UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING AND THE ENVIRONMENT
AERONAUTICS, ASTRONAUTICS AND COMPUTATIONAL ENGINEERING

Multidisciplinary and Multiobjective Design Optimisation of Coronary Stents

by

Sanjay Pant

A thesis submitted in partial fulfilment for the degree of Doctor of Philosophy

August, 2012

To all that is important for the survival of humanity.

Men say they know many things;
But lo! They have taken wings,—
The arts and the sciences,
And a thousand appliances;
The wind that blows
Is all that anybody knows.

- Henry David Thoreau (1817–1862).

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING AND THE ENVIRONMENT
AERONAUTICS, ASTRONAUTICS AND COMPUTATIONAL ENGINEERING

Doctor of Philosophy

MULTIDISCIPLINARY AND MULTIOBJECTIVE DESIGN OPTIMISATION OF CORONARY STENTS

by Sanjay Pant

Coronary stents are tubular type scaffolds that are deployed, using an inflatable balloon on a catheter, most commonly to recover the lumen size of narrowed (diseased) arterial segments. Even though numerous stent designs, of varying geometrical and material complexity, are used in clinical practice today, the adverse biological responses post-stenting are not completely eliminated. In-stent restenosis (IR), reduction in lumen size due to neointima formation within 12 months of procedure, and stent thrombosis (ST), formation of a blood clot inside a stented vessel, are the two most common adverse responses to stents. Such adverse responses are multifactorial and their causes are not completely understood. However, the geometric design of a stent, which is a common differentiating factor between the numerous commercially available stents, is known to be a key factor influencing adverse responses. In light of the above, this thesis exploits stent geometry parameterisation in both constrained and multiobjective optimisation. Gaussian process surrogate modelling is used to cost effectively (a) understand the influence of stent geometry parameters on metrics indicating adverse response, and (b) obtain families of stent designs which are potentially more resistant to such responses.

Various computational models are developed to evaluate the efficay of a stent in terms of the factors influencing the adverse responses. In particular, two finite element analysis (FEA) models and two computational fluid dynamics (CFD) models are developed. The FEA models are used to simulate the balloon-expansion of stents in a representative coronary artery and bending of stents on application of bending moments. On the other hand, the CFD models simulate haemodynamic flow in the stented artery and the associated drug-release into the tissue. The expansion FEA models are validated against manufacturer provided pressure-diameter relationship and the flexibility FEA models are validated against the numerical studies found in literature. The numerical models are then used to extract metrics which are related to the adverse responses. Six metrics are formulated: (i) acute recoil, which measures the radial strength of the stent; (ii) volume average stress, which measures potential arterial injury caused by the stenting procedure; (iii) haemodynamic low and reverse index, which measures the haemodynamic alteration relevant to IR; (iv) volume average drug, which measures the amount of anti-proliferative drug delivered into the tissue; (v) drug deviation, which measures the uniformity of drug-distribution in the tissue; and (vi) flexibility metric, which measures the deliverability of the stent. These metrics are then used to compare the performance of different geometric stent designs. Two parameterisation techniques – one for a generic ring and link topology of stents, and one for the commercial CYPHER (Cordis corporation, Johnson & Johnson company) – are proposed to study the effect of geometrical variation in stent design on the formulated metrics of efficacy. These techniques are then

combined with surrogate modelling to perform stent design optimisation studies and study the effect of stent geometry on the evaluation metrics. Finally, three paradigms to choose optimal stent designs from a set of non-dominated solutions, in terms of the evaluation metrics, are proposed, and optimal designs under such paradigms are identified.

The last part of this thesis concerns surrogate assisted optimisation, and is not specific to the problem of stent design. Here, the use of analytically available gradient information in widely used Kriging predictors is explored. A search algorithm to locate all stationary points of a Krig, using a combination of an iterative sequence of the Krig derivative and a low-discrepancy sequence is proposed.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Sanjay Pant, declare that the thesis entitled, 'Multidisciplinary and Multiobjective Design Optimisation of Coronary Stents' and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- Parts of this work have been published as:

  1. Pant, S., Limbert, G., Curzen, N. P., Bressloff, N. W. Multiobjective design optimisation of coronary stents. *Biomaterials*, 32:7755–7773, 2011.

  2. Pant, S., Bressloff, N. W., and Limbert, G., Geometry Parameterization and Multidisciplinary Constrained Optimisation of Coronary Stents. *Biomechanics and Modeling in Mechanobiology*, 11:61–82, 2012.

  3. Pant, S., Bressloff, N. W., Forrester, A. I. J. and Curzen, N. The influence of strut-connectors in stented vessels : A comparison of pulsatile flow through five different coronary stents. *Annals of Biomedical Engineering*, 38:1893–1907, 2010.

  4. Pant, S. and Bressloff, N. W., 2009, Effect of strut connectors on haemodynamics of stented vessels. *World Congress on Medical Physics and Biomedical Engineering, September 7 – 12, 2009, Munich, Germany, IFMBE Proceedings*, Volume 25/4, pages 680-683

Signed:

Date:

Acknowledgments

"THERE ARE ANGELS on this earth and they come in subtle forms"
- Lance Armstrong, in *It's not about the bike*

This space goes to all the ANGELS that have affected my life, at times in quite apparent forms and other times in the most subtle forms. I would like to acknowledge the influence of all such people on me and this work, not only during the three years of this thesis production, but all my life. I will inevitably miss a few names in the limited space that I have here, but if by any chance you find yourself reading this and do not find your name in what follows, remember that I have not forgotten you, and that your special contribution to this work and my life is most appreciated, and has left me indebted.

I will not even attempt to find words, because I know that language has not yet evolved as much, to be able to describe the perpetual support and love that I have received from my family. *Ma*, *Pa*, *Bhai*, and *Bhabhi* – you are the best people that I have in my life.

My sincere gratitude to Dr. Neil W. Bressloff, my supervisor, who has always been supportive of any ideas, no matter how weird, I have come up with during the course of this thesis. I am thankful for his patience, when things didn't work; his guidance, when things did work; and the general pleasantness of his conversations. I would also like to thank Dr. Georges Limbert, my co-supervisor, whose support and belief in me and this work has been nonpareil. I would also like to thank him for his expertise in continuum mechanics and finite element analysis, without which this work would not have been materialised. Professor Nick Curzen has been the person who has always put things into perspective, and reminded me, whenever I got lost in numbers (and various other things that engineers worry about), that *'the goal is to improve patient outcome'*. He is one of the most dedicated and professional surgeons that I have had the pleasure of knowing, and I thank him for teaching me things that a doctorate normally would not. Finally, I would like to thank Dr. Prasanth Nair, who introduced me to Gaussian Process modelling, whose expertise in optimisation has greatly enhanced this thesis, and who has been much more than just another academic at the university. Many thanks to Professor Andy J Keane, for generously letting me use the high performance compute cluster. Finally, thanks to Dr. Ivan I. Voutchkov for his help in IT matters and implementation of surrogate assisted multiobjective optimisation.

Thanks is also due to University of Southampton, for providing me the opportunity to conduct research in the university's premises, the then School of Engineering Sciences (SES)

and now Faculty of Engineering & the Environment (FEE) for providing me a studentship, and the Overseas Research Student Awards Scheme (ORSAS) for paying the fee difference between EU and overseas students. Special thanks to Rosalind Mizen and Alicja Ash for taking care of any administrative work that I have had in my research group with a smile.

Mr. James Douglas "Jim" Morrison said *"Friends can help each other. A true friend is someone who lets you have total freedom to be yourself - and especially to feel. Or, not feel. Whatever you happen to be feeling at the moment is fine with them. That's what real love amounts to – letting a person be what he really is."* This thesis would not have taken its current form, if it were not for my friends who have helped me beyond measure, and allowed me to be myself in all circumstances. I think I now understand, at least partly if not fully, what Mr. Morrison meant when he said the above. Joshua Jeeson Daniel, Aditya Deshpande, Alkin Nasuf, Moresh Wankhede, Dilesh Bhardwa, Athanasios Makrodimopoulos, Hamidreza Alidousti, Giorgos Ragousis, Dario Carugo, Adil Masroor, Iain Cooper, Sabrina Oliveira, Derek Eagan, Stephen Powell, James Parr, Florencia Angelica, Aniruddha Kaushal, Aditya Karnik, Hemant Singh, Louise Gilmour, Peter Pu, Aline Guitton, Lindsay Mabbutt, Claire Lin, Kay, Rosie Draisey, Bernard Dyke, and Christine Taylor: it has been a pleasure knowing you and spending time with you.

