric Propulsion. (May), 2018.

D. F. Landau and J. M. Longuski. A reassessment of trajectory options for human missions to Mars. *Collection of Technical Papers - AIAA/AAS Astrodynamics Specialist Conference*, 2: 836–871, 2004. doi: 10.2514/6.2004-5095.

D. F. Landau and J. M. Longuski. Trajectories for Human Missions to Mars, Part II: Low-Thrust Transfers. *Journal of Spacecraft and Rockets*, 43(5):1043–1048, 2006. ISSN 0022-4650. doi: 10.2514/1.a12253b.

H. W. Loeb, V. G. Petukhov, G. A. Popov, and A. I. Mogulkin. A realistic concept of a manned Mars mission with nuclear-electric propulsion. *Acta Astronautica*, 116:299–306, 2015. ISSN 00945765. doi: 10.1016/j.actaastro.2015.07.019. URL http://dx.doi.org/10.1016/j.actaastro.2015.07.019.

P. Lopez, E. Schultz, B. Mattfeld, C. Stromgren, and K. Goodliff. Logistics needs for potential deep space mission scenarios post Asteroid Redirect crewed Mission. *IEEE Aerospace Conference Proceedings*, 2015-June, 2015. ISSN 1095323X. doi: 10.1109/AERO.2015.7119161.

C. Maddock, B. Parsonage, L. Ricciardi, M. Vasile, and O. Cohen. Design optimisation and analysis of very high power transportation system to Mars. *Proceedings of the International Astronautical Congress, IAC*, A5(October):25–29, 2021. ISSN 00741795.

E. National Academies of Sciences and Medicine. *Space Nuclear Propulsion for Human Mars Exploration*. 2021. ISBN 0309684803.

S. R. Oleson, L. Burke, L. Dudzinski, J. Fittje, L. S. Mason, T. Packard, P. Schmitz, J. Gyekenyesi, and B. Faller. A Combined Nuclear Electric and Chemical Propulsion Vehicle Concept for Piloted Mars Opposition Class Missions. (January 2021), 2020. doi: 10.2514/6.2020-4055.

A. C. Owens, C. A. Jones, W. M. Cirillo, J. J. Klovstad, E. L. Judd, P. R. Chai, R. G. Merrill, A. Prasad, J. Cho, and C. Stromgren. Integrated crewed mars mission analysis, part ii: The trajectory strikes back. *Accelerating Space Commerce, Exploration, and New Discovery conference, ASCEND 2021*, 2021. doi: 10.2514/6.2021-4134.

B. Parsonage and C. Maddock. Biased Dyadic Crossover for Variable-Length Multi-Objective Optimal Control Problems. *2024 IEEE Congress on Evolutionary Computation, CEC 2024 - Proceedings*, pages 1–9, 2024a. doi: 10.1109/CEC60901.2024.10611888.

B. Parsonage and C. Maddock. Solving multi-objective optimal control problems using a multiresolution approach. *Journal of Guidance, Control, and Dynamics*, pages 1–14, 2024b. ISSN 1533-3884. doi: 10.2514/1.g008223.

B. Parsonage and C. Maddock. Multi-Objective Design Optimisation and Analysis of a Crewed Earth-Mars Transportation System using Nuclear Thermal Propulsion. *Proceedings of the International Astronautical Congress, IAC*, (October), 2024c.

J. M. Salotti, R. Heidmann, and E. Suhir. Crew size impact on the design, risks and cost of a human mission to mars. *IEEE Aerospace Conference Proceedings*, (March), 2014. ISSN 1095323X. doi: 10.1109/AERO.2014.6836241.

L. R. Shepherd and A. V. Cleaver. The atomic rocket. I. *J. Brit. Interplanetary Soc.*, 7, 1948.

M. A. Simon, M. Clowdsley, and S. Walker. Habitat design considerations for implementing solar particle event radiation protection. *43rd International Conference on Environmental Systems*, pages 1–13, 2013. doi: 10.2514/6.2013-3403.

J. T. Walton. An overview of tested and analyzed ntp concepts. *AIAA/NASA/OAI Conference on Advanced SEI Technologies, 1991*, 1991. doi: 10.2514/6.1991-3503.

E. A. Willis. Comparison of Trajectory Profiles and Nuclear-Propulsion-Module Arrangemenets for Manned Mars and Mars-Venus Missions. 1971.

P. D. Wooster, R. D. Braun, J. Ahn, and Z. R. Putnam. Mission design options for human Mars missions. *Mars*, 2007. ISSN 1548-1921. doi: 10.1555/mars.2007.0002.

V. Zakirov and V. Pavshook. Russian Nuclear Rocket Engine Design for Mars Exploration. *Tsinghua Science and Technology*, 12(3):256–260, 2007. ISSN 10070214. doi: 10.1016/S1007-0214(07)70038-X.

R. M. Zubrin, D. A. Baker, O. Gwynne, M. M. Astronautics, and P. O. Box. Mars Direct : A Simple , Robust , and Cost Effective Architecture for the Space Exploration Initiative. 1991.

Conclusions 8

The overall aim of this thesis was to contribute techniques for improving the numerical efficiency of multidisciplinary/multi-objective design optimisation methods. This was approached broadly by focusing on *multi-fidelity* methods that can utilise various sources of competing information to increase the attainable accuracy/reliability whilst simultaneously reducing the associated computational burden. Being particularly relevant within the field of aerospace engineering, an emphasis was placed on *Multi-Objective Optimal Control Problems* (MOCPs) and analyses thereof. This goal was broken down into a number of distinct research questions:

- How can alternative levels of numerical fidelity be mathematically defined and incorporated into the formulation of a MOCP?
- How can the optimisation process be designed to adaptively manage all available fidelity levels whilst balancing the inherent cost/accuracy trade off?
- What are the operational limits of including the developed methodology with respect to traditional single-fidelity approaches?
- Are the computational savings (if any) sufficient to justify the cost of incorporating such approaches?

These questions were first addressed by considering the source of the available information. That is, whether a particular representation of a physical process can be described as *external* or *internal* to the problem solution framework itself. External sources include numerical models, simulations, experimental data or other sources inherently separate from the design and analysis framework. External sources can, for example, be operated as *black-box* functions, simply returning the output values following a query by an outer-loop framework. On the other hand, internal sources of information are defined by the formulation, discretisation, or the general mathematical description of the problem (for instance, the dimensionality).

Accordingly, this work focused on the exploration and development of two distinct, but fundamentally linked, methodologies enabling the incorporation of multi-fidelity information sources. Both allow the systematic and adaptive upgrading of system/subsystem modelling fidelity for the purpose of solving *Multi-Objective Optimal Control Problems* (MOCPs) in a computationally efficient manner. These methodologies operate according to the following respective strategies according to the primary source of multi-fidelity information utilised therein:

- An **external** strategy, where alternative numerical models of the same subsystem(s), competing in accuracy and computational cost, are simultaneously incorporated into one single approximation.
- An **internal** strategy, where the numerical complexity of the problem itself can be systematically upgraded, concentrating computational resource where it is most needed.

Each strategy is incorporated into a separate numerical framework intended for application in scenarios where in-depth and exclusively high-fidelity models/analyses are not viable. For example, the preliminary design and optimisation phases of complex multidisciplinary engineering systems.

Part 1 was dedicated to a *model-based* framework for **external** multi-fidelity information. This framework assumes that any particular subsystem of interest of a complex engineering system is able to be represented by several alternative numerical models/formulations. It is assumed that these methods generally conform to an inherent hierarchical trade-off between the cost of evaluation and the relative accuracy to which the underlying physical process can be represented. This framework is therefore primarily concerned with utilising, in the most computationally efficient manner possible, the available information from all data sources such that the highest numerical accuracy can be achieved at a minimum cost.

Following a review of the relevant literature in Chapter 2, Chapter 3 introduced a complete *Multi-Fidelity Model Management* (MFMM) framework in which computationally inexpensive surrogate models of selected subsystems could be autonomously generated and progressively corrected using high-fidelity data. Several data correction methods were formulated and investigated, before being integrated into a modular outer-level structure (Section 3.1.2). Each method was designed to operate *cooperatively*, leveraging a particular aspect of the correlation(s) between the available data. Accordingly, each was tested in isolation and in several sequential combinations using a suite of numerical testing functions (Section 3.1.4).

The overall framework itself was designed to be as autonomous as possible, with a level of flexibility appropriate for application to general problems. For instances, any number of

competing/alternative models may be specified for a particular subsystem. Models themselves may have non-equal (or overlapping) input domains, and may be time-phase specific or shared across multiple phases (for dynamic problems). In addition to (or instead of) numerical models, existing data may be supplied to supplement the surrogate generation procedure. If none exists, initial sample points at which models are to be evaluated may be alternatively supplied, or simply determined autonomously.

Surrogate models were constructed using an *Artificial Neural Network* (ANN)-based modelling procedure (Section 3.1.3), again with the intention of allowing as general an application as possible. In this case, an autonomous hyper-parameter selection and training routine was implemented to allow the overall framework the ability to efficiently approximate any non-linear function to a high degree of accuracy, without requiring any *a priori* knowledge of the underlying relationship.

Finally, the complete MFMM framework was demonstrated in Chapter 4 with an engineering test case representing the preliminary design of an atmospheric sample re-entry vehicle, where multiple levels of fidelity are used to represent the vehicle aerodynamics. It was found that the sequential application of corrective stages allowed a greater attainable approximation accuracy relative to a single-fidelity equivalent model for a set number of high-fidelity samples up to some critical sample size, which matches expectation.

Part 2 was dedicated to a *resolution-based* framework for **internal** multi-fidelity information. Rather than using multiple external sources of information to enhance the design procedure, this approach instead systematically refined the numerical complexity at which the problem solution was represented. This allowed computational resource to be more appropriately allocated as the algorithm progressed. For example, within areas that experience high rates of change, or were particularly dominant with respect to the final solution. This approach was designed specifically for direct *Multi-Objective Optimal Control Problems* (MOCPs), where the time-dependant state and control profiles representing feasible solutions are evaluated by subdividing the time domain into a distinct grid of control nodes, across which approximating polynomial relationships may be fit.

Again following a review of the relevant literature in Chapter 5, this framework was introduced and described in Chapter 6. The focus was on adaptively increasing the *resolution* of a set of discretised Optimal Control solutions. This was achieved by concentrating discretisation nodes/grid-points around areas of high complexity unique to each solution (Section 6.2). Importantly, this meant that a set of independent solutions could be processed concurrently, allowing each to represent a particular balance between the objectives of a MOO problem. This was achieved using easily calculable relative local error values evaluated at finite-element boundaries, avoiding the need for any secondary optimisation or costly integrations.

A novel aspect of the proposed approach is the simultaneous handling of multiple such unique grid structures within an evolutionary architecture (Section 6.4). This allowed individual solutions to independently evolve concurrent with their propagation through objective space, aided via collaborative (social) action with neighbouring solutions. This was enabled by employing a unique social operator, designed to facilitate such actions between solutions of variable complexity. A complete MOCP may thus be treated as an *open-ended* evolutionary design problem, ultimately allowing a more complete, higher-fidelity representation of the trade-off surface in a more numerically efficient manner (Section 6.5).

This new social operator was first tested against several baseline and commonly applied approaches (Section 6.6). The complete *multi-resolution* framework was then applied to several demonstrative test cases, where the associated *Multi-Objective Optimal Control Problem* (MOCP) was treated as an open-ended evolutionary design problem (Section 6.7). The computational cost, achievable solution accuracy and population spread/diversity were measured in each case and compared against a traditional uniform-grid (analogous to single-fidelity) approach. An order of magnitude reduction in computational cost required to reach a desired level of local mesh error was observed for all cases. Furthermore, the progressive nature of population refinement (i.e. moving from coarse to detailed solutions, corresponding to a move from global exploration to local exploitation) was shown to facilitate a more complete and diverse representation of the Pareto-front. The new social operator was found to successfully promote offspring diversity throughout each open-ended evolutionary process by allowing a finer level of social action to occur between similar solutions.

Finally, as with Part 1, Chapter 7 presented a more in-depth engineering test case aimed at demonstrating a practical application of the proposed multi-resolution framework. In this example, the preliminary design and assessment of a high-power interplanetary transportation system using nuclear propulsion was presented. Continuous control profiles were employed to represent the propulsion system operation, with the proposed multi-resolution framework shown to successfully identify and refine an optimal throttle-switching structure for the complete outbound/inbound mission with respect to multiple, competing objectives.

It is clear from this work that the potential for introducing higher levels of autonomy into the design and optimisation process of complex engineering systems remains strong. With respect to both *external* and *internal* information sources, this work has shown that the collaborative incorporation of multiple levels of simulation fidelity represents a significant quantitative improvement over standard single-fidelity approaches, particularly considering cases (such as early stage analyses) where extensive high-fidelity simulation is impractical.

It may still be argued that the cost of establishing multi-fidelity approaches erodes their resultant savings over traditional single-fidelity analyses. This in part may be due to the inherent

problem-specific nature of many such analyses. This work, however, has demonstrated the application of such approaches within a more general and autonomous framework, thus the repeated cost of incorporation reduces substantially. That is, the wider range of problems that this framework, or others like it, can be used to solve, the lower the ultimate cost of applying such methods is in the first place.

With the wealth of numerical codes available to the modern designer (including those designed to take advantage of ever increasing computational processing power), methods to ensure operational efficiency are essential.

8.1 Recommendations for Future Work

During the course of this research, several promising areas of further work were identified:

- An important limitation of the proposed methodologies is an inherent assumption regarding the fidelity level assigned to each source. For *internal* sources, this distinction is perhaps trivial, a higher resolution mathematical description naturally implies a higher level of available fidelity. For *external* sources however, the picture may be less clear. In the presented formulation, at least one source must be designated the high-fidelity source. This can usually be linked to computational cost (i.e. higher fidelity data is more expensive to accumulate), but this may not always be true. Particularly when considering larger sets of models, the distinction between fidelity levels may become blurred, making it harder to assign relative values. An important area of research would then be to address this issue through means of, for example, automated confidence assignment, using measures such as both the agreement between alternative models and the actual cost of data acquisition.
- Direct quantification of objective function sensitivity relative to the included design variables via multi-fidelity analysis could be a valuable area of research. Such information could feasibly be used in conjunction with traditional monitoring techniques to provide measures of problem convergence and design confidence.
- In a wider, more fundamental sense, employing the examined MFMM methods within the context of *Uncertainty Quantification* (UQ) could also be a useful area of future study. In this case, the direct incorporation of multi-fidelity representations could help identify and quantify uncertainty with respect to the modelling of the simulated physical processes. This would allow the delivery of solutions with included confidence bounds, or designed specifically to be robust under uncertainty.
- Considering the presented investigation into variable-length crossover techniques for MOCPs, further analyses considering more complex design scenarios would add a

valuable level of assessment and verification. That is, cases where the effectiveness of the crossover operator plays a more definitive role in the quality/performance of the resulting solutions.

- Common to other areas of study within this field, the excessive NLP growth of open-ended optimisation problems remains a pressing issue outwith passive/indirect mitigation. Future study could include an adaptive approach to validate growth, explicitly detecting over-refinement and ultimately removing unnecessary nodes/segments without incurring additional expense or loss of crucial information.

- This work separately addressed methods to incorporate *internal* and *external* sources of information, albeit employing largely analogous management methods within each respective framework. The next step could then be to incorporate both types of source into the same framework, investigating whether the same management strategies can be successfully implemented and the consequences thereof. A formalised *Multidisciplinary Design Optimisation* (MDO) architecture could be of particular use here, considering that the resolution-based refinement presented in Part II could be treated as a multi-fidelity 'control' discipline. A single MDO framework could therefore feasibly include the traditional external multi-fidelity disciplines (i.e. disciplines dependant on information obtained from one or more external functions of varying fidelity) as well as internal multi-fidelity disciplines. The interaction of these components would be an interesting area for further study, most likely requiring a more robust effort to quantify and assess the confidence of the obtained solution(s).

- An aspect not addressed by the presented work is the adaptation of the polynomial basis functions used within the MOCPs. Refining the order of the approximating polynomials as well as the grid resolution would thus imply full *hp*-adaptivity, an established upgrade to *h*-adaptivity on it's own. For instance, the work by Patterson et al. (2015), in reference to the problem reproduced in Section 6.7.3, suggests a positive effect in solution accuracy and cost by increasing the polynomial degree (within user-defined applicability bounds) *as opposed* to the number of mesh elements.

- Whilst efforts were made throughout to compare the obtained results with those found in the literature (e.g. for several of the problems used as test cases across Parts I and II), a more targeted comparison between alternative multi-fidelity approaches applied to the same isolated test cases could be performed. For example, a particularly valuable direct comparison could be against the co-kriging methodology applied to the benchmark functions and test cases presented in Part I, given the prevalence of this approach in the statistics community and beyond.

Full List of References

O. Abdelkhalik. Hidden genes genetic optimization for variable-size design space problems. *Journal of Optimization Theory and Applications*, 156(2):450–468, 2013.

B. M. Adams, W. J. Bohnhoff, K. R. Dalbey, M. S. Ebeida, M. S. Eldred, J. D. Jakeman, L. P. Swiler, J. A. Stephens, D. M. Vigil, and T. M. Wildey. DAKOTA, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.16 User's Manual. *Sandia Technical Report SAND2010-2183*, (Updated November 7, 2013 Dakota,), 2022.

A. Aguilera and R. Pérez-Aguila. General n-dimensional rotations. *Proceedings of the WSCG SHORT Communications Papers*, pages 1–8, 2004. URL papers3://publication/uuid/6ADECEC1-F089-42FC-AA61-81C20D0E8B26.

T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.

N. Alexandrov, M. D. O. Branch, N. Larc, J. E. Dennis, R. Michael, L. Icase, and V. Torczon. A Trust Region Framework for Managing the Use of Approximation Models in Optimization. *Engineering*, (97), 1997.

N. M. Alexandrov, R. M. Lewis, C. R. Gumbert, L. L. Green, and P. A. Newman. Approximation and model management in aerodynamic optimization with variable-fidelity models. *Journal of Aircraft*, 38(6):1093–1101, 2001. ISSN 00218669. doi: 10.2514/2.2877.

M. H. Aliefa and S. Suyanto. Variable-Length Chromosome for Optimizing the Structure of Recurrent Neural Network. *2020 International Conference on Data Science and Its Applications, ICoDSA 2020*, 2020. doi: 10.1109/ICoDSA50139.2020.9213012.

D. Allaire and K. Willcox. A mathematical and computational framework for multifidelity design and analysis with computer models. *International Journal for Uncertainty Quantification*, 4(1):1–20, 2014. ISSN 21525099. doi: 10.1615/Int.J.UncertaintyQuantification.2013004121.

M. Alvarez and N. D. Lawrence. Sparse convolved Gaussian processes for multi-output regression. *Advances in Neural Information Processing Systems 21 - Proceedings of the 2008 Conference*, pages 57–64, 2009.

D. G. Anderson. Iterative Procedures for Nonlinear Integral Equations. *Journal of the ACM (JACM)*, 12(4):547–560, 1965. ISSN 1557735X. doi: 10.1145/321296.321305.

J. D. Anderson Jr. *Hypersonic and high-temperature gas dynamics*. American Institute of Aeronautics and Astronautics, 2006. ISBN 1600861954.

J. Andersson. *A survey of multiobjective optimization in engineering design*. PhD thesis, Linktjping University, Sweden, 2000.

J. Andrés Christen and C. Fox. Markov chain Monte Carlo using an approximation. *Journal of Computational and Graphical Statistics*, 14(4):795–810, 2005. ISSN 10618600. doi: 10.1198/106186005X76983.

F. Assassa and W. Marquardt. Dynamic optimization using adaptive direct multiple shooting. *Computers and Chemical Engineering*, 60:242–259, 2014. ISSN 00981354. doi: 10.1016/j.compchemeng.2013.09.017. URL http://dx.doi.org/10.1016/j.compchemeng.2013.09.017.

C. Audet, G. Savard, and W. Zghal. A mesh adaptive direct search algorithm for multiobjective optimization. *European Journal of Operational Research*, 204(3):545–556, 2010. ISSN 03772217. doi: 10.1016/j.ejor.2009.11.010. URL http://dx.doi.org/10.1016/j.ejor.2009.11.010.

A. Auger, J. Bader, D. Brockhoff, and E. Zitzler. Theory of the hypervolume indicator: Optimal μ-distributions and the choice of the reference point. *Proceedings of the 10th ACM SIGEVO Workshop on Foundations of Genetic Algorithms, FOGA'09*, pages 87–102, 2009. doi: 10.1145/1527125.1527138.

F. Aurenhammer. Voronoi Diagrams - A Survey of a Fundamental Data Structure. *ACM Computing Surveys (CSUR)*, 23(3):345–405, 1991.

V. Aute, K. Saleh, O. Abdelaziz, S. Azarm, and R. Radermacher. Cross-validation based single response adaptive design of experiments for Kriging metamodeling of deterministic computer simulations. *Structural and Multidisciplinary Optimization*, 48(3):581–605, 2013. ISSN 16151488. doi: 10.1007/s00158-013-0918-5.

I. Babuška and M. Suri. The p-and hp versions of the finite element method, an overview. *Computer methods in applied mechanics and engineering*, 80(1-3):5–26, 1990. ISSN 0045-7825. doi: https://doi.org/10.1016/0045-7825(90)90011-A.

J. Bader and E. Zitzler. HypE : An Algorithm for Fast Hypervolume-Based. *Evolutionary computation*, 19(1):45–76, 2011.

S. Bae, H. S. Shin, A. Savvaris, L. Vaios, and A. Tsourdos. Multi-objective suborbit/orbit trajectory optimisation for spaceplanes. *Acta Astronautica*, 170(February 2019):431–442, 2020. ISSN 00945765. doi: 10.1016/j.actaastro.2020.01.003. URL https://doi.org/10.1016/j.actaastro.2020.01.003.

B. K. Bairstow, O. De Weck, and J. Sobieszczanski-Sobieski. Multiobjective optimization of two-stage rockets for earth-to-orbit launch. *Collection of Technical Papers - AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, 3(May):1585–1600, 2006. ISSN 02734508. doi: 10.2514/6.2006-1720.

M. Bakr, J. W. Bandler, K. Madsen, and J. Sondergaard. An Introduction to the Space Mapping Technique. 103(3):239–248, 2002. ISSN 00138703. doi: 10.1023/A.

M. H. Bakr. A hybrid aggressive space-mapping algorithm for em optimization. *IEEE Transactions on Microwave Theory and Techniques*, 47(12):2440–2449, 1999. ISSN 00189480. doi: 10.1109/22.808991.

M. H. Bakr, J. W. Bandler, K. Madsen, J. E. Rayas-Sanchez, and J. Sondergaard. Space mapping optimization of microwave circuits exploiting surrogate models. *IEEE MTT-S International Microwave Symposium Digest*, 3(June):1785–1788, 2000. ISSN 0149645X. doi: 10.1109/mwsym.2000.862325.

J. W. Bandler. Electromagnetic Optimization Exploiting Aggressive Space Mapping. 1995.

J. W. Bandler, R. M. Biernacki, S. H. Chen, P. A. Grobelny, and R. H. Hemmers. Space mapping technique for electromagnetic optimization. *IEEE Transactions on Microwave Theory and Techniques*, 1994.

J. W. Bandler, Q. S. Cheng, S. A. Dakroury, A. S. Mohamed, M. H. Bakr, K. Madsen, and J. Søndergaard. Space Mapping: The State of The Art. *SBMO/IEEE MTT-S International Microwave and Optoelectronics Conference Proceedings*, 52(1):951–956, 2003.

J. W. Bandler, Q. S. Cheng, N. K. Nikolova, and M. A. Ismail. Implicit space mapping optimization exploiting preassigned parameters. *IEEE Transactions on Microwave Theory and Techniques*, 52(1 II):378–385, 2004. ISSN 00189480. doi: 10.1109/TMTT.2003.820892.

R. E. Bank, T. F. Dupont, and H. Yserentant. The hierarchical basis multigrid method. *Numerische Mathematik*, 52(4):427–458, 1988. ISSN 0029599X. doi: 10.1007/BF01462238.

W. Banzhaf, G. Beslon, S. Christensen, J. A. Foster, F. Képès, V. Lefort, J. F. Miller, M. Radman, and J. J. Ramsden. Guidelines: From artificial evolution to computational evolution: A

research agenda. *Nature Reviews Genetics*, 7(9):729–735, 2006. ISSN 14710056. doi: 10.1038/nrg1921.

J. F. Barthelemy and R. T. Haftka. Approximation concepts for optimum structural design - a review. *Structural Optimization*, 5(3):129–144, 1993. ISSN 09344373. doi: 10.1007/BF01743349.

R. R. Barton. Metamodels for simulation input-output relations. 9(x):49–54, 1992.

M. J. Bayarri, R. Paulo, J. O. Berger, J. Sacks, J. A. Cafeo, J. Cavendish, C. H. Lin, and J. Tu. A framework for validation of computer models. *Technometrics*, 49(2):138–154, 2007. ISSN 00401706. doi: 10.1198/004017007000000092.

B. Beachkofski and R. Grandhi. Improved distributed hypercube sampling. In *43rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, page 1274, 2002.

M. H. Beale, M. T. Hagan, and H. B. Demuth. *Neural Network Toolbox ™ User's Guide R2017b*. 2017. ISBN 0971732108.

R. Bellman. Dynamic programming. *Mathematics in Science and Engineering*, 40(P1):101–137, 1967. ISSN 00765392. doi: 10.1016/S0076-5392(08)61063-2.

R. E. Bellman. Dynamic Programming. 1957.

R. E. Bellman and S. E. Dreyfus. Applied dynamic programming. 1962.

A. Ben-Israel and T. N. Greville. *Generalized inverses: theory and applications*. Springer, 2003.

P. Benner, S. Gugercin, and K. Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Review*, 57(4):483–531, 2015. ISSN 00361445. doi: 10.1137/130932715.

P. S. Beran, D. Bryson, A. S. Thelen, M. Diez, and A. Serani. Comparison of multi-fidelity approaches for military vehicle design. In *AIAA AVIATION 2020 FORUM*, page 3158, 2020.

G. Berkooz, P. Holmes, and J. L. Lumley. The proper orthogonal, decomposition in the analysis of turbulent flows. (1971), 1993. ISSN 00664189. doi: 10.1146/annurev.fluid.25.1.539.

B. Betro. A Stochastic Technique for Global Optimization. *Computers and Mathematics with Applications*, 21(6):127–133, 1991.

J. T. Betts. Survey of Numerical Methods for Trajectory Optimization. *Journal of Guidance, Control, and Dynamics*, 21(2):193–207, 1998. ISSN 0731-5090. doi: 10.2514/2.4231.

J. T. Betts. *Practical methods for optimal control and estimation using nonlinear programming*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2010. ISBN 978-0-89871-688-7. doi: https://doi.org/10.1137/1.9780898718577.

J. T. Betts and P. D. Frank. A sparse nonlinear optimization algorithm. *Journal of Optimization Theory and Applications*, 82(3):519–541, 1994. ISSN 00223239. doi: 10.1007/BF02192216.

J. T. Betts and W. P. Huffman. Mesh refinement in direct transcription methods for optimal control. *Optimal Control Applications and Methods*, 19(1):1–21, 1998.

J. T. Betts, N. Biehn, S. L. Campbell, and W. P. Huffman. Compensating for order variation in mesh refinement for direct transcription methods II: Computational experience. *Journal of Computational and Applied Mathematics*, 143(2):237–261, 2002. ISSN 03770427. doi: 10.1016/S0377-0427(01)00509-X.

N. Beume, B. Naujoks, and M. Emmerich. SMS-EMOA: Multiobjective selection based on dominated hypervolume. *European Journal of Operational Research*, 181(3):1653–1669, 2007. ISSN 03772217. doi: 10.1016/j.ejor.2006.08.008.

A. Bhosekar and M. Ierapetritou. Advances in surrogate based modeling, feasibility analysis, and optimization: A review. *Computers and Chemical Engineering*, 108:250–267, 2018. ISSN 00981354. doi: 10.1016/j.compchemeng.2017.09.017. URL https://doi.org/10.1016/j.compchemeng.2017.09.017.

J. Biehler, M. W. Gee, and W. A. Wall. Towards efficient uncertainty quantification in complex and large-scale biomechanical problems based on a Bayesian multi-fidelity scheme. *Biomechanics and Modeling in Mechanobiology*, 14(3):489–513, 2015. ISSN 16177940. doi: 10.1007/s10237-014-0618-0. URL http://dx.doi.org/10.1007/s10237-014-0618-0.

T. Binder, L. Blank, W. Dahmen, and W. Marquardt. Grid refinement in multiscale dynamic optimization. *Computer Aided Chemical Engineering*, 8(C):31–36, 2000. ISSN 15707946. doi: 10.1016/S1570-7946(00)80007-3.

G. A. Bliss. *Lectures on the Calculus of Variations*. Number 504. University of Chicago press, 1963.

C. G. Boender and A. H. Rinnooy Kan. Bayesian stopping rules for multistart global optimization methods. *Mathematical Programming*, 37(1):59–80, 1987. ISSN 14364646. doi: 10.1007/BF02591684.

A. J. Booker, J. E. Dennis, P. D. Frank, D. B. Serafini, V. Torczon, and M. W. Trosset. A rigorous framework for optimization of expensive functions by surrogates. *Structural Optimization*, 17(1):1–13, 1999. ISSN 09344373. doi: 10.1007/BF01197708.

M. Borri, C. Bottasso, and P. Mantegazza. Basic features of the time finite element approach for dynamics. *Meccanica*, 27(2):119–130, 1992. ISSN 1572-9648.

C. L. Bottasso. A new look at finite elements in time: A variational interpretation of Runge-Kutta methods. *Applied Numerical Mathematics*, 25(4):355–368, 1997. ISSN 01689274. doi: 10.1016/S0168-9274(97)00072-X.

M. A. Bouhlel, N. Bartoli, A. Otsmane, and J. Morlier. Improving kriging surrogates of high-dimensional design models by Partial Least Squares dimension reduction. *Structural and Multidisciplinary Optimization*, 53(5):935–952, 2016. ISSN 1615-147X.

G. E. P. Box and N. R. Draper. *Empirical model-building and response surfaces*. John Wiley & Sons, 1987.

G. E. P. Box and K. B. Wilson. On the experimental attainment of optimum conditions. In *Breakthroughs in statistics: methodology and distribution*, pages 270–310. Springer, 1992.

J. H. Bramble, J. E. Pasciak, and J. Xu. Parallel Multilevel Preconditioners. *Mathematics of Computation*, 55(191):1, 1990. ISSN 00255718. doi: 10.2307/2008789.

M. Brameier. On Linear Genetic Programming. *Thesis*, 2004.

A. Brandt. Multi-Level Adaptive Solutions to Boundary-Value Problems. *Mathematics of Computation*, 31(138):333, 1977. ISSN 00255718. doi: 10.2307/2006422.

W. V. Braun. Manned Mars Landing: Presentation to the space task group. 1969.

W. L. Briggs, V. E. Henson, and S. F. Mccormick. A Multigrid Tutorial, 2nd Edition FOSLS/LL* View project Adaptive Algebraic Multigrid Methods View project SEE PROFILE. (January), 2000. URL https://www.researchgate.net/publication/220690328.

C. J. Brooks, A. I. Forrester, A. J. Keane, and S. Shahpar. Multi-fidelity design optimisation of a transonic compressor rotor. *9th European Conference on Turbomachinery: Fluid Dynamics and Thermodynamics, ETC 2011 - Conference Proceedings*, 2(2000):1267–1276, 2011. ISSN 24104833.

D. S. Broomhead and D. Lowe. Radial basis functions, multi-variable functional interpolation and adaptive networks. (4148), 1988.

C. G. Broyden. A Class of Methods for Solving Nonlinear Simultaneous Equations. *Mathematics of Computation*, 19(92):577, 1965. ISSN 00255718. doi: 10.2307/2003941.

C. Bruno and C. Dujarric. In-space nuclear propulsion. *Acta Astronautica*, 82(2):159–165, 2013. ISSN 00945765. doi: 10.1016/j.actaastro.2012.08.022. URL http://dx.doi.org/10.1016/j.actaastro.2012.08.022.

A. E. Bryson. Optimal control-1950 to 1985. *IEEE Control Systems Magazine*, 16(3):26–33, 1996.

D. E. Bryson and A. Earl. *Applied optimal control: optimization, estimation and control*. Routledge, 2018.

D. E. Bryson and M. P. Rumpfkeil. Variable-fidelity surrogate modeling of lambda wing transonic aerodynamic performance. *54th AIAA Aerospace Sciences Meeting*, (January):1–19, 2016a. doi: 10.2514/6.2016-0294.

D. E. Bryson and M. P. Rumpfkeil. Variable-fidelity surrogate modeling of lambda wing transonic aerodynamic performance. In *54th AIAA Aerospace Sciences Meeting*, page 294, 2016b.

D. E. Bryson, R. Haimes, and J. F. Dannenhoffer. Toward the realization of a highly integrated, multidisciplinary, multifidelity design environment. *AIAA Scitech 2019 Forum*, (January): 1–15, 2019. doi: 10.2514/6.2019-2225.

M. D. Buhmann. *Radial basis functions: theory and implementations*, volume 12. Cambridge university press, 2003. ISBN 1139435248.

H. J. Bungartz and M. Griebel. Sparse grids. *Acta Numerica*, 13:147–269, 2004. ISSN 09624929. doi: 10.1017/S0962492904000182.

J. Burkardt, M. Gunzburger, and H. C. Lee. POD and CVT-based reduced-order modeling of Navier-Stokes flows. *Computer Methods in Applied Mechanics and Engineering*, 196(1-3): 337–355, 2006. ISSN 00457825. doi: 10.1016/j.cma.2006.04.004.

D. S. Burke, K. A. De Jong, J. J. Grefenstette, C. L. Ramsey, and A. S. Wu. Putting more genetics into genetic algorithms. *Evolutionary computation*, 6(4):387–410, 1998. ISSN 10636560. doi: 10.1162/evco.1998.6.4.387.

R. H. Byrd, J. Nocedal, and R. A. Waltz. Knitro: An Integrated Package for Nonlinear Optimization. pages 35–59, 2006. doi: 10.1007/0-387-30065-1_4.

R. Calabretta, S. Nolfi, D. Parisi, and G. P. Wagner. Duplication of modules facilitates the evolution of functional specialization. *Artificial Life*, 6(1):69–84, 2000. ISSN 10645462. doi: 10.1162/106454600568320.

R. Caruana. Learning Many Related Tasks at the Same Time With Backpropagation. *NIPS 1994: Proceedings of the 7th International Conference on Neural Information Processing Systems*, pages 657–664, 1994.

P. R. Chai, K. T. McBrayer, A. C. McCrea, R. G. Merrill, A. Prasad, and M. Qu. Crewed mars mission mode options for nuclear electric/chemical hybrid transportation systems.

Accelerating Space Commerce, Exploration, and New Discovery conference, ASCEND 2021, 2021. doi: 10.2514/6.2021-4136.