Thanks is also due to Mr. Jamie Wilson, my music teacher, who introduced me to the saxophone and the wonderful world of Jazz and Swing music. Thanks also to the many musicians, artists, writers, poets, and philosophers, whose works have kept me going in tough times and always given me immense joy along with food for thought. A final thanks to the SUSU shop, for the wonderful people who make delicious bakes and always greet you with a smile, Chef China, for the famous Chicken Aubergine and sam-pui-kai, Sanjha, for delicious curries, and the Crown Inn, for the best burgers in Highfield.

. . . and I think to myself, what a wonderful world.

List of Abbreviations

Abbreviation	Description	Definition
BMS	Bare metal stents	page 12
CABG	Coronary artery bypass surgery	page 9
CAD	Computer aided design; coronary artery disease	page 19,page 7
CFD	Computational fluid dynamics	page 49
CG	Conjugate gradient	page 35
CST	Class-shape-transformation	page 246
DB	Dogboning	page 90
DD	Drug distribution	page 18
DES	Drug eluting stents	page 12
DHC	Dynamic hill climbing	page 210
EA	Evolutionary algorithm	page 37
FEA	Finite element analysis	page 71
FM	Flexibility metric	page 105
FP	Fixed point	page 182
FS	Foreshortening	page 90
GA	Genetic algorithm	page 37
GP	Gaussian Process	page 28
HLRFI	Haemodynamic low and reverse flow index	page 66
IVUS	Intravscular ultrasound	page 74
KKT	Karush Kuhn Tucker conditions	page 39
KNRS	Krige-Newton-Raphson-Sobol algorithm	page 206

Abbreviation	Description	Definition
LAD	Left anterior descending coronary artery	page 8
LCA	Left Coronary Artery	page 8
LDV	Laser doppler velocimetry	page 55
LL	Late loss	page 11
LP	Linear Programming	page 39
MFP	Modified fixed point iteration	page 192
MLD	Minimal lumen diameter	page 11
MOSI	Modified oscillatory shear index	page 53
MPC	Multipoint constraint	page 100
NBI	Normal boundary intersection	page 44
NC	Normal constraint	page 45
NR	Newton-Raphson iteration	page 192
NSGA-II	Non-dominated sorting genetic algorithm	page 45
NURBS	Non-uniform rational basis splines	page 19
PTCA	Percutaneous transluminal coronary angioplasty	page 10
QP	Quadratic Programming	page 39
RBF	Radial basis function	page 26
RCA	Right Coronary Artery	page 8
SBX	Simulated binary crossover	page 38
SCVR	Standardised cross validated residual	page 32
SFP	Standard fixed point iteration	page 191
SMA	Shape memory alloy	page 13
SMC	Smooth muscle cells	page 9
SPEA2	Strenth Pareto evolutionary algorithm	page 45
SQP	Sequential quadratic programming	page 43
TLR	Target lesion revascularization	page 11
VAS	Volume averaged stress	page 96
VAD	Volume averaged drug	page 110

Abbreviation	Description	Definition
WSS	Wall shear stress	page 50

List of Symbols

Greek symbols

Symbol	Description	Definition
τ_w^m	Wall shear stress magnitude	page 55
τ_w^x	Axial wall shear stress	page 55
ρ	Fluid density	page 55
ρ	Spectral radius	page 184
μ	Fluid viscosity; initial shear modulus	page 55; page 80
λ_i	Principal stretches	page 80
ν	Poisson's ratio	page 79
σ	von Mises stress	page 96
χ	Curvature index	page 101
ϕ	Bending angle	page 101
β	Mean of a random field	page 29
σ_z^2	Variance of a Gaussian Process	page 29
θ_i	Hyperparameter in Gaussian process correlation function	page 30
$\boldsymbol{\theta}$	Hyperparameter vector for θ_i	page 30
Γ	Covariance function	page 29
$\boldsymbol{\Gamma}$	Covariance matrix	page 30
$\boldsymbol{\gamma}$	Covariance vector between a new point and the points used to build a Gaussian process model	page 31
κ	Iteration counter	-

Symbol	Description	Definition
$\mathfrak{r}$	Correlation vector between a new point and the points used to build a Gaussian process model	page 31
$\boldsymbol{\Theta}$	Diagonal hyperparameter matrix for $\boldsymbol{\theta}$	page 186
Υ_i	Main effect of i^{th} variable	page 32
$s^{(\kappa)}$	Search direction at κ^{th} iteration	page 34
η_c	SBX crossover parameter	page 212
η_m	Polynomial mutation parameter	page 212

Operators

Symbol	Description
∇	Del operator
$\nabla(\cdot)$	Gradient of $(\cdot)$
$\nabla.(\cdot)$	Divergence of $(\cdot)$
$\nabla^2(\cdot)$	Divergence of the gradient of $(\cdot)$
$[\cdot]^T$	Matrix transpose of $[\cdot]$
$[\cdot]^{-1}$	Matrix inverse of $[\cdot]$
$E[\cdot]$	Expectation operator
$\lvert\cdot\rvert$	Absolute value and vector norm
$\lVert\cdot\rVert$	Matrix norm

Roman symbols

Symbol	Description	Definition
$\bar{I_1}$	First invariant of the Cauchy-Green tensor	page 79
$\mathbf{1_p}$	Matrix of size $1 \times p$ whose all elements are equal to one	page 183
c_i	Drug concentration in an element/cell	page 110
C_l	Drug concentration in blood	page 142
C_t	Drug concentration in tissue	page 143

Symbol	Description	Definition
d	Average diameter of stent	page 84
D_{dev}	Standard deviation of drug distribution	page 8
$\mathbf{D_i}$	i^{th} row of the Decision variable difference matrix $\mathbf{D}$	page 183
D_l	Drug diffusivity in blood	page 142
D_t	Drug diffusivity in tissue	page 143
$\mathbf{D}$	Matrix of difference between a decision variables of a new point and the points used to build a Gauusian process model	page 186
DB	Dogboning	page 90
E	Young's modulus	page 79
$\mathbb{F}$	Deformation gradient	page 80
$\mathbf{F}'$	Jacobian matrix for Newton-Raphson fixed point iteration	page 191
FM	Flexibility metric	page 105
FS	Foreshortening	page 90
$g(\mathbf{x})$	Inequality constraint	page 24
$\mathbf{G}'$	Jacobian matrix for standard fixed point iteration	page 185
$h(\mathbf{x})$	Equality constraint	page 24
h_c	axial length of the circumferential rings of a stent	page 111
$\mathbf{H}$	Hessian matrix	page 34
$HLRFI$	Haemodynamic low and reverse flow index	page 66
$\mathbf{I}$	Identity matrix	page 189
J	Total volume ratio	page 80
$\mathbf{J}$	Jacobian matrix for modified fixed point iteration	page 189
K	Initial bulk modulus	page 80
k_p	Drug permeability	page 143
$\mathbf{K}$	Gram Matrix	page 27
$\mathcal{L}$	negative log likelihood; Lagrangian	page 30; page 38
L_{stent}	Length of semi-crimped stent	page 110

Symbol	Description	Definition
m_j	Exponential hyperparameter in Gaussian process correlation function	page 30
M	Moment applied	page 101
n_{cell}	number of volume cells	page 145
n_d	Dividing parameter in the modified fixed point iterative sequence	page 189
$\mathcal{N}$	Normal (Gaussian) process	page 29
p	Balloon pressure	page 86
p_1	first link parameter for flex connectors	page 111
p_2	second link parameter for flex connectors	page 111
P	Fluid pressure	page 55
q	Drug flux	page 109
R	Correlation function	page 29
$\mathbf{R}$	Correlation matrix	page 30
R_c	Percentage acute recoil	page 91
Re	Reynolds number	page 55
$Recoil$	Absolute Acute recoil	page 95
R_{stent}	Radius of semi-crimped stent	page 110
$\mathbb{S}$	SOBOL sequence	page 209
$\mathbf{S^i}$	i^{th} point in the SOBOL sequence, $\mathbb{S}$	page 209
S_i	Sensitivity index of i^{th} variable	page 33
T	Time-period of the caridac pulse	page 56
U	Strain energy density	page 79
$\mathbf{v}$	Fluid velocity vector	page 55
VAD	Volume averaged drug	page 110
VAS	Volume averaged stress	page 96
$\mathbf{w}$	Weight vector for Gaussian process predictor	page 31
$\mathbf{W}$	Diagonal weight matrix for $\mathbf{w}$	page 186
W_{strut}	Strut width (circumferential dimension) of a stent	page 111
x_i	i^{th} decision/design variable	page 24

Symbol	Description	Definition
$\mathbf{x}$	Decision/design vector	page 24
$X(t)$	Random variable, indexed by t	page 28
$\mathbf{X_i}$	i^{th} row of the decision variable matrix $\mathbf{X}$	page 183
$\mathbf{X}$	Matrix of decision variables of all the points used to build a Gausian process model	page 184
$y(\mathbf{x})$	Objective function	page 24
$\mathbf{Z}$	Diagonal correlation matrix	page 181

Chapter 1

Aims & objectives

Coronary stents are tubular, often mesh-like, structures which are deployed in diseased (stenosed) artery segments to provide a scaffolding feature that compresses atheromatus plaque, hence restoring luminal area and maintaining vessel patency. Despite the widespread clinical use of stents in cardiovascular intervention, the presence of such devices can cause adverse responses leading to fatality or to the need for further treatment. The most common unwanted responses of inflammation, in-stent restenosis and thrombosis, are multifactorial. In-stent restenosis is caused by a cascade of events triggered by vessel injury during the balloon angioplasty procedure whereas late thrombosis (usually associated with drug eluting stents) typically occurs as a result of incomplete healing whereby inhibition of intimal proliferation results in exposed parts of the stent providing ideal sites for thrombogenesis. Both patient-specific factors, such as the geometry and morphology of the disease, combine with procedural factors, such as the size, shape, material and other design properties of the stent, to induce such responses.

This thesis aims to evaluate the effect of stent design parameters on the factors that determine the severity of the aforementioned adverse biological responses, primarily in-stent restenosis. Using such evaluations, design optimisation studies are conducted to obtain a potentially optimum family of stent designs that are more resistant to the adverse responses.

1.1 Aims

Currently, an "ideal stent" – that recovers arterial shape with no adverse response – does not exist, even though, as a multi-billion dollar industry, stent design has witnessed a fairly rapid evolution from bare metal stents of increasing complexity, through shape memory alloy stents, polymer coated, drug eluting stents to biodegradable (or bioresorbable or bioabsorbable) stents made from polymers or corrodible metals. In recent years, drug eluting stents, which elute an anti-proliferative drug to suppress smooth muscle cell proliferation, have witnessed a major increase in popularity following early trials and approvals in 2002-2003, largely due to their effectiveness in reducing in-stent restenosis. However, more recently, late thrombosis has been identified following the discontinuation of anti-platelet therapy. Increasing concerns over late thrombosis with drug eluting stents has led to a significant return to the use of bare metal stents and to further impetus in the quest for improved alternatives.

The coupling between arterial injury, which triggers the adverse responses, blood flow, which leads to differential shear stress distribution on the artery wall, and the distribution of an anti-proliferative drug in the arterial tissue, all three as a function of stent geometry, constitute a complex multi-objective problem that is poorly understood. Thus, this thesis aims to determine how, for typical stents or a new contemplated design, geometric variation affects arterial injury, blood flow, and drug distribution. Moreover, flexibility of a stent, which is extremely important for deliverability, is also a function of stent-geometry. Once a given stent can be evaluated for the physical behaviour during and post implantation, measures relating to the efficacy of the stent in arresting in-stent restenosis and deliverability can be extracted. These measures can then be fed back to the geometrical design of the stent to improve stent performance. With this background, the aims of this thesis are

- to assess the performance, i.e. deliverability and resistance to in-stent restenosis, of coronary stents in relation to variations in geometric design;
- to use this assessment in order to find a family of stent designs that minimise the adverse responses and maximise deliverability.

The final part of this thesis deals with the development of an optimisation algorithm, the aim of which is

- to evaluate if the analytically available derivative information for the widely used Kriging predictor, can be used for effective search of the surrogate model.

1.2 Objectives

The measurable objectives of this thesis in relation to the aims outlined above are

1. to develop a computational fluid dynamics model in order to evaluate flow features in a stented coronary artery;
2. to formulate measures relating to in-stent restenosis from the above model so that stents can be compared on their haemodynamic performance;
3. to develop a finite element analysis model in order to evaluate the process of balloon-expansion of coronary stents;
4. to formulate measures relating to radial strength of a stent and arterial injury caused by the procedure from the above model, which are shown to be related to in-stent restenosis;
5. to develop a computational fluid dynamics model in order to evaluate the drug-distribution achieved by a stent in a stented coronary artery;
6. to formulate measures relating both the amount of drug delivered and uniformity of drug-distribution from the above model, which are both relevant to arresting in-stent restenosis;
7. to develop a finite element analysis model in order to evaluate flexibility of stents on application of bending loads;

8. to formulate a measure of flexibility of a stent from the above model;

9. to construct surrogate models for each of the above extracted measures of performance;

10. to study trends in the measures of performance by variations in stent geometry with the aid of surrogate models;

11. to propose a technique to parameterise stent geometrical design;

12. to conduct design optimisation studies – both constrained and multi-objective – to demonstrate design improvement in coronary stents.

All the components of the above mentioned objectives can be seen in Figure 1.1, where the optimisation methodology is depicted. In this figure, the boxes inside the dashed boundary represent the engineering analyses part of the objectives. The 'simulations' columns show the computational models that are developed and the 'physical quantity measured' columns show the corresponding attribute which is related to deliverability for flexibility and resistance to in-stent restenosis for all other quantities. The boxes in the non-dashed part show the optimisation loop. The loop starts with a parametric representation of the stent geometry, such that different values of the parameters result in different stent geometries. The design search space is defined by setting up appropriate bounds on such parameters. This design space is then sampled at a number of points, defined by a sampling plan. For each point in the sampling plan, engineering analysis is performed to evaluate the physical response. Surrogate models are constructed for each of the measured attributes, and a search of these surrogates is made. The results of the search are used to add more points for surrogate improvement using an *infill* criterion. The analyses for these added points is conducted and the results are used to update the surrogate. The resulting surrogates are searched again and this process is repeated until a satisfactory surrogate is constructed or the required design improvement has been obtained or the available computational budget has been exhausted.

As mentioned in section 1.1, the last part of this thesis concerns how the analytically available gradient information of a Kriging predictor (surrogate model) can be used for an effective search of the Krig. For this part of the thesis, the objectives are

1. to derive the equations for the derivative of a Kriging predictor;

2. to formulate an iterative sequence which can search all stationary points of a Krig;

3. to formulate an optimisation algorithm which combines the above iterative sequence with a low-discrepancy sequence for an effective search of the Krig.