R. Chai, A. Savvaris, A. Tsourdos, and S. Chai. Multi-objective trajectory optimization of Space Manoeuvre Vehicle using adaptive differential evolution and modified game theory. *Acta Astronautica*, 136(February):273–280, 2017. ISSN 00945765. doi: 10.1016/j.actaastro.2017.02.023. URL http://dx.doi.org/10.1016/j.actaastro.2017.02.023.

R. Chai, A. Savvaris, A. Tsourdos, S. Chai, and Y. Xia. A review of optimization techniques in spacecraft flight trajectory design. *Progress in Aerospace Sciences*, 109(June):100543, 2019. ISSN 03760421. doi: 10.1016/j.paerosci.2019.05.003. URL https://doi.org/10.1016/j.paerosci.2019.05.003.

K. J. Chang, R. T. Haftka, G. L. Giles, and P.-J. Kao. Sensitivity-based scaling for approximating structural response. *Journal of Aircraft*, 30(2):283–288, 1993.

S. Chaturantabut and D. C. Sorensen. Nonlinear model reduction via discrete empirical interpolation. 32(5):2737–2764, 2010.

A. Chaudhuri and R. T. Haftka. A stopping criterion for surrogate based optimization using ego. 72, 2013.

A. Chaudhuri and R. T. Haftka. Efficient global optimization with adaptive target setting. *AIAA Journal*, 52(7):1573–1577, 2014. ISSN 00011452. doi: 10.2514/1.J052930.

R. Chen, J. Xu, S. Zhang, C.-H. Chen, and L. H. Lee. An effective learning procedure for multi-fidelity simulation optimization with ordinal transformation. In *2015 IEEE International Conference on Automation Science and Engineering (CASE)*, pages 702–707. IEEE, 2015. ISBN 1467381837.

V. C. Chen, K. L. Tsui, R. R. Barton, and M. Meckesheimer. A review on design, modeling and applications of computer experiments. *IIE Transactions (Institute of Industrial Engineers)*, 38 (4):273–291, 2006. ISSN 0740817X. doi: 10.1080/07408170500232495.

B. Cheng and D. M. Titterington. Neural Networks : A Review from a Statistical Perspective. *Statistical Sinica*, 9(1):2–54, 1994.

Q. S. Cheng, S. Koziel, and J. W. Bandler. Simplified space-mapping approach to enhancement of microwave device models. *International Journal of RF and Microwave Computer-Aided Engineering*, 16(5):518–535, 2006. ISSN 10964290. doi: 10.1002/mmce.20172.

R. Cheng, Y. Jin, M. Olhofer, and B. Sendhoff. A Reference Vector Guided Evolutionary Algorithm for Many-Objective Optimization. *IEEE Transactions on Evolutionary Computation*, 20(5):773–791, 2016. ISSN 1089778X. doi: 10.1109/TEVC.2016.2519378.

V. Cherkassky and F. M. Mulier. *Learning from data: concepts, theory, and methods*. John Wiley & Sons, 2007.

S. Choi, J. J. Alonso, and I. M. Kroo. Multi-fidelity design optimization studies for supersonic jets using surrogate management frame method. *Collection of Technical Papers - AIAA Applied Aerodynamics Conference*, 2:960–977, 2005. ISSN 10485953. doi: 10.2514/6.2005-5077.

S. Choi, J. J. Alonso, and H. M. Kroo. Two-level multifidelity design optimization studies for supersonic jets. *Journal of Aircraft*, 46(3):776–790, 2009. ISSN 00218669. doi: 10.2514/1.34362.

D. E. Christensen and K. Willcox. *Multifidelity Methods for Multidisciplinary Design Under Uncertainty*. PhD thesis, 2012.

H. S. Chung and J. J. Alonso. Using gradients to construct cokriging approximation models for high-dimensional design optimization problems. *40th AIAA Aerospace Sciences Meeting and Exhibit*, 2002.

S. J. Citron. *Elements of Optimal Control*. Holt, Rinehart and Winston, 1969. ISBN 9780030733703.

D. L. Clark, H. R. Bae, K. Gobal, and R. Penmetsa. Engineering design exploration using locally optimized covariance kriging. *AIAA Journal*, 54(10):3160–3175, 2016. ISSN 00011452. doi: 10.2514/1.J054860.

C. A. Coello Coello and M. S. Lechuga. MOPSO: A proposal for multiple objective particle swarm optimization. *Proceedings of the 2002 Congress on Evolutionary Computation, CEC 2002*, 2:1051–1056, 2002. doi: 10.1109/CEC.2002.1004388.

J. L. Cohon. *Multiobjective programming and planning*, volume 140. Courier Corporation, 2004.

A. R. Conn. *Introduction to Derivative-Free Optimization*, volume 01. 2009.

A. R. Conn, N. I. M. Gould, and P. L. Toint. *Trust region methods*. SIAM, 2000.

A. R. Conn, K. Scheinberg, and I. S. N. Vicente. Downloaded 11 / 24 / 23 to 130 . 159 . 237 . 115 . Redistribution subject to SIAM license or copyright ; see https://epubs.siam.org/terms-privacy Copyright © by SIAM . Unauthorized reproduction of this article is prohibited . Downloaded 11 / 24 / 23 to 13. 20(1):387–415, 2009.

I. Couckuyt, A. Forrester, D. Gorissen, F. De Turck, and T. Dhaene. Blind Kriging: Implementation and performance analysis. *Advances in Engineering Software*, 49:1–13, 2012. ISSN 0965-9978.

V. Coverstone-Carroll, J. W. Hartmann, and W. J. Mason. Optimal multi-objective low-thrust spacecraft trajectories. *Computer Methods in Applied Mechanics and Engineering*, 186(2-4): 387–402, 2000. ISSN 00457825. doi: 10.1016/S0045-7825(99)00393-X.

D. D. Cox, J. S. Park, and C. E. Singer. A statistical method for tuning a computer code to a data base. *Computational Statistics and Data Analysis*, 37(1):77–92, 2001. ISSN 01679473. doi: 10.1016/S0167-9473(00)00057-8.

P. Craven and G. Wahba. Smoothing noisy data with spline functions - Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik*, 31(4):377–403, 1978. ISSN 0029599X. doi: 10.1007/BF01404567.

N. Cressie. *Statistics for spatial data*. John Wiley & Sons, 2015.

K. Crombecq. Surrogate modeling of computer experiments with sequential experimental design. 2011.

K. Crombecq, D. Gorissen, D. Deschrijver, and T. O. M. Dhaene. A Novel Hybrid Sequential Design Strategy for Global Surrogate Modelling of Computer Experiments. 33(4):1948–1974, 2011.

L. Cruz-Piris, I. Marsa-Maestre, and M. A. Lopez-Carmona. A Variable-Length Chromosome Genetic Algorithm to Solve a Road Traffic Coordination Multipath Problem. *IEEE Access*, 7: 111968–111981, 2019. ISSN 21693536. doi: 10.1109/ACCESS.2019.2935041.

C. Currin, T. Mitchell, M. Morris, and D. Ylvisaker. Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association*, 86(416):953–963, 1991. ISSN 1537274X. doi: 10.1080/01621459.1991.10475138.

J. E. Cuthrell and L. T. Biegler. On the optimization of differential-algebraic process systems. *AIChE Journal*, 33(8):1257–1270, 1987. ISSN 15475905. doi: 10.1002/aic.690330804.

J. E. Cuthrell and L. T. Biegler. Simultaneous optimization and solution methods for batch reactor control profiles. *Computers and Chemical Engineering*, 13(1-2):49–62, 1989. ISSN 00981354. doi: 10.1016/0098-1354(89)89006-4.

A. Da Ronch, M. Panzeri, M. A. A. Bari, R. D'Ippolito, and M. Franciolini. Adaptive design of experiments for efficient and accurate estimation of aerodynamic loads. *Aircraft Engineering and Aerospace Technology*, 89(4):558–569, 2017. ISSN 00022667. doi: 10.1108/AEAT-10-2016-0173.

W. Dahmen and A. Kunoth. Multilevel preconditioning. *Numerische Mathematik*, 63(1):315–344, 1992. ISSN 0029599X. doi: 10.1007/BF01385864.

G. B. Dantzig, A. Orden, and P. Wolfe. The generalized simplex method for minimizing a linear form under linear inequality restraints. *Pacific Journal of Mathematics*, 5(2):183–195, 1955. ISSN 00308730. doi: 10.2140/pjm.1955.5.183.

C. L. Darby, W. W. Hager, and A. V. Rao. An hp-adaptive pseudospectral method for solving optimal control problems. *Optimal Control Applications and Methods*, 32(4):476–502, 2011. ISSN 01432087. doi: 10.1002/oca.957.

I. Das and J. Dennis. Normal-boundary intersection: A new method for generating the pareto surface in nonlinear multicriteria optimization problems. *SIAM Journal of Optimisation*, 8 (3):631–657, 1998.

I. Daubechies. Ten Lectures on Wavelets - Ingrid Daubechies (1992).pdf, 1998.

W. C. Davidon. Variable Metric Method for Minimization. Technical report, 1959.

S. De, J. Britton, M. Reynolds, R. Skinner, K. Jansen, and A. Doostan. On transfer learning of neural networks using bi-fidelity data for uncertainty propagation. *International Journal for Uncertainty Quantification*, 10(6), 2020.

V. A. de Oliveira and G. N. Silva. On sufficient optimality conditions for multiobjective control problems. *Journal of Global Optimization*, 64(4):721–744, 2016. ISSN 15732916. doi: 10.1007/s10898-015-0351-y.

O. L. de Weck. Multiobjective Optimization: History and Promise. In *Invited Keynote Paper, GL2-2, The Third China-Japan-Korea Joint Symposium on Optimization of Structural and Mechanical Systems, Kanazawa, Japan*, 2004.

K. Deb and H. Jain. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, Part I: Solving problems with box constraints. *IEEE Transactions on Evolutionary Computation*, 18(4):577–601, 2014. ISSN 1089778X. doi: 10.1109/TEVC.2013.2281535.

K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002. ISSN 1089778X. doi: 10.1109/4235.996017.

J. Decraene, M. Chandramohan, F. Zeng, M. Yoke Hean Low, and W. Cai. Evolving Agent-Based Model Structures using Variable-Length Genomes. In *Proceedings of the Fourth International workshop on Optimisation in Multi-Agent Systems*, 2011. URL http://pdcc.ntu.edu.sg/EVOSIM/.

J. E. Dennis Jr and R. B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*. SIAM, 1996.

D. Deschrijver, K. Crombecq, H. M. Nguyen, and T. Dhaene. Adaptive sampling algorithm for macromodeling of parameterized S-parameter responses. *IEEE Transactions on Microwave Theory and Techniques*, 59(1):39–45, 2011. ISSN 00189480. doi: 10.1109/TMTT.2010.2090407.

D. Ding, H. Guo, L. Huang, and Y. Wang. Multi-objective optimization of reentry trajectory for Hypersonic Gliding Vehicle. *CGNCC 2016 - 2016 IEEE Chinese Guidance, Navigation and Control Conference*, pages 1436–1443, 2017. doi: 10.1109/CGNCC.2016.7829000.

B. G. Drake. *Human Exploration of Mars Design Reference Architecture 5 . 0 Addendum*. Number July. 2009. ISBN 3016210134.

Q. Du, V. Faber, and M. Gunzburger. Centroidal Voronoi tessellations: Applications and algorithms. *SIAM Review*, 41(4):637–676, 1999. ISSN 00361445. doi: 10.1137/S0036144599352836.

J. Duchon. Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In *Constructive Theory of Functions of Several Variables: Proceedings of a Conference Held at Oberwolfach April 25–May 1, 1976*, pages 85–100. Springer, 1977. ISBN 3540080694.

C. Dujarric. A new propulsion concept for interplanetary missions. *ESA Bulletin*, 108 (november):66–71, 2001.

C. Dujarric, A. Santovincenzo, and L. Summerer. The Nuclear Thermal Electric Rocket: A Proposed Innovative Propulsion Concept for Manned Interplanetary Missions. *Progress in Prpulsion Physics*, 4:293–312, 2013. doi: 10.1051/eucass/201304.

M. Durante. Space radiation protection: Destination Mars. *Life Sciences in Space Research*, 1(1): 2–9, 2014. ISSN 22145532. doi: 10.1016/j.lssr.2014.01.002. URL http://dx.doi.org/10.1016/j.lssr.2014.01.002.

M. Durante and C. Bruno. Impact of rocket propulsion technology on the radiation risk in missions to Mars. *European Physical Journal D*, 60(1):215–218, 2010. ISSN 14346060. doi: 10.1140/epjd/e2010-00035-6.

M. Durante and F. A. Cucinotta. Physical basis of radiation protection in space travel. *Reviews of Modern Physics*, 83(4), 2011. ISSN 00346861. doi: 10.1103/RevModPhys.83.1245.

J. Eason and S. Cremaschi. Adaptive sequential sampling for surrogate model generation with artificial neural networks. *Computers and Chemical Engineering*, 68:220–232, 2014. ISSN 00981354. doi: 10.1016/j.compchemeng.2014.05.021. URL http://dx.doi.org/10.1016/j.compchemeng.2014.05.021.

D. Echeverría and P. W. Hemker. Space Mapping and Defect Correction. *Computational Methods in Applied Mathematics*, 5(2):107–136, 2005. ISSN 16099389. doi: 10.2478/cmam-2005-0006.

D. Echeverría and P. W. Hemker. Manifold mapping: A two-level optimization technique. *Computing and Visualization in Science*, 11(4-6):193–206, 2008. ISSN 14329360. doi: 10.1007/s00791-008-0096-y.

F. Y. Edgeworth. *Mathematical psychics: An essay on the application of mathematics to the moral sciences*, volume 10. CK Paul, 1881.

G. Eichfelder. An Adaptive Scalarization Method in Multiobjective Optimisation. *SIAM Journal of Optimisation*, 19(4):1694–1718, 2009.

M. S. Eldred, A. A. Giunta, and S. S. Collis. Second-order corrections for surrogate-based optimization with model hierarchies. *Collection of Technical Papers - 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, 3(October):1754–1768, 2004. doi: 10.2514/6.2004-4457.

J. A. Englander, M. A. Vavrina, and A. R. Ghosh. Aas 15-227 Multi-Objective Hybrid Optimal Control for Multiple-Flyby Low-Thrust Mission Design. pages 1–20, 2019.

P. J. Enright and B. A. Conway. Discrete approximations to optimal trajectories using direct transcription and nonlinear programming. *Journal of Guidance, Control, and Dynamics*, 15 (4):994–1002, 1992. ISSN 07315090. doi: 10.2514/3.20934.

F. Fahroo and I. M. Ross. A spectral patching method for direct trajectory optimization. *The Journal of the Astronautical Sciences*, 48:269–286, 2000.

F. Fahroo and I. M. Ross. Convergence of the costates does not imply convergence of the control. *Journal of Guidance, Control, and Dynamics*, 31(5):1492–1496, 2008. ISSN 15333884. doi: 10.2514/1.37331.

A. Falchi, E. Minisci, M. Vasile, D. Rastelli, and N. Bellini. DSMC-based Correction Factor for Low-fidelity Hypersonic Aerodynamics of Re-entering Objects and Space Debris. *7th European Conference of Aeronautics and Aerospace Sciences*, 10(July), 2017a.

A. Falchi, M. Vasile, A. Falchi, V. Renato, E. Minisci, and M. Vasile. FOSTRAD : An Advanced Open Source Tool for Re-entry Analysis. *Reinventing Space Conference*, (October):1–15, 2017b.

A. Farhang-Mehr and S. Azarm. Bayesian meta-modeling of engineering design simulations: A sequential approach with adaptation to irregularities in the response behaviour. *International Journal for Numerical Methods in Engineering*, 62(15):2104–2126, 2005. ISSN 00295981. doi: 10.1002/nme.1261.

P. Feldmann and R. W. Freund. Efficient linear circuit analysis by Pade approximation via the Lanczos process, 1994.

G. Fernandez, C. Park, N. H. Kim, and R. Haftka. Review of multi-fidelity models. (March), 2017.

M. Fernández-Godino, C. Park, N. H. Kim, and R. T. Haftka. Issues in deciding whether to use multifidelity surrogates. *AIAA Journal*, 57(5):2039–2054, 2019. ISSN 00011452. doi: 10.2514/1.J057750.

M. G. Fernández-Godino, C. Park, N.-H. Kim, and R. T. Haftka. Review of multi-fidelity models. 2016. URL http://arxiv.org/abs/1609.07196.

S. Ficini, U. Iemma, R. Pellegrini, A. Serani, and M. Diez. Assessing the Performance of an Adaptive Multi-Fidelity Gaussian Process with Noisy Training Data: A Statistical Analysis. *AIAA Aviation and Aeronautics Forum and Exposition, AIAA AVIATION Forum 2021*, pages 1–12, 2021. doi: 10.2514/6.2021-3098.

J. Finseth and NASA. Overview of ROVER Engine Tests: Final Report. *Discovery*, (3), 1991.

C. C. Fischer, R. V. Grandhi, and P. S. Beran. Bayesian-enhanced low-fidelity correction approach to multifidelity aerospace design. *AIAA Journal*, 56(8):3295–3306, 2018. ISSN 00011452. doi: 10.2514/1.J056529.

R. Fisher. Arrangment of Field Experiments. *Journal of the Ministry of Agriculture*, (33):503–513, 1926.

R. Fletcher. *Practical methods of optimization*. John Wiley & Sons, 2013.

C. Fleury and V. Braibant. Structural optimization: A new dual method using mixed variables. *International Journal for Numerical Methods in Engineering*, 23(3):409–428, 1986. ISSN 10970207. doi: 10.1002/nme.1620230307.