1.3 Thesis overview

In this section an overview of the thesis is presented. This overview is divided into the contents of each chapter as follows:

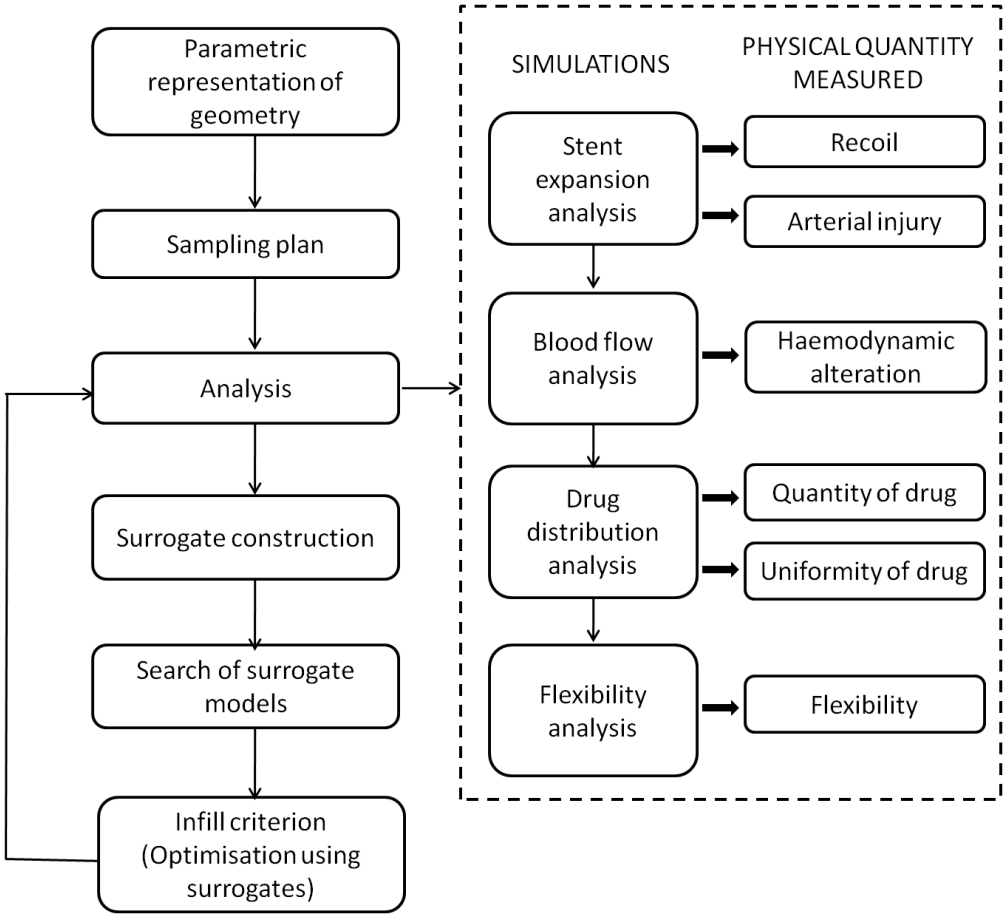


Figure 1.1: Optimisation methodology

1.3.1 Chapter 2

Chapter 2 introduces coronary artery disease. Anatomy and histology of the coronary artery is presented, and the available treatments for coronary artery disease are outlined. Thereafter, the two main adverse responses of coronary stenting: in-stent restenosis and thrombosis, are introduced. Finally, a classification of the variety of coronary stents available today are presented, and computer modelling approaches to create stent geometry are outlined.

1.3.2 Chapter 3

Chapter 3 presents an introduction to surrogate modelling and optimisation. For the former, the equations for a Gaussian process predictor are explained and for the latter, an overview of constrained, unconstrained, single objective, and multiobjective optimisation methods – both classical and evolutionary – is presented.

1.3.3 Chapter 4

In Chapter 4 haemodynamic evaluation of stents is presented. Pulsatile computational fluid dynamics (CFD) simulations are performed over five different coronary stents. Based on the results obtained, a numerical index to quantify the haemodynamic flow features that influence in-stent restenosis is formulated.

1.3.4 Chapter 5

Chapter 5 presents a model to simulate balloon expansion of a stent in a representative diseased artery using finite element analysis (FEA). These models are validated against the manufacturer provided experimental pressure-diameter relationship during the expansion of the CYPHER stent, Cordis corporation, Johnson & Johnson company. Using the results of the FEA analysis, arterial injury is quantified in a numerical index, and recoil is measured to evaluate the radial strength of the stent.

1.3.5 Chapter 6

Chapter 6 presents a finite element model to measure the flexibility of a stent, and proposes a numerical index to quantify flexibility in a numeric quantity which can be used to compare stents based on deliverability. The FEA model used in this chapter is validated against the numerical studies found in literature. In particular, a comparison of the moment-curvature plot for the CYPHER stent is made against the model of De Beule [11].

1.3.6 Chapter 7

Chapter 7 presents a constrained optimisation study. A parameterisation technique to create generic stent designs is proposed. A finite element model to evaluate drug-distribution is described. This model, combined with the analyses of chapters 5 and 6, and the proposed parameterisation, is used in a constrained optimisation study to obtain design improvement from the baseline geometry.

1.3.7 Chapter 8

Chapter 8 uses the expanded geometry obtained from Chapter 5. In this expanded geometry, first a haemodynamic analysis is performed (using the model developed in Chapter 4), and then a drug release simulation is performed. Unlike the drug release model of chapter 7, the model used in this chapter includes haemodynamic flow in the lumen. This chapter also proposes numerical indices to measure both the quantity of the drug transported to the tissue, and the uniformity of the resulting distribution.

1.3.8 Chapter 9

This chapter brings together the contents of chapters 4, 5, 6, and 8 in a multiobjective optimisation study for the CYPHER stent. A three parameter technique to represent CYPHER like stents is proposed and from the results of surrogate assisted multiobjective optimisation results, several conflicts between various pairs of desired attributes are shown. Features in the geometric design of stents which effect each of the measurable attributes are also identified. Finally, three paradigms to choose optimal stent designs from a set of non-dominated solutions are presented, and optimal stents under such paradigms are identified.

1.3.9 Chapter 10

Chapter 10 presents the development of a new optimisation algorithm for effective search of a Kriging predictor. The chapter presents how iterative sequences can be formed using the analytically available derivative information for the Kriging predictor, to locate stationary

points. Such a sequence is combined with the space-filling properties of quasi-random sequences to propose the Krige-Newton-Raphson-Sobol (KNRS) algorithm for both global and multimodal optimisation. Finally, the performance of this algorithm is compared with (i) a standard genetic algorithm and a dynamic hill climbing algorithm for global optimisation on 10 test-functions, and (ii) a fitness sharing genetic algorithm and a dynamic hill climbing algorithm for multimodal optimisation on five test functions.

1.3.10 Chapter 11

Chapter 11 concludes the thesis with a list of contributions made to (i) the areas coronary stent design, analysis, and optimisation, and (ii) the area of surrogate assisted search and optimisation. Finally, recommendations for further work in the aforementioned areas are made.

Chapter 2

Introduction to coronary artery disease and stents

2.1 Introduction

Coronary artery disease (CAD), also known as atherosclerotic heart disease, is a condition caused by the accumulation of lipids and fibrous tissue (collectively referred as *atherosclerotic plaque*) on the inner walls of a coronary artery [4]. This accumulation leads to narrowing of the arteries, thereby resulting in reduced blood flow to the downstream heart muscles (myocardium), and can consequently result in chest pain (*angina pectoris*) or heart attack (*myocardial infraction*).

CAD is a leading cause of death in western countries. According to the British Heart Foundation [21], in 2008, CAD was the cause of 88,000 deaths in the UK (one in five male and one in eight female deaths). Similarly, according to the American Heart Association [22], CAD caused 425,425 deaths in the United States of America, in 2006. Even though the treatment of CAD has evolved significantly in the past two decades, a treatment with no adverse effects does not yet exist. This chapter has the following aims:

1. to introduce CAD and its available treatments,
2. to present the adverse issues associated with the most common treatment (coronary stenting) for CAD,
3. to present a survey of coronary stent designs, outline the properties that are desirable in an *ideal stent*, and introduce the basic stent design problem, and
4. to present computer approaches for modelling the geometry of coronary stents.