C. M. Fonseca, L. Paquete, and M. López-Ibáñez. An improved dimension-sweep algorithm for the hypervolume indicator. *2006 IEEE Congress on Evolutionary Computation, CEC 2006*, pages 1157–1163, 2006. doi: 10.1109/cec.2006.1688440.

F. D. Foresee and M. T. Hagan. Gauss-Newton Approximation to Bayesian Learning. *School of Electrical and Computer Engineering*, pages 1930–1935, 1997.

A. Forrester, A. Sobester, and A. Keane. *Engineering design via surrogate modelling: a practical guide*. John Wiley & Sons, 2008. ISBN 0470770791.

A. I. Forrester. Black-box calibration for complex-system simulation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1924):3567–3579, 2010. ISSN 1364503X. doi: 10.1098/rsta.2010.0051.

A. I. Forrester and A. J. Keane. Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences*, 45(1-3):50–79, 2009. ISSN 03760421. doi: 10.1016/j.paerosci.2008.11.001.

A. I. Forrester, N. W. Bressloff, and A. J. Keane. Optimization using surrogate models and partially converged computational fluid dynamics simulations. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 462(2071):2177–2204, 2006. ISSN 14712946. doi: 10.1098/rspa.2006.1679.

A. I. Forrester, A. Sóbester, and A. J. Keane. Multi-fidelity optimization via surrogate modelling. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 463 (2088):3251–3269, 2007. ISSN 14712946. doi: 10.1098/rspa.2007.1900.

Y. Freund, H. S. Seung, E. Shamir, and N. Tishby. Selective Sampling Using the Query by Committee Algorithm. *Machine Learning*, 28(2-3):133–168, 1997. ISSN 08856125. doi: 10.1023/A:1007330508534.

N. Frischauf and B. A. Hamilton. Space nuclear power and propulsion - A basic tool for the manned exploration of the solar system. *Proceedings of the 2004 International Congress on Advances in Nuclear Power Plants, ICAPP'04*, (July 2004):2210–2216, 2004.

A. Gambier and E. Badreddin. Multi-objective Optimal Control: An Overview. (October):1–3, 2007.

S. E. Gano, J. E. Renaud, and B. Sanders. Hybrid variable fidelity optimization by using a kriging-based scaling function. *AIAA Journal*, 43(11):2422–2430, 2005. ISSN 00011452. doi: 10.2514/1.12466.

C. Geuzaine and J.-F. Remacle. Gmsh: A 3-D finite element mesh generator with built-in pre- and post-processing facilities. *International Journal for Numerical Methods in Engineering*, 79: 1309–1331, 2009. doi: 10.1002/nme. URL http://onlinelibrary.wiley.com/doi/10.1002/nme.3279/full.

M. Ghoreyshi, K. J. Badcock, and M. A. Woodgate. Accelerating the numerical generation of aerodynamic models for flight simulation. *Journal of Aircraft*, 46(3):972–980, 2009. ISSN 15333868. doi: 10.2514/1.39626.

P. E. Gill, W. Murray, and M. A. Saunders. SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM Review*, 47(1):99–131, 2005. ISSN 00361445. doi: 10.1137/S0036144504446096.

A. A. Giunta, R. Narducci, S. Burgee, B. Grossman, W. H. Mason, L. T. Watson, and R. T. Haftka. Variable-complexity response surface aerodynamic design of an HSCT wing. *13th Applied Aerodynamics Conference*, pages 994–1002, 1995. doi: 10.2514/6.1995-1886.

A. A. Giunta, S. F. Wojtkiewicz, and M. S. Eldred. Overview of modern design of experiments methods for computational simulations. *41st Aerospace Sciences Meeting and Exhibit*, (January), 2003. doi: 10.2514/6.2003-649.

F. Glover and M. Laguna. Tabu Search. *Handbook of Combinatorial Optimization*, pages 2093–2229, 1998. doi: 10.1007/978-1-4613-0303-9_33.

T. Goel, R. T. Haftka, W. Shyy, and N. V. Queipo. Ensemble of surrogates. *Structural and Multidisciplinary Optimization*, 33(3):199–216, 2007. ISSN 1615147X. doi: 10.1007/s00158-006-0051-9.

D. Goldberg. Messy Genetic Algorithms : Motivation, Analysis, and First Results. *Complex Systems*, 3(5):493–530, 1989. ISSN 0891-2513. URL http://www.complex-systems.com/abstracts/v03_i05_a05.html.

H. H. Goldstine. A history of the calculus of variations from the 17th through the 19th century. 1980.

J. Golinski. An adaptive optimization system applied to machine synthesis. *Mechanism and Machine Theory*, 8(4):419–436, 1973. ISSN 0094114X. doi: 10.1016/0094-114X(73)90018-9.

Q. Gong and I. M. Ross. Autonomous pseudospectral knotting methods for space mission optimization. In *16th AAS/AIAA Space Flight Mechanics Meeting, American Astronautical Society Paper*, pages 06–151, 2006.

D. Gorissen, I. Couckuyt, P. Demeester, T. Dhaene, and K. Crombecq. A surrogate modeling and adaptive sampling toolbox for computer based design. *Journal of Machine Learning Research*, 11:2051–2055, 2010. ISSN 15324435.

M. Grabe, S. Fasoulas, and K. Hannemann. Minimal Induced Drag for Non-planar Lifting Surfaces with Moderate and Small Aspect Ratio Zusammenfassung Fulltext Preview Abstract - SpringerLink. *Notes on Numerical Fluid Mechanics*, (May):84–85, 2012. doi: 10.1007/978-3-642-14243-7.

R. B. Gramacy and H. K. Lee. Bayesian treed Gaussian process models with an application to computer modeling. *Journal of the American Statistical Association*, 103(483):1119–1130, 2008. ISSN 01621459. doi: 10.1198/016214508000000689.

R. B. Gramacy and H. K. Lee. Adaptive design and analysis of supercomputer experiments. *Technometrics*, 51(2):130–145, 2009. ISSN 00401706. doi: 10.1198/TECH.2009.0015.

L. L. Gratiet and J. Garnier. Recursive co-kriging model for design of computer experiments with multiple levels of fidelity. *International Journal for Uncertainty Quantification*, 4(5):365–386, 2014. ISSN 21525099. doi: 10.1615/int.j.uncertaintyquantification.2014006914.

J. Guenther, H. K. Lee, and G. A. Gray. Sequential design for achieving estimated accuracy of global sensitivities. *Applied Stochastic Models in Business and Industry*, 31(6):782–800, 2015. ISSN 15264025. doi: 10.1002/asmb.2091.

S. Gunn. *Support Vector Machines for classification and regression.* PhD thesis, 1998.

M. Guo, A. Manzoni, M. Amendt, P. Conti, and J. S. Hesthaven. Multi-fidelity regression using artificial neural networks: Efficient approximation of parameter-dependent output quantities. *Computer Methods in Applied Mechanics and Engineering*, 389:114378, 2022. ISSN 00457825. doi: 10.1016/j.cma.2021.114378. URL https://doi.org/10.1016/j.cma.2021.114378.

H. M. Gutmann. A Radial Basis Function Method for Global Optimization. *Journal of Global Optimization*, 19(3):201–227, 2001. ISSN 09255001. doi: 10.1023/A:1011255519438.

R. T. Haftka. Optimization of flexible wing structures subject to strength and induced drag constraints. *AIAA Journal*, 15(8):1101–1106, 1977. ISSN 00011452. doi: 10.2514/3.7400.

R. T. Haftka. Combining global and local approximations. *AIAA Journal*, 29(9), 1991.

R. T. Haftka, J. A. Nachlas, L. T. Watson, T. Rizzo, and R. Desai. Two-point constraint approximation in structural optimization, 1987. ISSN 00457825.

R. T. Haftka, E. P. Scott, and J. R. Cruz. Optimization and experiments: a survey. 1998.

P. Hajela and L. Berke. Neural networks in structural analysis and design: An overview. *Computing Systems in Engineering*, 3(1-4):525–538, 1992. ISSN 09560521. doi: 10.1016/0956-0521(92)90138-9.

Z. H. Han, S. Görtz, and R. Hain. A Variable-Fidelity Modeling Method for Aero-Loads Prediction. *New results in numerical and experimental fluid mechanics*, (7):17–25, 2010.

Z. H. Han, S. Görtz, and R. Zimmermann. Improving variable-fidelity surrogate modeling via gradient-enhanced kriging and a generalized hybrid bridge function. *Aerospace Science and Technology*, 25(1):177–189, 2013. ISSN 12709638. doi: 10.1016/j.ast.2012.01.006. URL http://dx.doi.org/10.1016/j.ast.2012.01.006.

A. Harten. Multiresolution representation of data: A general framework. *SIAM Journal on Numerical Analysis*, 33(3):1205–1256, 1996. URL https://www.jstor.org/stable/2158503.

J. K. Hartman. Some experiments in global optimization. *Naval Research Logistics Quarterly*, 20(3):569–576, 1973. ISSN 00281441. doi: 10.1002/nav.3800200316.

I. Harvey. The SAGA Cross: The Mechanics of Recombination for Species with Variable-Length Genotypes. *Parallel Problem Solving from Nature*, 2:269–278, 1992.

T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

G. Heath. Mathworks Community Profile - Greg Heath, 2020. URL https://uk.mathworks.com/matlabcentral/profile/authors/2929937-greg-heath.

D. O. Hebb. The Organization of Behavior; A Neuropsychological Theory. *The American Journal of Psychology*, 63(4):633, 1949. ISSN 00029556. doi: 10.2307/1418888.

P. W. Hemker and D. Echeverría. A trust-region strategy for manifold-mapping optimization. *Journal of Computational Physics*, 224(1):464–475, 2007. ISSN 10902716. doi: 10.1016/j.jcp.2007.04.003.

W. Hendrickx and T. Dhaene. Sequential design and rational metamodelling. *Proceedings - Winter Simulation Conference*, 2005(June):290–298, 2005. ISSN 08917736. doi: 10.1109/WSC.2005.1574263.

A. L. Herman and B. A. Conway. Direct optimization using collocation based on high-order Gauss-Lobatto quadrature rules. *Journal of Guidance, Control, and Dynamics*, 19(3):592–599, 1996. ISSN 15333884. doi: 10.2514/3.21662.

N. Hitomi and D. Selva. Constellation optimization using an evolutionary algorithm with a variable-length chromosome. *IEEE Aerospace Conference Proceedings*, 2018-March:1–12, 2018. ISSN 1095323X. doi: 10.1109/AERO.2018.8396743.

Y. Hochberg and A. C. Tamhane. *Multiple comparison procedures*. John Wiley & Sons, Inc., 1987.

D. H. Hodges and R. R. Bless. Weak hamiltonian finite element method for optimal control problems. *Journal of Guidance, Control, and Dynamics*, 14(1):148–156, 1991. ISSN 07315090. doi: 10.2514/3.20616.

C. M. Holden. *Visualization methodologies in aircraft design*. PhD thesis, 2004.

J. H. Holland. Genetic Algorithms. *Scientific American*, 267(1):66–73, jun 1992. ISSN 00368733, 19467087. URL http://www.jstor.org/stable/24939139.

J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982. ISSN 0027-8424.

A. A. Howard, M. Perego, G. E. Karniadakis, and P. Stinis. Multifidelity deep operator networks for data-driven and physics-informed problems. *Journal of Computational Physics*, 493: 112462, 2023.

D. Huang, T. T. Allen, W. I. Notz, and R. A. Miller. Sequential kriging optimization using multiple-fidelity evaluations. *Structural and Multidisciplinary Optimization*, 32(5):369–382, 2006. ISSN 1615147X. doi: 10.1007/s00158-005-0587-0.

L. Huang, Z. Gao, and D. Zhang. Research on multi-fidelity aerodynamic optimization methods. *Chinese Journal of Aeronautics*, 26(2):279–286, 2013. ISSN 10009361. doi: 10.1016/j.cja.2013.02.004. URL http://dx.doi.org/10.1016/j.cja.2013.02.004.

D. G. Hull. Conversion of optimal control problems into parameter optimization problems. *Journal of guidance, control, and dynamics*, 20(1):57–60, 1997.

B. Hutt and K. Warwick. Synapsing variable-length crossover: Meaningful crossover for variable-length genomes. *IEEE Transactions on Evolutionary Computation*, 11(1):118–131, 2007. ISSN 1089778X. doi: 10.1109/TEVC.2006.878096.

S. Jain and P. Tsiotras. Trajectory optimization using multiresolution techniques. *Journal of Guidance, Control, and Dynamics*, 31(5):1424–1436, 2008. ISSN 15333884. doi: 10.2514/1.32220.

M. J. Jansen. Analysis of variance designs for model output. *Computer Physics Communications*, 117(1):35–43, 1999. ISSN 00104655. doi: 10.1016/S0010-4655(98)00154-4.

P. Jiang, Q. Zhou, and X. Shao. *Surrogate-Model-Based Design and Optimization*. 2020. ISBN 9789811507304. doi: 10.1007/978-981-15-0731-1_7.

R. Jin, W. Chen, and T. W. Simpson. Comparative studies of metamodelling techniques under multiple modelling criteria. *Structural and Multidisciplinary Optimization*, 23(1):1–13, 2001. ISSN 1615147X. doi: 10.1007/s00158-001-0160-4.

D. R. Jones. A Taxonomy of Global Optimization Methods Based on Response Surfaces. *Journal of Global Optimization*, 21(4):345–383, 2001. ISSN 09255001. doi: 10.1023/A:1012771025575.

D. R. Jones, M. Schonlau, and W. J. Welch. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization*, 13:455–492, 1998.

I. M. Jonsson, L. Leifsson, S. Koziel, Y. A. Tesfahunegn, and A. Bekasiewicz. Shape optimization of trawl-doors using variable-fidelity models and space mapping. *Procedia Computer Science*, 51(1):905–913, 2015. ISSN 18770509. doi: 10.1016/j.procs.2015.05.223.

V. R. Joseph, Y. Hung, and A. Sudjianto. Blind kriging: A new method for developing metamodels. *Journal of mechanical design*, 130(3), 2008. ISSN 1050-0472.

C. R. Joyner, M. Eades, J. Horton, T. Jennings, T. Kokan, D. J. Levack, B. J. Muzek, and C. B. Reynolds. LEU NTP Engine System Trades and Mission Options. *Nuclear Technology*, 206(8):1140–1154, 2020. ISSN 19437471. doi: 10.1080/00295450.2019.1706982. URL https://doi.org/10.1080/00295450.2019.1706982.

S. Kameswaran and L. T. Biegler. Convergence rates for direct transcription of optimal control problems using collocation at Radau points. *Computational Optimization and Applications*, 41(1):81–126, 2008. ISSN 09266003. doi: 10.1007/s10589-007-9098-9.

W. Karush. Minima of functions of several variables with inequalities as side constraints. *M. Sc. Dissertation. Dept. of Mathematics, Univ. of Chicago*, 1939.

M. Kaufman, V. Balabanov, S. L. Burgee, A. A. Giunta, B. Grossman, R. T. Haftka, W. H. Mason, and L. T. Watson. Variable-complexity response surface approximations for wing structural weight in HSCT design. *Computational Mechanics*, 18(2):112–126, 1996. ISSN 01787675. doi: 10.1007/BF00350530.

C. Y. Kaya and H. Maurer. A numerical method for nonconvex multi-objective optimal control problems. *Computational Optimization and Applications*, 57(3):685–702, 2014. ISSN 15732894. doi: 10.1007/s10589-013-9603-2.

A. Keane and P. Nair. *Computational approaches for aerospace design: the pursuit of excellence*. John Wiley & Sons, 2005. ISBN 0470855479.

A. J. Keane. Wing optimization using design of experiment, response surface, and data fusion methods. *Journal of Aircraft*, 40(4):741–750, 2003. ISSN 0021-8669.

A. J. Keane. Statistical improvement criteria for use in multiobjective design optimization. *AIAA Journal*, 44(4):879–891, 2006. ISSN 00011452. doi: 10.2514/1.16875.

A. J. Keane. Cokriging for robust design optimization. *AIAA Journal*, 50(11):2351–2364, 2012. ISSN 00011452. doi: 10.2514/1.J051391.

R. L. Keeney and H. Raiffa. *Decisions with multiple objectives: preferences and value trade-offs*. Cambridge university press, 1993.

H. B. Keller. *Numerical methods for two-point boundary-value problems*. Courier Dover Publications, 1968.

J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95-international conference on neural networks*, volume 4, pages 1942–1948. IEEE, 1995.

M. C. Kennedy and A. O'Hagan. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87(1):1–13, 2000. ISSN 00063444. doi: 10.1093/biomet/87.1.1.

M. C. Kennedy and A. O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001a. ISSN 1369-7412. doi: 10.1111/1467-9868.00294.

M. C. Kennedy and A. O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001b. ISSN 1369-7412. doi: 10.1111/1467-9868.00294.

M. C. Kennedy, A. O'Hagan, and N. Higgins. Bayesian Analysis of Computer Code Outputs. *Quantitative Methods for Current Environmental Issues*, pages 227–243, 2002. doi: 10.1007/978-1-4471-0657-9_11.

B. T. Kien, N. C. Wong, and J. C. Yao. Necessary conditions for multiobjective optimal control problems with state constraints. *Dynamics of Continuous, Discrete and Impulsive Systems Series B: Applications and Algorithms*, 19(3):431–446, 2012. ISSN 14928760.

B. Kim, Y. Lee, and D. H. Choi. Construction of the radial basis function based on a sequential sampling approach using cross-validation. *Journal of Mechanical Science and Technology*, 23 (12):3357–3365, 2009. ISSN 1738494X. doi: 10.1007/s12206-009-1014-z.

J. B. Kim and D. Wallace. A goal-oriented design evaluation model. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 80456, page V003T30A014. American Society of Mechanical Engineers, 1997.

S. Kirkpatrick, C. D. Gelatt Jr, and M. P. Vecchi. Optimization by simulated annealing. *science*, 220(4598):671–680, 1983. ISSN 0036-8075.

J. P. Kleijnen and W. C. Van Beers. Application-driven sequential designs for simulation experiments: Kriging metamodelling. *Journal of the Operational Research Society*, 55(8): 876–883, 2004. ISSN 01605682. doi: 10.1057/palgrave.jors.2601747.

D. L. Knill, A. A. Giunta, C. A. Baker, B. Grossman, W. H. Mason, R. T. Haftka, and L. T. Watson. HSCT configuration design using response surface approximations of supersonic euler aerodynamics. *36th AIAA Aerospace Sciences Meeting and Exhibit*, (January 1998), 1998. doi: 10.2514/6.1998-905.

P. N. Koch, T. W. Simpson, J. K. Allen, and F. Mistree. Statistical approximations for multidisciplinary design optimization: The problem of size. *Journal of Aircraft*, 36(1):275–286, 1999. ISSN 15333868. doi: 10.2514/2.2435.

P. N. Koch, R. J. Yang, and L. Gu. Design for six sigma through robust optimization. *Structural and Multidisciplinary Optimization*, 26(3-4):235–248, 2004. ISSN 1615147X. doi: 10.1007/s00158-003-0337-0.

D. Koenig. Experience Gained from the Space Nuclear Rocket Program (Rover). Technical report, 1986.

T. Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59–69, 1982. ISSN 03401200. doi: 10.1007/BF00337288.

T. G. Kolda, R. M. Lewis, and V. Torczon. Optimization by direct search: New perspectives on some classical and modern methods. *SIAM Review*, 45(3):385–482, 2003. ISSN 00361445. doi: 10.1137/S003614450242889.

A. N. Kolmogorov. On the representation of continuous functions of several variables as superpositions of continuous functions of one variable and addition. *Doklady Akademii. Nauk USSR*, 114:679–681, 1957. doi: 10.18411/lj-12-2018-148.

A. S. Koroteev, V. N. Akimov, L. E. Zakharenkov, A. V. Karevskiy, E. Y. Kuvshinova, A. V. Semenkin, and A. E. Solodukhin. Architecture Features and Application Aspects of High-Power Spacecrafts With Electric Propulsion. (May), 2018.

R. Kosonen and J. Shemeikka. The use of a simple simulation tool for energy analysis. *Proceeding of Building Simulation, 5h Conference of International Building Performance Simulation Association, Prague, Czech Republic, 8-10 September*, pages 103–110, 1997.

J. R. Koza. The genetic programming paradigm: Genetically breeding populations of computer programs to solve problems. *Dynamic, Genetic, and Chaotic Programming*, (June):203–321, 1992. doi: doi:10.1109/TAI.1990.130444.

S. Koziel. Efficient optimization of microwave circuits using shape-preserving response prediction. *IEEE MTT-S International Microwave Symposium Digest*, pages 1569–1572, 2009. ISSN 0149645X. doi: 10.1109/MWSYM.2009.5166010.

S. Koziel and J. W. Bandler. Space-mapping-based modeling utilizing parameter extraction with variable weight coefficients and a data base. *IEEE MTT-S International Microwave Symposium Digest*, 1(2):1763–1766, 2006. ISSN 0149645X. doi: 10.1109/MWSYM.2006.249380.

S. Koziel and J. W. Bandler. Microwave device modeling using space-mapping and radial basis functions. *IEEE MTT-S International Microwave Symposium Digest*, 3:799–802, 2007a. ISSN 0149645X. doi: 10.1109/MWSYM.2007.380079.

S. Koziel and J. W. Bandler. A Space-Mapping Approach to Microwave Device Modeling Exploiting Fuzzy Systems. *IEEE Transactions on Microwave Theory and Techniques*, 55(12): 2539–2547, 2007b. ISSN 15579670. doi: 10.1109/TMTT.2007.909605.

S. Koziel and L. Leifsson. Shape-preserving response prediction for engineering design optimization. *IEEE Transactions on Microwave Theory and Techniques*, 58(11), 2010. ISSN 18770509. doi: 10.1016/j.procs.2013.05.253.

S. Koziel and L. Leifsson. Knowledge-based airfoil shape optimization using space mapping. In *30th AIAA applied aerodynamics conference*, page 3016, 2012.

S. Koziel and L. Leifsson. Multi-level CFD-based airfoil shape optimization with automated low-fidelity model selection. *Procedia Computer Science*, 18:889–898, 2013a. ISSN 18770509. doi: 10.1016/j.procs.2013.05.254. URL http://dx.doi.org/10.1016/j.procs.2013.05.254.

S. Koziel and L. Leifsson. Shape-preserving response prediction for engineering design optimization. *Procedia Computer Science*, 18(December):879–888, 2013b. ISSN 18770509. doi: 10.1016/j.procs.2013.05.253. URL http://dx.doi.org/10.1016/j.procs.2013.05.253.

S. Koziel and L. Leifsson. Shape-Preserving Response Prediction for Surrogate Modeling and Engineering Design Optimization. *Springer Proceedings in Mathematics and Statistics*, 97: 241–253, 2014.

S. Koziel and L. Leifsson. *Simulation-driven design by knowledge-based response correction techniques*. 2016. ISBN 9783319301150. doi: 10.1007/978-3-319-30115-0.

S. Koziel and X.-S. Yang. *Computational Optimization, Methods and Algorithms*. 2011. ISBN 9783642175565.

S. Koziel, J. W. Bandler, and K. Madsen. A space-mapping framework for engineering optimization - Theory and implementation. *IEEE Transactions on Microwave Theory and Techniques*, 54(10):3721–3730, 2006a. ISSN 00189480. doi: 10.1109/TMTT.2006.882894.

S. Koziel, J. W. Bandler, and K. Madsen. A space-mapping framework for engineering optimization - Theory and implementation. *IEEE Transactions on Microwave Theory and Techniques*, 54(10):3721–3730, 2006b. ISSN 00189480. doi: 10.1109/TMTT.2006.882894.

S. Koziel, J. W. Bandler, and K. Madsen. Theoretical justification of space-mapping-based modeling utilizing a database and on-demand parameter extraction. *IEEE Transactions on Microwave Theory and Techniques*, 54(12):4316–4321, 2006c. ISSN 00189480. doi: 10.1109/TMTT.2006.884648.

S. Koziel, Q. Cheng, and J. W. Bandler. Space mapping. *IEEE Microwave Magazine*, (December): 105–122, 2008.

S. Koziel, J. W. Bandler, and K. Madsen. Space mapping with adaptive response correction for microwave design optimization. *Society*, 57(2):478–486, 2009.

S. Koziel, Q. S. Cheng, and J. W. Bandler. Implicit space mapping with adaptive selection of preassigned parameters. *IET Microwaves, Antennas and Propagation*, 4(3):361–373, 2010. ISSN 17518725. doi: 10.1049/iet-map.2009.0080.

S. Koziel, S. Ogurtsov, and S. Szczepanski. Rapid antenna design optimization using shape-preserving response prediction. *Bulletin of the Polish Academy of Sciences: Technical Sciences*, 60(1):143–149, 2012. ISSN 02397528. doi: 10.2478/v10175-012-0020-y.

A. Krogh and J. Vedelsby. Neural Network Ensembles, Cross Validation, and Active Learning Anders. pages 6–7, 1995.

H. W. Kuhn and A. W. Tucker. Nonlinear programming 2nd Berkeley Symposium on Mathematical Statistics and Probability (Berkeley, CA, 1951.

H. Kwakernaak, R. Sivan, and B. N. D. Tyreus. Linear optimal control systems. 1974.

C. Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operator, 1950.

D. F. Landau and J. M. Longuski. A reassessment of trajectory options for human missions to Mars. *Collection of Technical Papers - AIAA/AAS Astrodynamics Specialist Conference*, 2: 836–871, 2004. doi: 10.2514/6.2004-5095.

D. F. Landau and J. M. Longuski. Trajectories for Human Missions to Mars, Part II: Low-Thrust Transfers. *Journal of Spacecraft and Rockets*, 43(5):1043–1048, 2006. ISSN 0022-4650. doi: 10.2514/1.a12253b.

W. B. Langdon and R. Poli. Fitness causes bloat: Mutation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 1391(July 2006):37–48, 1998. ISSN 16113349. doi: 10.1007/BFb0055926.

D. S. Lazzara, R. Haimes, and K. Willcox. Multifidelity geometry and analysis in aircraft conceptual design. *19th AIAA Computational Fluid Dynamics Conference*, (June):1–22, 2009.

L. Le Gratiet. *Multi-fidelity Gaussian process regression for computer experiments*. PhD thesis, 2013.

L. Le Gratiet and C. Cannamela. Cokriging-Based Sequential Design Strategies Using Fast Cross-Validation Techniques for Multi-Fidelity Computer Codes. *Technometrics*, 57(3): 418–427, 2015. ISSN 15372723. doi: 10.1080/00401706.2014.928233.

R. H. Leary. Global Optimization on Funneling Landscapes. *Journal of Global Optimization*, 18 (4):367–383, 2000. ISSN 09255001. doi: 10.1023/A:1026500301312.

S. Leary, A. Bhaskar, and A. Keane. A constraint mapping approach to structural optimization of an expensive model uisng surrogates, 2001.

S. J. Leary, A. Bhaskar, and A. J. Keane. A knowledge-based approach to response surface modelling in multifidelity optimization. *Journal of Global Optimization*, 26(3):297–319, 2003. ISSN 09255001. doi: 10.1023/A:1023283917997.

E. S. Lee and R. J. Li. Fuzzy multiple objective programming and compromise programming with Pareto optimum. *Fuzzy sets and systems*, 53(3):275–288, 1993.

L. Leifsson, S. Koziel, and Y. Tesfahunegn. Fast Multi-Objective Aerodynamic Optimization Using Space-Mapping-Corrected Multi-Fidelity Models and Kriging Interpolation. 153, 2016. doi: 10.1007/978-3-319-27517-8. URL http://link.springer.com/10.1007/978-3-319-27517-8.

K. Levenberg. A Method for the Solution of Certain Non-Linear Problems in Least Squares. *Quarterly of Applied Mathematics*, 2(2):164–168, 1944. ISSN 0022-0574.

J.-C. Lévesque, A. Durand, C. Gagné, and R. Sabourin. Bayesian optimization for conditional hyperparameter spaces. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 286–293. IEEE, 2017.

R. Lewis and S. Nash. A multigrid approach to the optimization of systems governed by differential equations. In *8th symposium on multidisciplinary analysis and optimization*, page 4890, 2000.

B. Li, L. Peng, and B. Ramadass. Accurate and efficient processor performance prediction via regression tree based modeling. *Journal of Systems Architecture*, 55(10-12):457–467, 2009. ISSN 13837621. doi: 10.1016/j.sysarc.2009.09.004. URL http://dx.doi.org/10.1016/j.sysarc.2009.09.004.

B. Li, H. Zhang, W. Zheng, and L. Wang. Spacecraft close-range trajectory planning via convex optimization and multi-resolution technique. *Acta Astronautica*, 175(November 2019):421–437, 2020. ISSN 00945765. doi: https://doi.org/10.1016/j.actaastro.2020.05.051.

G. Li, S. Azarm, A. Farhang-Mehr, and A. R. Diaz. Approximation of multiresponse deterministic engineering simulations: A dependent metamodeling approach. *Structural and Multidisciplinary Optimization*, 31(4):260–269, 2006. ISSN 1615147X. doi: 10.1007/s00158-005-0574-5.

G. Li, V. Aute, and S. Azarm. An accumulative error based adaptive design of experiments for offline metamodeling. *Structural and Multidisciplinary Optimization*, 40(1-6):137–155, 2010. ISSN 1615147X. doi: 10.1007/s00158-009-0395-z.

Y. Lin, F. Mistree, J. K. Allen, K. L. Tsui, and V. C. Chen. A sequential exploratory experimental design method: Development of appropriate empirical models in design. *Proceedings of the ASME Design Engineering Technical Conference*, 1:1021–1035, 2004. doi: 10.1115/detc2004-57527.

D. J. Linse and R. F. Stengel. Identification of aerodynamic coefficients using computational neural networks. *Journal of Guidance, Control, and Dynamics*, 16(6):1018–1025, 1993. ISSN 07315090. doi: 10.2514/3.21122.

D. Liu and Y. Wang. Multi-fidelity physics-constrained neural network and its application in materials modeling. *Journal of Mechanical Design*, 141(12):121403, 2019.

H. Liu, S. Xu, and X. Wang. Sequential sampling designs based on space reduction. *Engineering Optimization*, 47(7):867–884, 2015. ISSN 10290273. doi: 10.1080/0305215X.2014.928816. URL https://doi.org/10.1080/0305215X.2014.928816.

H. Liu, S. Xu, Y. Ma, X. Chen, and X. Wang. An Adaptive Bayesian Sequential Sampling Approach for Global Metamodeling. *Journal of Mechanical Design*, 138(1), 2016. ISSN 10500472. doi: 10.1115/1.4031905.

H. Liu, Y. S. Ong, and J. Cai. A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design, 2018a. ISSN 16151488.

H. Liu, Y. S. Ong, and J. Cai. A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design. *Structural and Multidisciplinary Optimization*, 57(1):393–416, 2018b. ISSN 16151488. doi: 10.1007/s00158-017-1739-8.

H. Liu, S. Xu, X. Wang, S. Yang, and J. Meng. A multi-response adaptive sampling approach for global metamodeling. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 232(1):3–16, 2018c. ISSN 20412983. doi: 10.1177/0954406216672250.

J. Liu, Z. Han, and W. Song. Comparison of Infill Sampling Criteria in Kriging-Based Aerodynamic Optimization. *28th ICAS*, pages 1–10, 2014.

W. Liu and S. M. Batill. Gradient-enhanced neural network response surface approximations. *8th Symposium on Multidisciplinary Analysis and Optimization*, (February), 2000. doi: 10.2514/6.2000-4923.

W. Liu and S. M. Batill. Gradient-enhanced response surface approximations using kriging models. *9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, pages 1–11, 2002. doi: 10.2514/6.2002-5456.

H. W. Loeb, V. G. Petukhov, G. A. Popov, and A. I. Mogulkin. A realistic concept of a manned Mars mission with nuclear-electric propulsion. *Acta Astronautica*, 116:299–306, 2015. ISSN 00945765. doi: 10.1016/j.actaastro.2015.07.019. URL http://dx.doi.org/10.1016/j.actaastro.2015.07.019.

F. Logist, P. M. Van Erdeghem, and J. F. Van Impe. Efficient deterministic multiple objective optimal control of (bio)chemical processes. *Chemical Engineering Science*, 64(11):2527–2538, 2009. ISSN 00092509. doi: 10.1016/j.ces.2009.01.054.

M. Long, E. Betrán, K. Thornton, and W. Wang. The origin of new genes: Glimpses from the young and old. *Nature Reviews Genetics*, 4(11):865–875, 2003. ISSN 14710056. doi: 10.1038/nrg1204.

P. Lopez, E. Schultz, B. Mattfeld, C. Stromgren, and K. Goodliff. Logistics needs for potential deep space mission scenarios post Asteroid Redirect crewed Mission. *IEEE Aerospace Conference Proceedings*, 2015-June, 2015. ISSN 1095323X. doi: 10.1109/AERO.2015.7119161.

A. Lovison and E. Rigoni. Adaptive sampling with a Lipschitz criterion for accurate metamodeling. *Communications in Applied and Industrial Mathematics*, 1(2):110–126, 2010. doi: 10.1685/2010CAIM545.

D. J. C. MacKay. A Practical Bayesian Framework for Backpropagation Networks. *Neural Computation*, 472(1):448–472, 1992. URL http://authors.library.caltech.edu/13793/1/MACnc92b.pdf.

T. J. MacKman, C. B. Allen, M. Ghoreyshi, and K. J. Badcock. Comparison of adaptive sampling methods for generation of surrogate aerodynamic model. *AIAA Journal*, 51(4):797–808, 2013. ISSN 00011452. doi: 10.2514/1.J051607.

N. K. Madavan. Multiobjective Optimization Using a Pareto Differential Evolution Approach. *Proceedings of the 2002 Congress on Evolutionary Computation*, 2, 2002. doi: https://doi.org/10.1109/CEC.2002.1004404.

C. Maddock, B. Parsonage, L. Ricciardi, M. Vasile, and O. Cohen. Design optimisation and analysis of very high power transportation system to Mars. *Proceedings of the International Astronautical Congress, IAC*, A5(October):25–29, 2021. ISSN 00741795.

K. Madsen, J. Bandler, S. Dakroury, M. Bakr, J. Sondergaard, Q. Cheng, and A. Mohamed. Space Mapping: The State of the Art. *IEEE Transactions on Microwave Theory and Techniques*, 52 (1):337–361, 2004. ISSN 0018-9480. doi: 10.1109/tmtt.2003.820904.

M. W. Maier. *The art of systems architecting*. CRC press, 2009.

L. Mainini, A. Serani, M. P. Rumpfkeil, E. Minisci, D. Quagliarella, H. Pehlivan, S. Yildiz, S. Ficini, R. Pellegrini, F. Di Fiore, D. Bryson, M. Nikbay, M. Diez, and P. Beran. Analytical Benchmark Problems for Multifidelity Optimization Methods. pages 1–14, 2022. URL http://arxiv.org/abs/2204.07867.

S. G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *Fundamental Papers in Wavelet Theory*, II(7), 1989. doi: https://doi.org/10.1109/34.192463.