2.2 CAD and its treatments

2.2.1 Circulation in the heart: anatomical features

Coronary circulation refers to the circulation that supplies oxygen-rich blood and nutrients to the myocardium, the muscle tissue of the heart. The vessels that supply blood to the

myocardium are called coronary arteries and can, in general, be classified as one of the following –

- Left Coronary Artery (LCA)
- Right Coronary Artery (RCA)

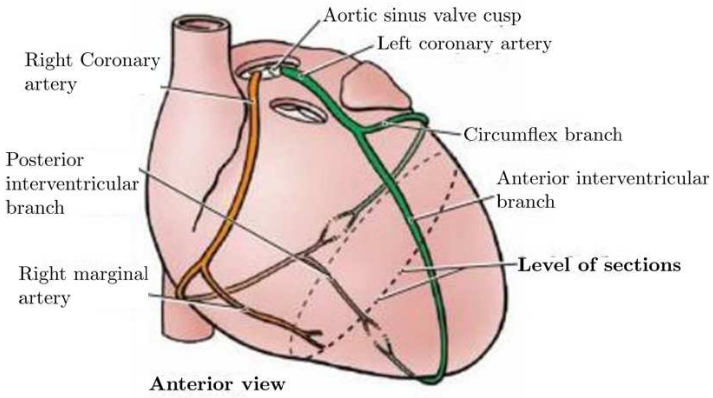


Figure 2.1: Coronary circulation: Left and Right Coronary Arteries [1]

Figure 2.1 shows the two coronary arteries, both beginning at the root of the aorta (LCA and RCA originating from the left and right aortic sinus, respectively) and travelling down forming a complex tree structure with numerous bifurcations. Shortly after its origination the LCA divides into two main branches: the anterior interventricular branch (also known as the left anterior descending (LAD) artery) and the circumflex branch. Similarly, the RCA divides into the right marginal artery and, in approximately 67% cases, into the posterior interventricular branch [1].

2.2.2 Coronary artery disease

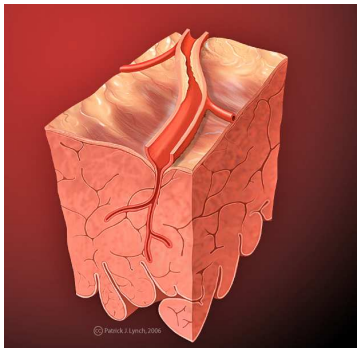


Figure 2.2: An illustration of plaque deposition inside a coronary artery [2]

Coronary artery disease refers to the condition when one or more branches, either the main branch or subsequent bifurcations, of the LCA and/or the RCA become narrowed (or get blocked) by gradual deposition of plaque. The deposition of plaque gradually causes the artery to harden, i.e. become less elastic. This phenomenon is called atherosclerosis. Plaque consists mainly of atheroma (composed of macrophage white blood cells), cholesterol, and calcium deposits. The deposition of plaque leads to a reduction of lumen area. This lumen area reduction, also known as stenosis, reduces the blood supply to the myocardium, leading to *angina pectoris*, chest pain, and sometimes to myocardial infraction (MI), or heart attack. MI is generally a result of the complete blockage of an artery, usually caused by a formation of blood clot (thrombus) over a ruptured plaque [23]. Figure 2.2 shows a picture of a coronary artery that has narrowed down due to the deposition of plaque.

Stenosis is detected with the help of angiography, an imaging technique used to visualize the lumen of an artery. In this process a radio-opaque agent, called contrast-agent, is injected into the blood and then visualized using X-ray based techniques. Figure 2.3 shows an angiogram where the stenosed region has been circled.

2.2.3 Histology of coronary artery

The artery wall has a complex structure composed of various layers with different mechanical properties. This makes the properties of the arterial tissue highly non-linear and anisotropic. In general, the artery wall is composed of the following layers, proportions of which differ in different parts of the circulation:

- The intima (inner)
- The media (middle)
- The adventitia (outer)

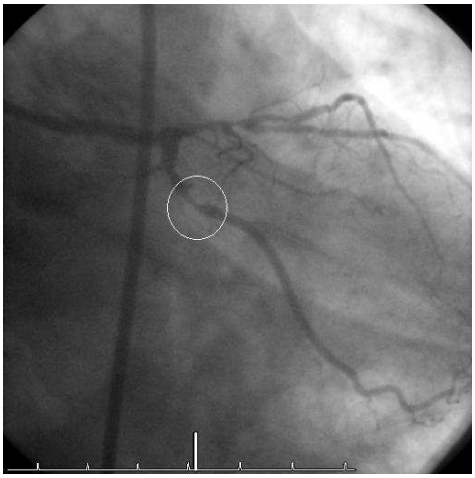


Figure 2.3: Angiogram of a coronary artery: the circle shows the stenosed region [3]

Figure 2.4 shows the three layers of the artery wall structure. The inner-most layer, intima, is composed of two layers: endothelium, which is a single layer of cells that acts as a barrier, and the internal elastic lamina that is composed of elastic fibres. The central layer, media, is the thickest layer composed of elastin, collagen, smooth muscle cells (SMC), and ground substance (glycosaminoglycans) [24]. The outer-most layer, adventitia, is composed largely of collagen I with admixed elastic, fibroplasts, and nerves [24]. The adventitia merges into the surrounding tissue thereby limiting the longitudinal movement of the artery.

2.2.4 Treatments for CAD

CAD, if not severe, can be treated by changes in lifestyle: healthy eating, low-saturated fat diet, regular exercise, and not smoking [23]. However, if the disease is severe then either coronary artery bypass graft (CABG) surgery, angioplasty, or angioplasty with stenting is used. The following sections describe each of these procedures.

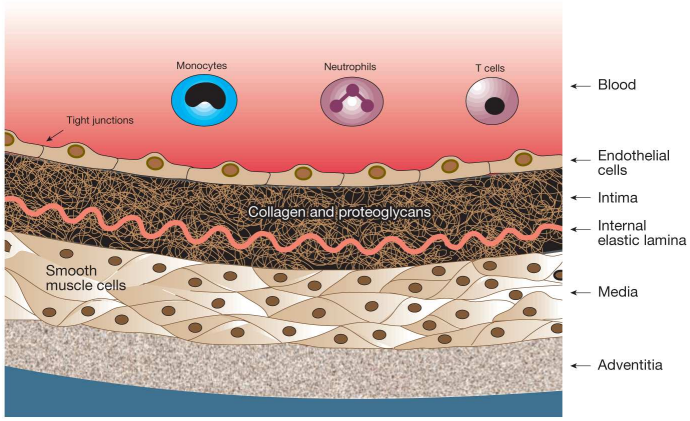
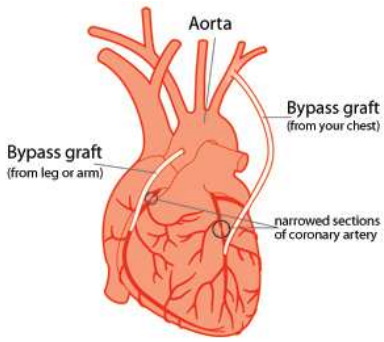


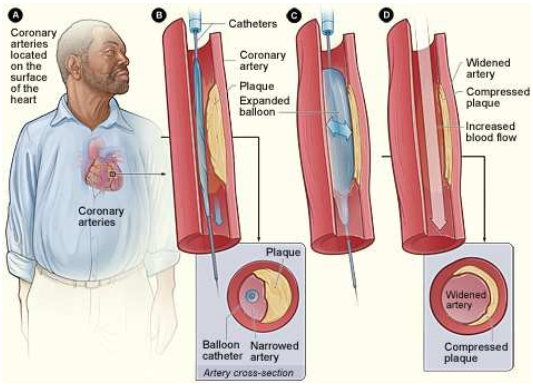
Figure 2.4: Artery wall structure: the three layers [4]

2.2.4.1 CABG

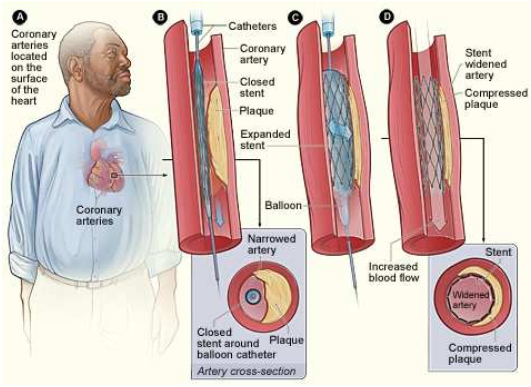
This is a surgical procedure in which arteries/veins from other parts of the body (usually the leg) are grafted into the coronary artery to bypass the narrowed (stenotic) region. The bypass graft is connected from the aorta to the post-stenotic region, thereby bypassing the blockage to maintain the downstream blood-supply. Figure 2.5a shows the CABG procedure. CABG, owing to its highly invasive nature is used only if the disease is severe or can't be treated with angioplasty/stenting.