A. March and K. Willcox. Provably convergent multifidelity optimization algorithm not requiring high-fidelity derivatives. *AIAA Journal*, 50(5):1079–1089, 2012a. ISSN 00011452. doi: 10.2514/1.J051125.

A. March and K. Willcox. Constrained multifidelity optimization using model calibration. *Structural and Multidisciplinary Optimization*, 46(1):93–109, 2012b. ISSN 1615147X. doi: 10.1007/s00158-011-0749-1.

A. March, K. Willcox, and Q. Wang. Gradient-based multifidelity optimisation for aircraft design using Bayesian model calibration. *Aeronautical Journal*, 115(1174):729–738, 2011. ISSN 00019240. doi: 10.1017/S0001924000006473.

A. I. March. *Multifidelity Methods for Multidisciplinary System Design*. PhD thesis, 2012. URL https://dspace.mit.edu/bitstream/handle/1721.1/76145/820204781-MIT.pdf?sequence=2.

D. Marquardt. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.

A. L. Marsden, M. Wang, J. E. Dennis, Jr., and P. Moin. Optimal Aeroacoustic Shape Design Using the Surrogate Management Framework. *Optimization and Engineering*, 5(2):235–262, 2004. doi: 10.1023/b:opte.0000033376.89159.65.

J. Martínez-Frutos and D. Herrero-Pérez. Kriging-based infill sampling criterion for constraint handling in multi-objective optimization. *Journal of Global Optimization*, 64(1):97–115, 2016. ISSN 15732916. doi: 10.1007/s10898-015-0370-8.

J. Martins. A Short Course on Multidisciplinary Design Optimization. *University of Michigan*, 2012.

J. R. R. A. Martins and A. B. Lambe. Multidisciplinary Design Optimization: A Survey of Architectures. *AIAA Journal*, 51(9):2049–2075, 2013. ISSN 0001-1452. doi: 10.2514/1.j051895.

B. H. Mason, R. T. Haftka, E. R. Johnson, and G. L. Farley. Variable complexity design of composite fuselage frames by response surface techniques. *Thin-Walled Structures*, 32(4): 235–261, 1998. ISSN 02638231. doi: 10.1016/S0263-8231(98)00016-0.

H. Maurer and M. Wiegand. Numerical solution of a drug displacement problem with bounded state variables. *Optimal Control Applications and Methods*, 13(1):43–55, 1992. ISSN 10991514. doi: 10.1002/oca.4660130104.

J. F. Mccloskey. The Beginnings of Operations Research: 1934-1941. *Operations Research*, 35(1): 143–152, 1987. URL https://www.jstor.org/stable/170920.

W. S. McCulloch and W. H. Pitts. Logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 1943.

J. McFarland, S. Mahadevan, V. Romero, and L. Swileir. Calibration and uncertainty analysis for computer simulations with multivariate output. *AIAA Journal*, 46(5):1253–1265, 2008. ISSN 00011452. doi: 10.2514/1.35288.

J. Meditch. On the problem of optimal thrust programming for a soft lunar landing. 1964. doi: https://doi.org/10.1109/TAC.1964.1105758.

P. M. Mehta, E. Minisci, M. Vasile, A. C. Walker, and M. Brown. An Open-source Hypersonic Aerodynamic And Aerothermodynamic Modeling Tool. *8th European Symposium on Aerothermodynamics for Space Vehicles*, (1):9, 2015.

X. Meng and G. E. Karniadakis. A composite neural network that learns from multi-fidelity data: Application to function approximation and inverse pde problems. *Journal of Computational Physics*, 401:109020, 2020.

T. Mengistu and W. Ghaly. Aerodynamic optimization of turbomachinery blades using evolutionary methods and ANN-based surrogate models. *Optimization and Engineering*, 9(3): 239–255, 2008. ISSN 13894420. doi: 10.1007/s11081-007-9031-1.

A. Merlevede, H. Åhl, and C. Troein. Homology and linkage in crossover for linear genomes of variable length. *PLoS ONE*, 14(1):1–16, 2019. ISSN 19326203. doi: 10.1371/journal.pone.0209712.

A. Messac. Physical programming: Effective optimization for computational design. *AIAA Journal*, 34(1):149–158, 1996. ISSN 00011452. doi: 10.2514/3.13035.

A. Messac and C. A. Mattson. Normal constraint method with guarantee of even representation of complete pareto frontier. *AIAA Journal*, 42(10):2101–2111, 2004. ISSN 00011452. doi: 10.2514/1.8977.

E. Mezura-Montes and C. A. C. Coello. Constrained optimization via multiobjective evolutionary algorithms. *Multiobjective Problem Solving from Nature: From Concepts to Applications*, pages 53–75, 2008.

A. Miele. Recent advances in gradient algorithms for optimal control problems. *Journal of Optimization Theory and Applications*, 17:361–430, 1975.

E. Minisci and M. Vasile. Robust Design of a Re-entry Unmanned Space Vehicle by Multi-Fidelity Evolution Control. 51:4–6, 2013.

E. Minisci, M. Vasile, and H. Liqiang. Robust multi-fidelity design of a micro re-entry unmanned space vehicle. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, 225(11):1195–1209, 2011. ISSN 20413025. doi: 10.1177/0954410011410124.

D. C. Montgomery. *Desgin And Analysis Of Experiments*. 2001. ISBN 0-471-31469-0.

G. E. Moore and Others. Cramming more components onto integrated circuits, 1965.

M. D. Morris, T. J. Mitchell, and D. Ylvisaker. Bayesian design and analysis of computer experiments: Use of derivatives in surface prediction. *Technometrics*, 35(3):243–255, 1993. ISSN 15372723. doi: 10.1080/00401706.1993.10485320.

A. A. Mullur and A. Messac. Metamodeling using extended radial basis functions: A comparative approach. *Engineering with Computers*, 21(3):203–217, 2006. ISSN 01770667. doi: 10.1007/s00366-005-0005-7.

M. Munetomo, Y. Takai, and Y. Sato. Migration scheme for the genetic adaptive routing algorithm. *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, 3:2774–2779, 1998. ISSN 08843627. doi: 10.1109/icsmc.1998.725081.

D. E. Myers. Co-kriging—new developments. In *Geostatistics for Natural Resources Characterization Part 1*, pages 295–305. Springer, 1984.

R. H. Myers and D. C. Montgomery. Response Surface Methodology: Process and Product in Optimization Using Designed Experiments. 1995. ISSN 0471581003.

A. Narayan, C. Gittelson, and D. Xiu. A stochastic collocation algorithm with multifidelity models. *SIAM Journal of Scientific Computing*, 36(2):495–521, 2014.

E. National Academies of Sciences and Medicine. *Space Nuclear Propulsion for Human Mars Exploration*. 2021. ISBN 0309684803.

J. A. Nelder and R. Mead. A simplex method for function minimization. *The computer journal*, 7(4):308–313, 1965. ISSN 1460-2067.

Y. Nesterov and A. Nemirovskii. *Interior-point polynomial algorithms in convex programming*. SIAM, 1994.

L. W. Ng and K. E. Willcox. Multifidelity approaches for optimization under uncertainty. *International Journal for Numerical Methods in Engineering*, (February):1102–1119, 2014. doi: 10.1002/nme. URL http://onlinelibrary.wiley.com/doi/10.1002/nme.3279/full.

T. N. Ngo and N. Hayek. Necessary conditions of Pareto optimality for multiobjective optimal control problems under constraints. *Optimization*, 66(2):149–177, 2017. ISSN 10294945. doi: 10.1080/02331934.2016.1261349.

S. Ober-Blobaum, M. Ringkamp, and G. Zum Felde. Solving multiobjective Optimal Control problems in space mission design using Discrete Mechanics and reference point techniques. *Proceedings of the IEEE Conference on Decision and Control*, pages 5711–5716, 2012. ISSN 25762370. doi: 10.1109/CDC.2012.6426285.

S. R. Oleson, L. Burke, L. Dudzinski, J. Fittje, L. S. Mason, T. Packard, P. Schmitz, J. Gyekenyesi, and B. Faller. A Combined Nuclear Electric and Chemical Propulsion Vehicle Concept for Piloted Mars Opposition Class Missions. (January 2021), 2020. doi: 10.2514/6.2020-4055.

A. C. Owens, C. A. Jones, W. M. Cirillo, J. J. Klovstad, E. L. Judd, P. R. Chai, R. G. Merrill, A. Prasad, J. Cho, and C. Stromgren. Integrated crewed mars mission analysis, part ii: The trajectory strikes back. *Accelerating Space Commerce, Exploration, and New Discovery conference, ASCEND 2021*, 2021. doi: 10.2514/6.2021-4134.

R. M. Paiva, A. R. Carvalho, C. Crawford, and A. Suleman. Comparison of surrogate models in a multidisciplinary optimization framework for wing design. *AIAA Journal*, 48(5):995–1006, 2010. ISSN 00011452. doi: 10.2514/1.45790.

F. Palacios, M. R. Colonno, A. C. Aranake, A. Campos, S. R. Copeland, T. D. Economon, A. K. Lonkar, T. W. Lukaczyk, T. W. Taylor, and J. J. Alonso. Stanford University Unstructured (SU2): An open-source integrated computational environment for multi-physics simulation and design. *51st AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition 2013*, (January):1–60, 2013. doi: 10.2514/6.2013-287.

K. Palmer and K. Tsui. A minimum bias Latin hypercube design. *IIE Transactions (Institute of Industrial Engineers)*, 33(9):793–808, 2001. ISSN 0740817X. doi: 10.1080/07408170108936873.

S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 2010.

V. Pareto. *Manuale di economia politica: con una introduzione alla scienza sociale*, volume 13. Società editrice libraria, 1919.

C. Park, R. T. Haftka, and N. H. Kim. Remarks on multi-fidelity surrogates. *Structural and Multidisciplinary Optimization*, 55(3):1029–1050, 2017. ISSN 16151488. doi: 10.1007/s00158-016-1550-y. URL http://dx.doi.org/10.1007/s00158-016-1550-y.

C. Park, R. T. Haftka, and N. H. Kim. Low-fidelity scale factor improves Bayesian multi-fidelity prediction by reducing bumpiness of discrepancy function. *Structural and Multidisciplinary Optimization*, 58(2):399–414, 2018. ISSN 16151488. doi: 10.1007/s00158-018-2031-2.

J. S. Park. Tuning complex computer codes to data and optimal designs. 1991.

J. M. Parr, A. I. Forrester, A. J. Keane, and C. M. Holden. Enhancing infill sampling criteria for surrogate-based constrained optimization. *Journal of Computational Methods in Sciences and Engineering*, 12(1-2):25–45, 2012a. ISSN 14727978. doi: 10.3233/JCM-2012-0402.

J. M. Parr, A. J. Keane, A. I. Forrester, and C. M. Holden. Infill sampling criteria for surrogate-based optimization with constraint handling. *Engineering Optimization*, 44(10):1147–1166, 2012b. ISSN 0305215X. doi: 10.1080/0305215X.2011.637556.

B. Parsonage and C. Maddock. A multi-fidelity model management framework for multi-objective aerospace design optimisation. *Frontiers in Aerospace Engineering*, 2(February): 1–21, 2023. doi: 10.3389/fpace.2023.1046177.

B. Parsonage and C. Maddock. Biased Dyadic Crossover for Variable-Length Multi-Objective Optimal Control Problems. *2024 IEEE Congress on Evolutionary Computation, CEC 2024 - Proceedings*, pages 1–9, 2024a. doi: 10.1109/CEC60901.2024.10611888.

B. Parsonage and C. Maddock. Solving multi-objective optimal control problems using a multiresolution approach. *Journal of Guidance, Control, and Dynamics*, pages 1–14, 2024b. ISSN 1533-3884. doi: 10.2514/1.g008223.

B. Parsonage and C. Maddock. Multi-Objective Design Optimisation and Analysis of a Crewed Earth-Mars Transportation System using Nuclear Thermal Propulsion. *Proceedings of the International Astronautical Congress, IAC*, (October), 2024c.

B. Parsonage and C. A. Maddock. Multi-stage Multi-fidelity Information Correction for Artificial Neural Network Based Meta-modelling. *2020 IEEE Symposium Series on Computational Intelligence, SSCI 2020*, pages 950–957, 2020. doi: 10.1109/SSCI47803.2020.9308255.

A. Pascoletti and P. Serafini. Scalarizing vector optimization problems. *Journal of Optimization Theory and Applications*, 42(4):499–524, 1984. ISSN 00223239. doi: 10.1007/BF00934564.

M. A. Patterson, W. W. Hager, and A. V. Rao. A ph mesh refinement method for optimal control. *Optimal Control Applications and Methods*, 36(4):398–421, 2015. ISSN 10991514. doi: 10.1002/oca.2114.

W. R. Pearson. An introduction to sequence similarity ("homology") searching. *Current Protocols in Bioinformatics*, (SUPPL.42):1–8, 2013. ISSN 19343396. doi: 10.1002/0471250953.bi0301s42.

P. Pedersen. *The integrated approach of FEM-SLP for solving problems of optimal design*. Technical University of Denmark, 1980.

B. Peherstorfer, T. Cui, Y. Marzouk, and K. Willcox. Multi-fidelity Monte Carlo: importance sampling method using high-fidelity and surrogate model. 7831:1–23, 2015.

B. Peherstorfer, K. Willcox, and M. Gunzburger. Optimal model management for multifidelity Monte Carlo estimation. *SIAM Journal on Scientific Computing*, 38(5):A3163–A3194, 2016. ISSN 10957200. doi: 10.1137/15M1046472.

B. Peherstorfer, K. Willcox, and M. Gunzburger. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *SIAM Review*, 60(3):550–591, 2018. ISSN 00361445. doi: 10.1137/16M1082469.

M. Penwarden, S. Zhe, A. Narayan, and R. M. Kirby. Multifidelity modeling for physics-informed neural networks (pinns). *Journal of Computational Physics*, 451:110844, 2022.

P. Perdikaris, D. Venturi, J. O. Royset, and G. E. Karniadakis. Multi-fidelity modelling via recursive co-kriging and Gaussian-Markov random fields. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2179), 2015. ISSN 14712946. doi: 10.1098/rspa.2015.0018.

H. J. Pesch. Real-time computation of feedback controls for constrained optimal control problems. part 2: A correction method based on multiple shooting. *Optimal Control Applications and Methods*, 10(2):147–171, 1989. ISSN 10991514. doi: 10.1002/oca.4660100206.

E. Polak. An Historical Survey of Computational Methods in Optimal Control. 15(2):553–584, 1973. doi: https://doi.org/10.1137/1015071.

Pomax. A Primer on Bezier Curves, 2023. URL https://pomax.github.io/bezierinfo/.

M. J. Powell. A fortran subroutine for of non-linear. 1968.

M. J. Powell. A new algorithm for unconstrained optimization. In *Nonlinear programming*, pages 31–65. Elsevier, 1970.

M. J. D. Powell. Convergence properties of a class of minimization algorithms. In *Nonlinear programming 2*, pages 1–27. Elsevier, 1975.

Y. Qi, X. Ma, F. Liu, L. Jiao, J. Sun, and J. Wu. MOEA/D with Adaptive Weight Adjustment. *Evolutionary Computation*, 22(2):231–264, 2014. ISSN 15309304. doi: 10.1162/EVCO_a_00109.

P. Z. Qian and C. F. Wu. Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments. *Technometrics*, 50(2):192–204, 2008. ISSN 00401706. doi: 10.1198/004017008000000082.

P. Z. Qian, B. Tang, and C. F. Jeff Wu. Nested space-filling designs for computer experiments with two levels of accuracy. *Statistica Sinica*, 19(1):287–300, 2009. ISSN 10170405.

P. Z. Qian, Y. Hwang, M. Ai, and H. Su. Asymmetric nested lattice samples. *Technometrics*, 56 (1):46–54, 2014. ISSN 00401706. doi: 10.1080/00401706.2013.800002.

Z. Qiongbing and D. Lixin. A new crossover mechanism for genetic algorithms with variable-length chromosomes for path optimization problems. *Expert Systems with Applications*, 60: 183–189, 2016. ISSN 09574174. doi: 10.1016/j.eswa.2016.04.005.

N. V. Queipo, R. T. Haftka, W. Shyy, T. Goel, R. Vaidyanathan, and P. Kevin Tucker. Surrogate-based analysis and optimization. *Progress in Aerospace Sciences*, 41(1):1–28, 2005. ISSN 03760421. doi: 10.1016/j.paerosci.2005.02.001.

M. M. Rai and N. K. Madavan. Aerodynamic design using neural networks. *AIAA journal*, 38 (1):173–182, 2000. ISSN 00011452. doi: 10.2514/2.938.

D. Rajnarayan, A. Haas, and I. Kroo. A multifidelity gradient-free optimization method and application to aerodynamic design. *12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, MAO*, (September), 2008.

D. G. Rajnarayan. *Trading risk and performance for engineering design optimization using multifidelity analyses*. Stanford University, 2009. ISBN 1109240821.

A. Rao. A Survey of Numerical Methods for Optimal Control. *Solution Methods for Integral Equations*, pages 1–58, 2013. doi: 10.1007/978-1-4757-1466-1_1.

S. S. Rao. *Engineering Optimization: Theory and Practice*, volume 56. 2009. ISBN 9780470183526. doi: 10.1080/03052150500066646.

J. Rasmussen. Accumulated approximation: A new method for structural optimization by iterative improvement. In *NASA. Langley Research Center, The Third Air Force (NASA Symposium on Recent Advances in Multidisciplinary Analysis and Optimization*, 1990.

J. E. Rayas-Sánchez. The Aggressive Space Mapping (ASM) Algorithm : Over Two Decades of Development and Engineering Applications. pages 1–18, 2015.

S. Razavi, B. A. Tolson, and D. H. Burn. Numerical assessment of metamodelling strategies in computationally intensive optimization. *Environmental Modelling and Software*, 34:67–86, 2012. ISSN 13648152. doi: 10.1016/j.envsoft.2011.09.010. URL http://dx.doi.org/10.1016/j.envsoft.2011.09.010.

G. W. Reddien. Collocation at Gauss points as a discretization in optimal control. *SIAM Journal on Control and Optimization*, 17(2):298–306, 1979.

M. Redhe and L. Nilsson. Optimization of the new Saab 9-3 exposed to impact load using a space mapping technique. *Structural and Multidisciplinary Optimization*, 27(5):411–420, 2004. ISSN 1615147X. doi: 10.1007/s00158-004-0396-x.

M. Redhe and L. Nilsson. A multipoint version of space mapping optimization applied to vehicle crashworthiness design. *Structural and Multidisciplinary Optimization*, 31(2):134–146, 2006. ISSN 1615147X. doi: 10.1007/s00158-005-0544-y.

G. R. Reeck, C. de Haen, D. C. Teller, R. F. Doollttle, W. M. Fitch, R. E. Dickerson, P. Chambon, A. D. McLachlan, E. Margoliash, T. H. Jukes, and E. Zuckerkandl. "Homology" in Proteins and Nucleic Acids: A Terminology Muddle and a Way out of It. *Accountability in Research*, 1987.

F. Regazzoni, S. Pagani, A. Cosenza, A. Lombardi, and A. Quarteroni. A physics-informed multi-fidelity approach for the estimation of differential equations parameters in low-data or large-noise regimes. *Rendiconti Lincei*, 32(3):437–470, 2021.

R. Reichart, K. Tomanek, U. Hahn, and A. Rappoport. Multi-task active learning for linguistic annotations. *ACL-08: HLT - 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference*, (June):861–869, 2008.

P.-E. Rethore, P. Fugslang, G. C. Larsen, T. Buhl, T. J. Larsen, and H. A. Madsen. Multi-fidelity optimization of wind farms. *Wind Energy*, (November 2013), 2013. doi: 10.1002/we.

L. A. Ricciardi. *Multi-Objective Hybrid Optimal Control with Application to Space Systems*. Doctoral dissertation, University of Strathclyde, 2019.

L. A. Ricciardi and M. Vasile. MODHOC - Multi Objective Direct Hybrid Optimal Control. 2018. URL https://strathprints.strath.ac.uk/67085/.

L. A. Ricciardi and M. Vasile. Direct transcription of optimal control problems with finite elements on bernstein basis. *Journal of Guidance, Control, and Dynamics*, 42(2):229–243, 2019a. ISSN 07315090. doi: 10.2514/1.G003753.

L. A. Ricciardi and M. Vasile. Improved Archiving and Search Strategies for Multi Agent Collaborative Search. *Computational Methods in Applied Sciences*, 48(January):435–455, 2019b. ISSN 18713033. doi: 10.1007/978-3-319-89988-6_26.

L. A. Ricciardi and M. Vasile. A relaxation approach for hybrid multi-objective optimal control: Application to multiple debris removal missions. *Advances in the Astronautical Sciences*, 168: 681–700, 2019c. ISSN 00653438.

L. A. Ricciardi, M. Vasile, F. Toso, and C. A. Maddock. Multi-Objective Optimal Control of the Ascent Trajectories of Launch Vehicles. pages 1–14, 2016. doi: 10.2514/6.2016-5669.

L. A. Ricciardi, C. A. Maddock, and M. Vasile. Direct Solution of Multi-Objective Optimal Control Problems Applied to Spaceplane Mission Design. *Journal of Guidance, Control, and Dynamics*, 42(1):30–46, 2018a. ISSN 0731-5090. doi: 10.2514/1.g003839.

L. A. Ricciardi, C. A. Maddock, and M. Vasile. Multi-objective optimal control of re-entry and abort scenarios. In *Space Flight Mechanics Meeting, 2018*, number 210009, 2018b. ISBN 9781624105333. doi: 10.2514/6.2018-0218.

L. A. Ricciardi, C. A. Maddock, and M. Vasile. Direct Solution of Multi-Objective Optimal Control Problems Applied to Spaceplane Mission Design. *Journal of Guidance, Control, and Dynamics*, 42(1):30–46, 2019. ISSN 0731-5090. doi: 10.2514/1.g003839.

J. L. Ringuest. Compromise programming. *Multiobjective optimization: behavioral and computational considerations*, pages 51–59, 1992.

J. Robinson. An Overview of NASA'S Integrated Design and Engineering Analysis (IDEA) Environment. pages 1–16, 2012. doi: 10.2514/6.2011-2392.

T. D. Robinson, M. S. Eldred, K. E. Willcox, and R. Haimes. Surrogate-based optimization using multifidelity models with variable parameterization and corrected space mapping. *AIAA Journal*, 46(11):2814–2822, 2008a. ISSN 00011452. doi: 10.2514/1.36043.

T. D. Robinson, M. S. Eldred, K. E. Willcox, and R. Haimes. Surrogate-based optimization using multifidelity models with variable parameterization and corrected space mapping. *AIAA Journal*, 46(11):2814–2822, 2008b. ISSN 00011452. doi: 10.2514/1.36043.

J. F. Rodriguez, J. E. Renaud, and L. T. Watson. Trust region augmented Lagrangian methods for sequential response surface approximation and optimization. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 80449, page V002T29A004. American Society of Mechanical Engineers, 1997. ISBN 0791880443.

J. F. Rodríguez, J. E. Renaud, and L. T. Watson. Convergence of trust region augmented Lagrangian methods using variable fidelity approximation data. *Structural Optimization*, 15 (3-4):141–156, 1998. ISSN 09344373. doi: 10.1007/bf01203525.

J. F. Rodríguez, V. M. Pérez, D. Padmanabhan, and J. E. Renaud. Sequential approximate optimization using variable fidelity response surface approximations. *Structural and Multidisciplinary Optimization*, 22(1):24–34, 2001. ISSN 1615147X. doi: 10.1007/s001580100122.

A. Rodríguez-Molina, J. Solís-Romero, M. G. Villarreal-Cervantes, O. Serrano-Pérez, and G. Flores-Caballero. Path-Planning for Mobile Robots Using a Novel Variable-Length Differential Evolution Variant. *Mathematics*, 9(4):357, 2021. ISSN 22277390. doi: 10.3390/math9040357.

B. Rosenbaum. Efficient Global Surrogate Models for Responses of Expensive Simulations. *PhD Thesis*, page 171, 2013.

F. Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958. ISSN 1939-1471.

J. Roshanian, A. A. Bataleblu, and M. Ebrahimi. Robust ascent trajectory design and optimization of a typical launch vehicle. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 232(24):4601–4614, 2018. ISSN 20412983. doi: 10.1177/0954406217753460.

J. Roshanian, A. A. Bataleblu, and M. Ebrahimi. A novel metamodel management strategy for robust trajectory design of an expendable launch vehicle. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, 0(0):1–18, 2019. ISSN 20413025. doi: 10.1177/0954410019860638.

I. M. Ross and F. Fahroo. Pseudospectral knotting methods for solving optimal control problems. *Journal of Guidance, Control, and Dynamics*, 27(3):397–405, 2004. ISSN 15333884. doi: 10.2514/1.3426.

I. M. Ross and M. Karpenko. A review of pseudospectral optimal control: From theory to flight. *Annual Reviews in Control*, 36(2):182–197, 2012.

W. J. Roux, N. Stander, and R. T. Haftka. Response surface approximations for structural optimization. *6th Symposium on Multidisciplinary Analysis and Optimization*, 534(August 1997):565–578, 1996. doi: 10.2514/6.1996-4042.

G. Rozza, D. B. P. Huynh, and A. T. Patera. Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations. *Archives of Computational Methods in Engineering*, 15(3):1–47, 2007. ISSN 1134-3060. doi: 10.1007/bf03024948.

D. E. Rumelhart, B. Widrow, and M. A. Lehr. The Basic Ideas in Neural Networks. *Communications of the ACM*, 37(3):87–92, 1994. ISSN 15577317. doi: 10.1145/175247.175256.

J. Sacks, W. J. Welch, T. J. M. And, and H. P. Wynn. Design and analysis of computer experiments. *Statistical SCience*, 4(4):409–423, 1989. ISSN 0002-6018. doi: 10.1002/9781118842003.ch9.

J. M. Salotti, R. Heidmann, and E. Suhir. Crew size impact on the design, risks and cost of a human mission to mars. *IEEE Aerospace Conference Proceedings*, (March), 2014. ISSN 1095323X. doi: 10.1109/AERO.2014.6836241.

A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259–270, 2010. ISSN 00104655. doi: 10.1016/j.cpc.2009.09.018. URL http://dx.doi.org/10.1016/j.cpc.2009.09.018.

T. J. Santner, B. J. Williams, W. I. Notz, and B. J. Williams. *The design and analysis of computer experiments*, volume 1. Springer, 2003.

L. d. S. Santos, K. M. F. de Souza, M. R. Bandeira, V. R. Ruiz Ahón, F. C. Peixoto, and D. M. Prata. Dynamic optimization of a continuous gas lift process using a mesh refining sequential method. *Journal of Petroleum Science and Engineering*, 165(September 2017):161–170, 2018. ISSN 09204105. doi: 10.1016/j.petrol.2018.02.019.

M. Santos, B. Mattos, and R. Girardi. Aerodynamic coefficient prediction of airfoils using neural networks. In *46th AIAA aerospace sciences meeting and exhibit*, page 887, 2008.

M. J. Sasena. Flexibility and Efficiency Enhancements for Constrained Global Design Optimization with Kriging Approximations. *University of Michigan*, pages 1–237, 2002. doi: 10.1.1.2.4697. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.2.4697.

R. R. Schaller. Moore's law: past, present, and future. *IEEE Spectrum*, 34(6):52–55, 57, 1997. ISSN 00189235. doi: 10.1109/6.591665.

K. J. Schlager. Systenas Engineering-Key to Modern Development. *Ire Transactions on Engineering Management*, pages 1953–1955, 1953. URL http://origins.sese.asu.edu/ses405/AdditionalReading/05007383.pdf.

M. Schlegel, K. Stockmann, T. Binder, and W. Marquardt. Dynamic optimization using adaptive control vector parameterization. *Computers and Chemical Engineering*, 29(8):1731–1751, 2005. ISSN 00981354. doi: 10.1016/j.compchemeng.2005.02.036.

P. J. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, 2010. ISSN 14697645. doi: 10.1017/S0022112010001217.

L. A. Schmit. Structural Design by Systematic Synthesis. *Proceedings, 2nd Conference on Electronic Computation, ASCE*, pages 105–122, 1960.

L. A. Schmit and B. Farshl. Some approximation concepts for structural synthesis. *AIAA Journal*, 12(5):692–699, 1974. ISSN 00011452. doi: 10.2514/3.49321.

M. Schonlau. Computer Experiments and Global Optimization. 1997.

O. Schütze, X. Esquivel, A. Lara, and C. A. Coello. Using the averaged hausdorff distance as a performance measure in evolutionary multiobjective optimization. *IEEE Transactions on Evolutionary Computation*, 16(4):504–522, 2012. ISSN 1089778X. doi: 10.1109/TEVC.2011.2161872.

A. L. Schwartz. *Theory and implementation of numerical methods based on Runge-Kutta integration for solving optimal control problems.* PhD thesis, University of California, Berkeley, 1996. URL http://www2.eecs.berkeley.edu/Pubs/TechRpts/1996/2995.html.

B. Settles. Computer Sciences Active Learning Literature Survey. (January), 2009.

H. S. Seung, M. Opper, and H. Sompolinsky. Query by committee. In *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory*, pages 287–294, 1992. ISBN 089791497X. doi: 10.1145/130385.130417.

L. R. Shepherd and A. V. Cleaver. The atomic rocket. I. *J. Brit. Interplanetary Soc.*, 7, 1948.

M. C. Shewry and H. P. Wynn. Maximum entropy sampling. *Journal of applied statistics*, 14(2): 165–170, 1987. ISSN 0266-4763.

M. Shi, L. Lv, W. Sun, and X. Song. A multi-fidelity surrogate model based on support vector regression. *Structural and Multidisciplinary Optimization*, 2020a. ISSN 16151488. doi: 10.1007/s00158-020-02522-6.

R. Shi, T. Long, and H. Baoyin. Multi-fidelity and multi-objective optimization of low-thrust transfers with control strategy for all-electric geostationary satellites. *Acta Astronautica*, 177 (June):577–587, 2020b. ISSN 00945765. doi: 10.1016/j.actaastro.2020.08.013. URL https://doi.org/10.1016/j.actaastro.2020.08.013.

M. A. Simon, M. Clowdsley, and S. Walker. Habitat design considerations for implementing solar particle event radiation protection. *43rd International Conference on Environmental Systems*, pages 1–13, 2013. doi: 10.2514/6.2013-3403.

T. W. Simpson. Design and Analysis of Computer Experiments in Multidisciplinary Design Optimization: A Review of How Far We Have Come - or Not. *AIAA*, (September):1–21, 2008. doi: doi:10.2514/6.2008-5802.

T. W. Simpson, J. D. Peplinski, P. N. Koch, and J. K. Allen. Metamodels for computer-based engineering design: Survey and recommendations. *Engineering with Computers*, 17(2): 129–150, 2001. ISSN 01770667. doi: 10.1007/PL00007198.

G. Singh and R. V. Grandhi. Mixed-Variable Optimization Strategy Employing Multifidelity Simulation and Surrogate Models. *AIAA Journal*, 48(1):215–223, 2009. ISSN 0001-1452. doi: 10.2514/1.43469.

P. Singh, D. Deschrijver, and T. Dhaene. A BALANCED SEQUENTIAL DESIGN STRATEGY FOR GLOBAL SURROGATE MODELING. In *Winter Simulation Conference*, 2013.

S. F. Smith. *A learning system based on genetic adaptive algorithms*. PhD thesis, University of Pittsburgh, 1980.

A. Sóbester, S. J. Leary, and A. J. Keane. A parallel updating scheme for approximating and optimizing high fidelity computer simulations. *Structural and Multidisciplinary Optimization*, 27(5):371–383, 2004. ISSN 1615147X. doi: 10.1007/s00158-004-0397-9.

J. Sobieszczanski-Sobieski and R. T. Haftka. Multidisciplinary aerospace design optimization: Survey of recent developments. *Structural Optimization*, 14(1):1–23, 1997. ISSN 09344373. doi: 10.1007/BF01197554.

A. I. M. Sobol. On the Systematic Search in a Hypercube Published by : Society for Industrial and Applied Mathematics Stable URL : https://www.jstor.org/stable/2156633. 16(5):790–793, 1979.

I. M. Sobol. Distribution of points in a cube and approximate evaluation of integrals. *USSR Computational mathematics and mathematical physics*, 7:86–112, 1967.

D. H. Song and D. M. Tartakovsky. Transfer learning on multifidelity data. *Journal of Machine Learning for Modeling and Computing*, 3(1), 2022.

E. Spyromitros-Xioufis, G. Tsoumakas, W. Groves, and I. Vlahavas. Multi-target regression via input space expansion: treating targets as inputs. *Machine Learning*, 104(1):55–98, 2016. ISSN 15730565. doi: 10.1007/s10994-016-5546-z.

R. P. Starkey and M. J. Lewis. Analytical off-design lift-to-drag-ratio analysis for hypersonic waveriders. *Journal of Spacecraft and Rockets*, 37(5):684–691, 2000. ISSN 00224650. doi: 10.2514/2.3618.

R. B. Statnikov and J. B. Matusov. *Multicriteria optimization and engineering*. 2012.

J. Steck, K. Rokhsaz, J. Steck, and K. Rokhsaz. Some applications of artificial neural networks in modeling of nonlinear aerodynamics and flight dynamics. In *35th Aerospace Sciences Meeting and Exhibit*, page 338, 1997.

R. E. Steuer. Multiple criteria optimization. *Theory, Computation, and Application*, 1986.

G. Sun, G. Li, M. Stone, and Q. Li. A two-stage multi-fidelity optimization procedure for honeycomb-type cellular materials. *Computational Materials Science*, 49(3):500–511, 2010. ISSN 0927-0256.

H. J. Sussmann and J. C. Willems. 300 Years of Optimal Control: From The Brachystochrone to the Maximum Principle, 1997. ISSN 1066033X.

J. A. K. Suykens, J. P. L. Vandewalle, and B. L. de Moor. *Artificial neural networks for modelling and control of non-linear systems*. 1996. ISBN 1475724934.

K. Svanberg. The method of moving asymptotes—a new method for structural optimization. *International Journal for Numerical Methods in Engineering*, 24(2):359–373, 1987. ISSN 10970207. doi: 10.1002/nme.1620240207.

M. A. Taddy, H. K. Lee, G. A. Gray, and J. D. Griffin. Bayesian guided pattern search for robust local optimization. *Technometrics*, 51(4):389–401, 2009. ISSN 00401706. doi: 10.1198/TECH.2009.08007. URL https://doi.org/10.1198/TECH.2009.08007.

E.-G. Talbi. A taxonomy of hybrid metaheuristics. *Journal of heuristics*, 8:541–564, 2002. ISSN 1381-1231.

D. J. Toal. Some considerations regarding the use of multi-fidelity Kriging in the construction of surrogate models. *Structural and Multidisciplinary Optimization*, 51(6):1223–1245, 2015. ISSN 16151488. doi: 10.1007/s00158-014-1209-5.

D. J. Toal and A. J. Keane. Non-stationary kriging for design optimization. *Engineering Optimization*, 44(6):741–765, 2012. ISSN 0305215X. doi: 10.1080/0305215X.2011.607816.