(a) Coronary artery bypass graft surgery [25]



(b) Angioplasty: procedure [26]



(c) Stenting: procedure [26]

Figure 2.5: Treatment options for CAD

2.2.4.2 Angioplasty

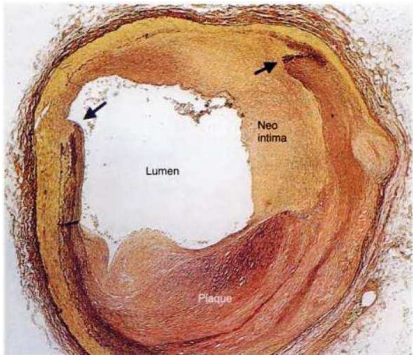
Percutaneous transluminal coronary angioplasty (PTCA), is a minimally invasive procedure in which a catheter, with a balloon mounted on the end, is inserted through the femoral/brachial arteries to the stenotic region. Once the catheter is positioned in the stenotic region, the balloon is inflated/deflated multiple times to compress the plaque against the artery wall. The catheter, along with the deflated balloon, is then withdrawn without leaving any permanent object inside the artery. Figure 2.5b shows the angioplasty procedure.

2.2.4.3 Stenting

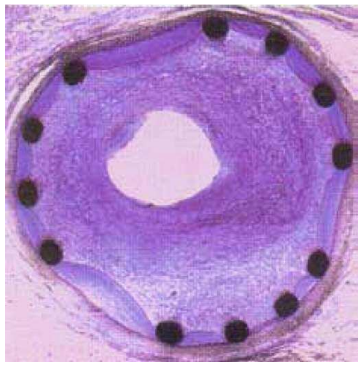
Coronary artery stents are tubular metal structures (often meshes) which are inserted in the stenotic region through a balloon catheter, usually after angioplasty, and then expanded until they deform plastically to provide scaffolding support that prevents arterial recoil. After the procedure, the metal stent remains inside the artery wall to prevent its recoil. Cells grow over the stent after the procedure, making it a permanent part of the artery. Figure 2.5c shows the procedure of stenting.

2.3 Issues with stenting: restenosis and thrombosis

The two most common issues that the use of angioplasty, with or without stenting, face today are restenosis and thrombosis. The following sections describe both these issues.



(a) Restenosis in an artery cross section after angioplasty [5]



(b) Restenosis in an artery cross section after angioplasty followed by stenting [30]

Figure 2.6: Restenosis

2.3.1 Restenosis

Although balloon angioplasty and stenting are widely used procedures today, restenosis continues to be a major problem associated with it. Dangas and Fuster [27] define restenosis as the reduction in lumen size at the cite of an angioplasty/stenting procedure. Restenosis is a result of arterial damage that leads to the formation and proliferation of neointima, a new thick layer of intima, at the procedure cite and occurs in 40-50% of cases within six months of the procedure [27, 28]. Figure 2.6 shows the occurrence of restenosis in coronary arteries. Although, the advent of drug eluting stents (see section 2.4 for types of stents) has significantly reduced the rates of restenosis to a level just above 10%, its presence can not be neglected as the number of patients treated with drug eluting stents is large [28]. Restenosis is a complex multifactorial biological process, the causes and mechanism of which are not completely understood [28, 29]. However, there are several factors which have been identified to contribute towards restenosis. These factors are discussed individually in future chapters. Restenosis is usually measured in the following three ways:

- Angiographically (binary restenosis)
- Clinically (target lesion revascularization (TLR))
- Late loss (LL)

Angiographic (binary) restenosis refers to more than 50% diameter stenosis at follow-up. TLR is defined as clinically driven repeat percutaneous intervention (PCI) of the lesion. It is driven by clinical signs of ischemia, reduced downstream blood-flow. TLR is most relevant to the patients as it reflects the risk of them needing a repeat interventional procedure [5]. Late loss, measured in mm, is the most quantitative definition of the restenosis rate. It is defined as

$$\text{Late Loss} = (\text{MLD immediately after procedure}) - (\text{MLD at follow-up})$$

where MLD denotes minimal lumen diameter.

2.3.2 Thrombosis

Thrombosis refers to the formation of a blood clot, thrombus, inside a blood vessel. The presence of a thrombus may either reduce the supply of downstream blood or completely occlude the blood vessel. The extent of non-occlusive thrombosis depends on the extent of vessel injury [5]. It is a result of incomplete healing where the exposed parts of the stent, or parts of ruptured plaque, provide ideal sites for thrombogenesis. Figure 2.7 shows a picture of thrombus along with the lumen, neo-intima, and the plaque on a cross section of a human artery. Rabbat et. al. [31] identify several procedural and patient specific risk-factors that contribute towards thrombosis. Late thrombosis (occurring after 30 days of stent implantation) [31] is generally associated with drug eluting stents, which elute an anti-proliferative drug into the arterial tissue to prevent restenosis. As opposed to the bare metal stents, drug eluting stents (see section 2.4 for different types of stents) delay the process of endothelialisation [32], the process of generation of endothelial cell layer post stenting procedure, and can trigger a thrombogenic response leading to late thrombosis [33].

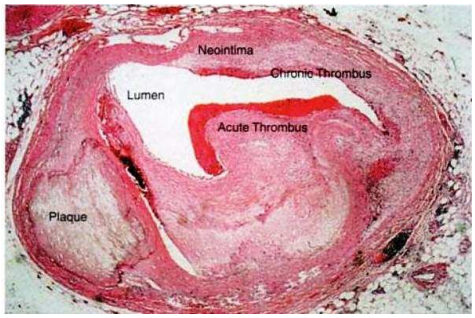


Figure 2.7: Thrombosis: artery cross section of a patient that died after 10 months of balloon angioplasty [5]

2.4 Stents: classification and desirable properties

This section presents a classification of stents based on various design parameters such as materials, geometry etc. After discussing the classification of stents, the properties that are desirable in an *ideal stent* are outlined.

Stents can broadly be classified as bare metal stents (BMS), drug eluting stents (DES), and bioabsorbable stents. BMS are made of metal only, and they may or may not have a biocompatible polymer coating. DES, on the other hand, necessarily have a drug coating, which is most commonly bound within a polymer. The two most commonly used drugs are sirolimus and paclitaxel. Both of these drugs are anti-proliferative which means that they interfere with the cell growth/division cycle [30] and hence help in reducing restenosis. Biosbsorbable stents are those made up of biodegradable materials that gradually degrade in roughly 12 months after the implant procedure depending on the type of biodegradable material used. They may or may not have a drug coating. Early results showed vigorous inflammatory response to bioabsorbable stents, but active research is currently being undertaken to develop bioabsorbable stents [33].

Apart from the above mentioned broad classification stents can further be categorized according to various properties. The *handbook of coronary stents* [34] lists the details of 43 commercially available stents. Stoeckel et. al. [6], in 2002, classified nearly 100 different commercially available stents to differentiate them by their engineering properties. The following sub-sections detail their classification with some modifications.

2.4.1 Nature of expansion

The most obvious classification of stents is based on the nature of expansion. Stents can either be balloon expandable or self expanding. Balloon expandable stents are made of metal, usually stainless steel (316L), or alloys, such as platinum-chromium or cobalt-chromium, that can plastically deform through balloon inflation. Self-expanding stents, on the other hand, either rely on the elastic properties of the metal or are made up of shape memory alloys (SMA), such as Nitinol (Nickel-Titanium), which can expand autonomously after release from the delivery system.

2.4.2 Materials

The material of the stent depends on the nature of expansion and its bio-compatibility. While most balloon-expandable stents are made of 316L stainless steel, a majority of self-expandable stents are made of Nitinol. 316L stainless steel is a corrosion resistant material with low carbon content with additions of molybdenum and niobium. Nitinol is an alloy composed of 55% weight percent nickel and 45% titanium. In addition to stainless steel and nitinol there are a number of other materials used to manufacture stents as tabulated in table 2.1. Recently, various metal alloys have emerged as a good alternative to stainless steel 316L, for e.g. the latest Boston Scientific's Promus^{TM} Element^{TM}coronary stent is made from a Platinum-Chromium alloy and Medtronic's Integrity stent is made from a cobalt-chromium alloy. The advantage of using alloys is that they allow relatively thinner stent struts without compromising structural strength.

Table 2.1: Stents classification: materials

Type	Material	Example
Balloon expandable	Stainless steel (316L)	Bx VELOCITY stent
	Tantalum	Wiktor
	Martensitic Nitinol	Paragon
	Paladium Iridium	Angio stent
	Polymers	Ingaki-Tamai stent
	Niobium alloy Iridium	Lunar StarFlex
	Cobalt-Chromium alloys	Integrity and Xience stents
	Platinum Chromium alloy	Promus ELEMENT stent
Self Expanding	Super elastic Nickel-Titanium (Nitinol)	Cordis SMART
	Cobalt alloy Iridium	Wallstent
	Full Hard Stainless Steel	Cook Z-Stent

2.4.3 Manufacturing form

Stents can be made from sheet metal, wires, or slotted tubes. For sheet-metal stents, the pattern is made on the sheet which is then rolled to form a tubular structure. Alternatively, wires can be knitted or braided together to form tubular meshes. The majority of stents available today are made from tubes, which are laser-cut to carve specific patterns on the tube. Table 2.2 classifies the stents based on their form.

Table 2.2: Stents classification: form

Type	Description	Example
Sheet	rolled to make a stent	NIR stent
Wire	knitted or braided together	Wallstent
Slotted-Tube	laser-cut	Bx VELOCITY stent

2.4.4 Fabrication method

Depending on the form of the stent, different fabrication methods can be used for manufacturing. Conventional wire-forming techniques like coiling, knitting or braiding are used to form stents with wires. All coil stents available today are self-expanding and made of Nitinol. Sometimes the wires after coiling are welded at certain locations to produce closed-cell wire stents (e.g. the Symphony stent). For slotted-tube stents, laser cutting is typically used. Balloon expandable stents are usually laser-cut in a crimped or near-crimped state and then surface treated (for example electropolished). Alternatively, waterjet cutting can be used for cutting out tubes (e.g. SCS stainless steel stent). This process does not produce a heat-affected zone along the cutting edge like that produced in the process of laser cutting. Lastly, photochemical etching can also be used to manufacture stents. This process is currently used to produce stents from tubing, but is also applicable in sheet processing to produce a large number of parts in a single run.

2.4.5 Geometry

Classification of stents based on geometry is the most interesting aspect of stent design. A vast variety of stent designs are available today with contrasting geometrical features. One of the main objectives of this thesis is to identify geometric properties that lead to better results for restenosis rates. Initial stent designs started with simplistic geometries/patterns, which over a period of time have evolved into more complex shapes. The following high level categories were used for geometrical classification of stents by Stoeckel et. al. [6]

2.4.5.1 Coil

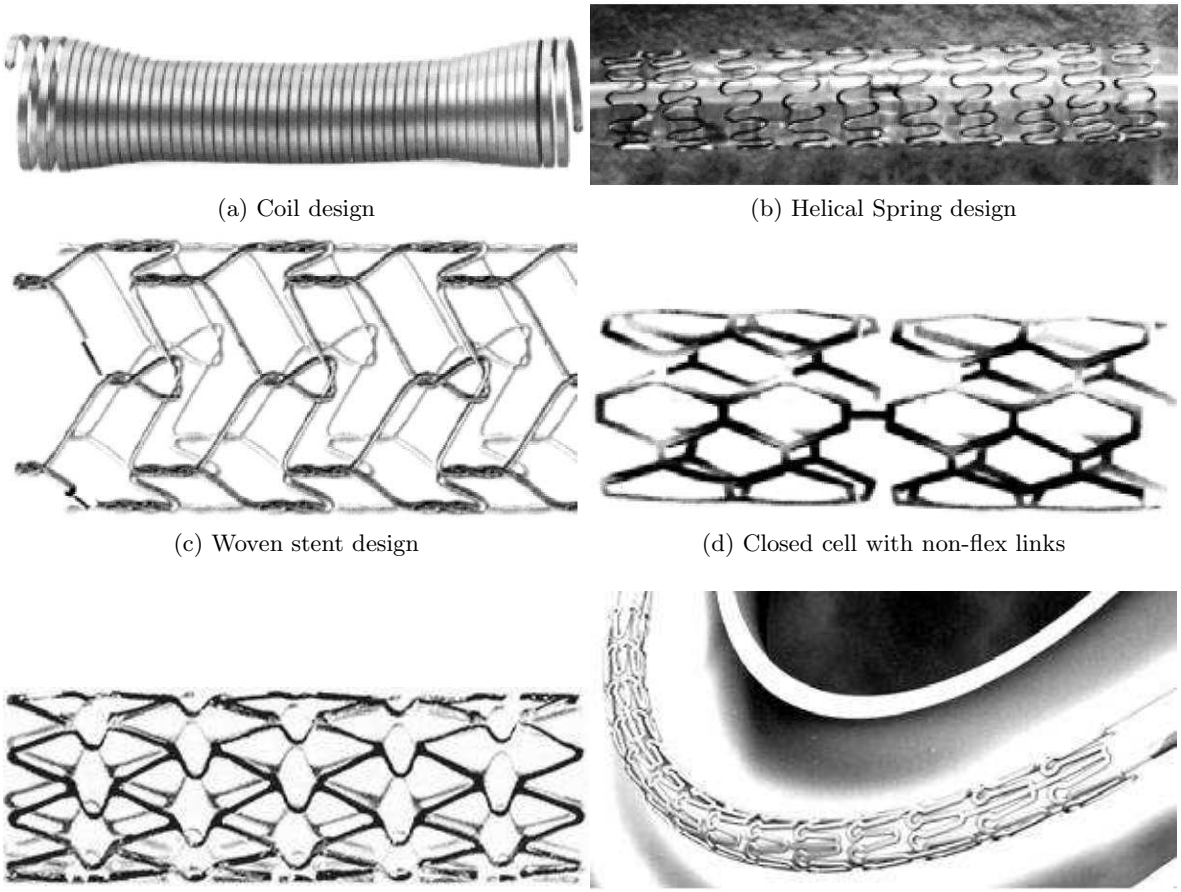
Coil design is most common in non-vascular applications as a coil stent can be retrieved after implantation. Coil stents are extremely flexible. However, their strength is limited and they have a low expansion ratio. Figure 2.8a shows the Esophacoil device with a coil design.

2.4.5.2 Helical spiral

These are helix shaped stents with no or minimal connections. Helical designs produce highly flexible stents but compromise on longitudinal support. Internal connections help the longitudinal stability by compromising on flexibility. Figure 2.8b shows the Crossflex stent with a helical spiral design.

2.4.5.3 Woven

Woven stents are typically wire stents which have been knitted/braided together. Self-expanding stents are often made of nitinol wires. Woven stents provide excellent wall coverage but typically shorten during expansion. Moreover, their radial strength is highly dependent



(a) Coil design

(b) Helical Spring design

(c) Woven stent design

(d) Closed cell with non-flex links

(e) Closed cell with flex links

(f) Open cell design

Figure 2.8: Stent classification based on geometry [6]

on axial fixation at the ends. Figure 2.8c shows the Cook ZA stent which has a knitted nitinol wire design.