V. Toropov, F. Va, and V. Markine. Refinements in the multi-point approximation method to reduce the effects of noisy structural responses. In *6th Symposium on Multidisciplinary Analysis and Optimization*, page 4087, 1996.

V. V. Toropov. Simulation approach to structural optimization. *Structural Optimization*, 1(1): 37–46, 1989. ISSN 09344373. doi: 10.1007/BF01743808.

J. D. Toscano, V. Oommen, A. J. Varghese, Z. Zou, N. Ahmadi Daryakenari, C. Wu, and G. E. Karniadakis. From pinns to pikans: Recent advances in physics-informed machine learning. *Machine Learning for Computational Science and Engineering*, 1(1):1–43, 2025.

N. Tsiantis and J. R. Banga. Using optimal control to understand complex metabolic pathways. *BMC Bioinformatics*, 21(1):1–33, 2020. ISSN 14712105. doi: 10.1186/s12859-020-03808-8.

C. J. Turner, R. H. Crawford, and M. I. Campbell. Multidimensional sequential sampling for NURBs-based metamodel development. *Engineering with Computers*, 23(3):155–174, 2007. ISSN 01770667. doi: 10.1007/s00366-006-0051-9.

M. Tyan and J. W. Lee. Efficient multi-response adaptive sampling algorithm for construction of variable-fidelity aerodynamic tables. *Chinese Journal of Aeronautics*, 32(3):547–558, 2019. ISSN 10009361. doi: 10.1016/j.cja.2018.12.012. URL https://doi.org/10.1016/j.cja.2018.12.012.

W. C. van Beers and J. P. Kleijnen. Customized sequential designs for random simulation experiments: Kriging metamodeling and bootstrapping. *European Journal of Operational Research*, 186(3):1099–1113, 2008. ISSN 03772217. doi: 10.1016/j.ejor.2007.02.035.

E. R. Van Dam, B. Husslage, D. Den Hertog, and H. Melissen. Maximin Latin hypercube designs in two dimensions. *Operations Research*, 55(1):158–169, 2007. ISSN 0030364X. doi: 10.1287/opre.1060.0317.

V. Vapnik. *The nature of statistical learning theory*. 1999.

M. Vasile. Finite elements in time: A direct transcription method for optimal control problems. *AIAA/AAS Astrodynamics Specialist Conference 2010*, (August 2010), 2010a. doi: 10.2514/6.2010-8275.

M. Vasile. Finite elements in time: A direct transcription method for optimal control problems. *AIAA/AAS Astrodynamics Specialist Conference 2010*, pages 1–25, 2010b. doi: 10.2514/6.2010-8275.

M. Vasile and F. Bernelli-Zazzera. Trajectory design combining low-thrust and gravity assist manoeuvres. In *Operations Research in Space and Air*, pages 203–229. Springer, 2003.

M. Vasile and A. Finzi. Direct lunar descent optimisation by finite elements in time approach. *International Journal of Mechanics and Control*, 1(1):1–15, 2000. ISSN 1590-8844.

M. Vasile and F. Zuiani. Multi-agent collaborative search: An agent-based memetic multi-objective optimization algorithm applied to space trajectory design. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, 225(11):1211–1227, 2011. ISSN 20413025. doi: 10.1177/0954410011410274.

S. Venkataraman and R. T. Haftka. Structural optimization complexity: What has Moore's law done for us? *Structural and Multidisciplinary Optimization*, 28(6):375–387, 2004. ISSN 1615147X. doi: 10.1007/s00158-004-0415-y.

S. Venkataraman, R. Haftka, and T. Johnson. Design of shell structures for buckling using correction response surface approximations. In *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, page 4855, 1998.

G. Venter, R. T. Haftka, and J. H. Starnes. Construction of response surface approximations for design optimization. *AIAA Journal*, 36(12):2242–2249, 1998. ISSN 00011452. doi: 10.2514/2.333.

N. Vertovec, S. Ober-Blobaum, and K. Margellos. Multi-objective minimum time optimal control for low-thrust trajectory design. *2021 European Control Conference, ECC 2021*, pages 1975–1980, 2021. doi: 10.23919/ECC54610.2021.9654919.

F. A. Viana and R. T. Haftka. Using multiple surrogates for minimization of the RMS error in meta-modeling. In *Proceedings of the ASME Design Engineering Technical Conference*, volume 1, pages 851–860, 2008. ISBN 9780791843253. doi: 10.1115/DETC2008-49240.

F. A. Viana, T. W. Simpson, V. Balabanov, and V. Toropov. Metamodeling in multidisciplinary design optimization: How far have we really come? *AIAA Journal*, 52(4):670–690, 2014. ISSN 00011452. doi: 10.2514/1.J052375.

R. Vitali, R. Haftka, and B. Sankar. Correction response surface approximations for stress intensity factors of a composite stiffened plate. In *39th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference and Exhibit*, page 2047, 1998.

R. Vitali, O. Park, R. T. Haftka, B. V. Sankar, and C. A. Rose. Structural optimization of a hat-stiffened panel using response surfaces. *Journal of aircraft*, 39(1):158–166, 2002. ISSN 0021-8669.

O. von Stryk and R. Bulirsch. Direct and indirect methods for trajectory optimization. *Annals of Operations Research*, 37(1):357–373, 1992. ISSN 02545330. doi: 10.1007/BF02071065.

G. Wahba. *Spline models for observational data*. SIAM, 1990.

D. R. Wallace, M. J. Jakiela, and W. C. Flowers. Design search under probabilistic specifications using genetic algorithms. *Computer-Aided Design*, 28(5):405–421, 1996.

R. Wallach, B. S. de Mattos, and R. da Mota Girardi. Aerodynamic coefficient prediction of a general transport aircraft using neural network. *ICAS-Secretariat - 25th Congress of the International Council of the Aeronautical Sciences 2006*, 2(January):1199–1214, 2006. doi: 10.2514/6.2006-658.

J. T. Walton. An overview of tested and analyzed ntp concepts. *AIAA/NASA/OAI Conference on Advanced SEI Technologies, 1991*, 1991. doi: 10.2514/6.1991-3503.

G. G. Wang and S. Shan. Review of metamodeling techniques in support of engineering design optimization. *Journal of Mechanical Design, Transactions of the ASME*, 129(4):370–380, 2007. ISSN 10500472. doi: 10.1115/1.2429697.

H. Wang, F. Ye, E. Li, and G. Li. A comparative study of expected improvement-assisted global optimization with different surrogates. *Engineering optimization*, 48(8):1432–1458, 2016.

H. Wang, Y. Jin, and J. Doherty. A Generic Test Suite for Evolutionary Multifidelity Optimization. *IEEE Transactions on Evolutionary Computation*, 22(6):836–850, 2018. ISSN 1089778X. doi: 10.1109/TEVC.2017.2758360.

Y. Wang and C.-Y. Lai. Multi-stage neural networks: Function approximator of machine precision. *Journal of Computational Physics*, 504:112865, 2024.

E. W. Weisstein. Least squares fitting–polynomial, 2002. URL https://mathworld.wolfram.com/.

B. Widrow and M. Hoff. Adaptive Switching Circuits. Technical Report May, 1960.

S. M. Wild and C. Shoemaker. Global convergence of radial basis function trust-region algorithms for derivative-free optimization. *SIAM Review*, 55(2):349–371, 2013. ISSN 00361445. doi: 10.1137/120902434.

E. A. Willis. Comparison of Trajectory Profiles and Nuclear-Propulsion-Module Arrangemenets for Manned Mars and Mars-Venus Missions. 1971.

R. L. Winkler. Combining Probability Distributions from Dependent Information Sources. *Management Science*, 27(4):479–488, 1981. ISSN 0025-1909. doi: 10.1287/mnsc.27.4.479.

D. H. Wolpert and W. G. Macready. No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, 1(1):67–82, 1997. ISSN 1089-778X.

P. D. Wooster, R. D. Braun, J. Ahn, and Z. R. Putnam. Mission design options for human Mars missions. *Mars*, 2007. ISSN 1548-1921. doi: 10.1555/mars.2007.0002.

B. A. Wujek, J. E. Renaud, S. M. Batill, and J. B. Brockman. Concurrent subspace optimization using design variable sharing in a distributed computing environment. *Concurrent Engineering Research and Applications*, 4(4):361–376, 1996. ISSN 1063293X. doi: 10.1177/1063293x9600400405.

S. Xiong, P. Z. Qian, and C. F. Wu. Sequential design and analysis of high-accuracy and low-accuracy computer codes. *Technometrics*, 55(1):37–46, 2013. ISSN 00401706. doi: 10.1080/00401706.2012.723572.

Y. Xiong, W. Chen, D. Apley, and X. Ding. A non-stationary covariance-based Kriging method for metamodelling in engineering design. *International Journal for Numerical Methods in Engineering*, 2006. doi: 10.1002/nme. URL http://onlinelibrary.wiley.com/doi/10.1002/nme.3279/full.

S. Xu, H. Liu, X. Wang, and X. Jiang. A robust error-pursuing sequential sampling approach for global metamodeling based on voronoi diagram and cross validation. *Journal of Mechanical Design, Transactions of the ASME*, 136(7), 2014. ISSN 10500472. doi: 10.1115/1.4027161.

P. Yang and R. Qi. Reentry trajectory optimization for hypersonic vehicle based on improved mesh refinement techniques. *Chinese Control Conference, CCC*, 2016-Augus(61374116): 5473–5478, 2016. ISSN 21612927. doi: 10.1109/ChiCC.2016.7554207.

X.-S. Yang. *Engineering optimization: an introduction with metaheuristic applications*. 2010.

W. Yao, X. Chen, and W. Luo. A gradient-based sequential radial basis function neural network modeling method. *Neural Computing and Applications*, 18(5):477–484, 2009. ISSN 09410643. doi: 10.1007/s00521-009-0249-z.

K. Q. Ye. Orthogonal Column Latin Hypercubes and Their Application in Computer Experiments. *Journal of the American Statistical Association*, 93(444):1430–1439, 1998. ISSN 1537274X. doi: 10.1080/01621459.1998.10473803.

R. Yondo, E. Andrés, and E. Valero. A review on design of experiments and surrogate models in aircraft real-time and many-query aerodynamic analyses. *Progress in Aerospace Sciences*, 96(November 2017):23–61, 2018. ISSN 03760421. doi: 10.1016/j.paerosci.2017.11.003.

M. Y. Yoo and J. H. Choi. Probabilistic calibration of computer model and application to reliability analysis of elasto-plastic insertion problem. *Transactions of the Korean Society of Mechanical Engineers, A*, 37(9):1133–1140, 2013. ISSN 22885226. doi: 10.3795/KSME-A.2013.37.9.1133.

Y. Yue and K. Meerbergen. Accelerating optimization of parametric linear systems by model order reduction. *SIAM Journal of Optimisation*, 23(2):1344–1370, 2013.

M. K. Zahir and Z. Gao. Variable-fidelity optimization with design space reduction. *Chinese Journal of Aeronautics*, 26(4):841–849, 2013. ISSN 10009361. doi: 10.1016/j.cja.2013.06.002. URL http://dx.doi.org/10.1016/j.cja.2013.06.002.

V. Zakirov and V. Pavshook. Russian Nuclear Rocket Engine Design for Mars Exploration. *Tsinghua Science and Technology*, 12(3):256–260, 2007. ISSN 10070214. doi: 10.1016/S1007-0214(07)70038-X.

R. S. Zebulum, M. A. Pacheco, and M. Vellasco. Variable length representation in evolutionary electronics. *Evolutionary computation*, 8(1):93–120, 2000a. ISSN 10636560. doi: 10.1162/106365600568112.

R. S. Zebulum, M. Vellasco, and M. A. Pacheco. Variable length representation in evolutionary electronics. *Evolutionary Computation*, 8(1):93–120, 2000b.

L. E. Zerpa, N. V. Queipo, S. Pintos, and J. L. Salager. An optimization methodology of alkaline-surfactant-polymer flooding processes using field scale numerical simulation and multiple surrogates. *Journal of Petroleum Science and Engineering*, 47(3-4):197–208, 2005. ISSN 09204105. doi: 10.1016/j.petrol.2005.03.002.

L. Zhang, Q. J. Zhang, and J. Wood. Statistical neuro-space mapping technique for large-signal modeling of nonlinear devices. *IEEE Transactions on Microwave Theory and Techniques*, 56 (11):2453–2467, 2008. ISSN 00189480. doi: 10.1109/TMTT.2008.2004894.

L. Zhang, Y. Wu, P. Jiang, S. K. Choi, and Q. Zhou. A multi-fidelity surrogate modeling approach for incorporating multiple non-hierarchical low-fidelity data. *Advanced Engineering Informatics*, 51(September 2021):101430, 2022. ISSN 14740346. doi: 10.1016/j.aei.2021.101430. URL https://doi.org/10.1016/j.aei.2021.101430.

M. Zhang, M. Tomac, C. Wang, and A. Rizzi. Variable fidelity methods and surrogate modeling of critical loads on x-31 aircraft. In *51st AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition*, page 1081, 2013.

Y. Zhang, T. N. Hoang, K. H. Low, and M. Kankanhalli. Near-Optimal active learning of multi-output Gaussian processes. *30th AAAI Conference on Artificial Intelligence, AAAI 2016*, pages 2351–2357, 2016. ISSN 2159-5399. doi: 10.1609/aaai.v30i1.10209.

J. Zhao and S. Li. Modified multiresolution technique for mesh refinement in numerical optimal control. *Journal of Guidance, Control, and Dynamics*, 40(12):3324–3334, 2017. ISSN 15333884. doi: 10.2514/1.G002796.

J. Zhao and S. Li. Mars atmospheric entry trajectory optimization with maximum parachute deployment altitude using adaptive mesh refinement. *Acta Astronautica*, 160(June 2018): 401–413, 2019. ISSN 00945765. doi: 10.1016/j.actaastro.2019.03.027.

J. Zhao and S. Li. Adaptive mesh refinement method for solving optimal control problems using interpolation error analysis and improved data compression. *Journal of the Franklin Institute*, 357(3):1603–1627, 2020. ISSN 00160032. doi: 10.1016/j.jfranklin.2019.11.024.

J. Zhao and T. Shang. Dynamic optimization using local collocation methods and improved multiresolution technique. *Applied Sciences (Switzerland)*, 8(9), 2018. ISSN 20763417. doi: 10.3390/app8091680.

Y. Zhao and P. Tsiotras. Density functions for mesh refinement in numerical optimal control. *Journal of Guidance, Control, and Dynamics*, 34(1):271–277, 2011. ISSN 15333884. doi: 10.2514/1.45852.

J. Zheng, X. Shao, L. Gao, P. Jiang, and Z. Li. A hybrid variable-fidelity global approximation modelling method combining tuned radial basis function base and kriging correction. *Journal of Engineering Design*, 24(8):604–622, 2013. ISSN 09544828. doi: 10.1080/09544828.2013.788135.

Q. Zhou, X. Shao, P. Jiang, Z. Gao, H. Zhou, and L. Shu. An active learning variable-fidelity metamodelling approach based on ensemble of metamodels and objective-oriented sequential sampling. *Journal of Engineering Design*, 27(4-6):205–231, 2016. ISSN 14661837. doi: 10.1080/09544828.2015.1135236.

Q. Zhou, Y. Wang, S. K. Choi, P. Jiang, X. Shao, and J. Hu. A sequential multi-fidelity metamodeling approach for data regression. *Knowledge-Based Systems*, 134:199–212, 2017. ISSN 09507051. doi: 10.1016/j.knosys.2017.07.033. URL http://dx.doi.org/10.1016/j.knosys.2017.07.033.

Q. Zhou, Y. Wang, S. K. Choi, P. Jiang, X. Shao, J. Hu, and L. Shu. A robust optimization approach based on multi-fidelity metamodel. *Structural and Multidisciplinary Optimization*, 57(2):775–797, 2018. ISSN 16151488. doi: 10.1007/s00158-017-1783-4.

Q. Zhou, J. Wu, T. Xue, and P. Jin. A two-stage adaptive multi-fidelity surrogate model-assisted multi-objective genetic algorithm for computationally expensive problems. *Engineering with Computers*, (0123456789), 2019. ISSN 14355663. doi: 10.1007/s00366-019-00844-8. URL https://doi.org/10.1007/s00366-019-00844-8.

Q. Zhou, Y. Wu, Z. Guo, J. Hu, and P. Jin. A generalized hierarchical co-Kriging model for multi-fidelity data fusion. *Structural and Multidisciplinary Optimization*, 62(4):1885–1904, 2020. ISSN 16151488. doi: 10.1007/s00158-020-02583-7.

Q. J. Zhu. Hamiltonian necessary conditions for a multiobjective optimal control problem with endpoint constraints. *SIAM Journal on Control and Optimization*, 39(1):97–112, 2000. ISSN 03630129. doi: 10.1137/S0363012999350821.

D. Zimmerman, C. Pavlik, A. Ruggles, and M. P. Armstrong. An Experimental Comparison of Ordinary and Universal Kriging and Inverse Distance Weighting. *Mathematical Geology*, 31 (4), 1999. ISSN 00138703. doi: 10.1023/A.

W. Zou, Q. Zhang, Q. Gao, and Z. Feng. Wavelet multi-resolution approximation for multiobjective optimal control. *PLoS ONE*, 13(8):1–13, 2018. ISSN 19326203. doi: 10.1371/journal.pone.0201514.

R. M. Zubrin, D. A. Baker, O. Gwynne, M. M. Astronautics, and P. O. Box. Mars Direct : A Simple , Robust , and Cost Effective Architecture for the Space Exploration Initiative. 1991.