2.4.5.4 Sequential rings

Sequential rings can be joined together at various points to make a stent. The rings can take various shapes, the most common of which is a sequence of zig-zag Z-shaped elements to form a ring. These rings can have various types of connections between them to form a complete mesh. Commonly found connections are:

- Regular connections - connections at each tip of the Z-shaped elements
- Periodic connections - connections at a subset of the tips of Z-shaped elements which repeat perodically
- Peak-peak or peak-valley connections

These connectors either are straight segments (non-flex connectors) or flexible (flex connectors). Another way to consider the sequential ring stents, more commonly applicable to slotted-tube stents, is if they are closed cell or open cell:

- Closed cell designs - These are designs where all the tips of the structural members are connected by bridging elements/connectors. Connectors can be either flex or non-flex,

but the combination should make closed cells. Figure 2.8d and figure 2.8e shows stent designs with a non-flex and flex connector respectively.

- Open cell designs - These are designs produced by closed cell structures by eliminating some/all of the bridges to open-up the closed cells. Open cell stents are usually more flexible than their closed-cell counterparts because the unconnected elements increase longitudinal flexibility. This category also entails stents which have no tip-tip connections but connectors originating from the middle of the struts (eg. BeStent). Figure 2.8f shows a picture of a stent with open-cell architecture.

2.4.6 Additions

A number of enhancements are added to stents, to improve their performance or visibility. The following are the most commonly used enhancements:

- Radio-opaque markers - Gold markers to improve the visibility for stent delivery and follow-up diagnosis
- Radio opaque coating - Gold or silicon-carbide coating to improve visibility
- Biocompatibility coating - Coatings of tantalum, phosphorylcholine, carbon, or silicon-carbide

2.4.7 Based on major stent manufacturers

Stents can also be classified according the major manufacturers. Currently, the following are the major manufacturers operating in the stent market:

- Boston Scientific
- Abbott Vascular
- Medtronic
- Biosensors

In terms of engineering properties, the most widely studied drug eluting stent is the CYPHER stent, manufactured until 2011[1] by Cordis Corporation, Johnson & Johnson company. CYPHER, although now discontinued, is a sirolimus eluting stent on the Bx VELOCITY stent platform, i.e. the geometric shape of the Bx VELOCITY bare metal stent. Figure 2.9a shows a picture of the CYPHER stent. It has a closed cell design with 'n' shaped flex connectors. In contrast to the CYPHER stent, Boston Scientific's TAXUS Liberté stent is an open cell design and is a paclitaxel eluting stent. Figure 2.9b shows the TAXUS Liberté stent. Similarly, Boston Scientific's Promus ELEMENT stent, an open cell everolimus (derivative of sirolimus) eluting stent, is shown in Figure 2.9c. Xience V is the main DES produced by Abbott Vascular (shown in Figure 2.9d). It has an open cell design and elutes everolimus. Medtronic has a BMS stent, called the Integrity stent, and a DES, called the Resolute Integrity stent. The Resolute Integrity stent is based on the Integrity stent platform, and elutes zotarolimus, a synthetic derivative of sirolimus. The geometric platform

[1] Johnson & Johnson announced in June 2011 [35] that Cordis corporation will stop the production of CYPHER stent owing to their focus on other areas of the interventional cardiology market.

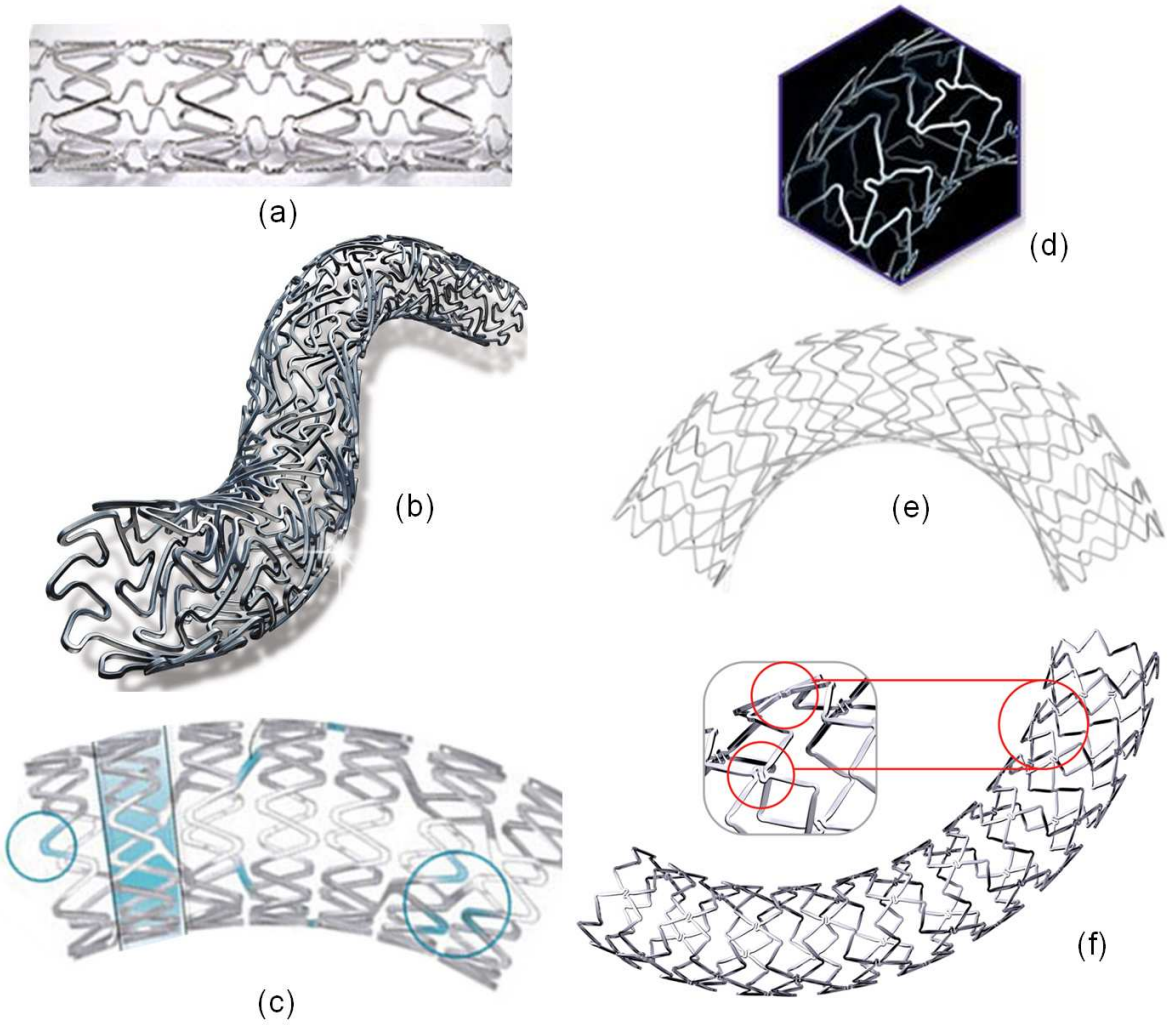


Figure 2.9: Contemporary stent designs [7, 8]

of these stents, which are made from a single wire in a sinusoidal form, is shown in Figure 2.9e. Biosensors' BioMatrix Flex stent, shown in Figure 2.9f, is a Biolimus A9 drug eluting stent. It has a unique stent design with quadrature S-shaped links between the Z-shaped crowns/rings. The quadrature link comprises two links per band that are axially rotated 90^o between successive crowns. A number of other manufacturers are active in research and clinical trials. Each stent manufacturer, however, has a unique stent design and a unique drug delivery method (polymer coating and drug combination) which is characteristic of the stent-manufacturer.

2.5 Properties of an ideal coronary stent

An ideal stent can be defined as: *'A stent which is easy to deliver, provides adequate arterial support, and minimises the associated adverse processes of restenosis and thrombosis, both in the short and long term.'*. This definition, although easy to understand in a general sense, provides many challenges in terms of what is precisely meant by 'ease of deliveability', 'adequate arterial support', and most importantly what is it in a stent that would minimise the 'associated adverse responses'. Consequently, there is a need to define these desirable attributes in measurable engineering terms. An ideal stent should

1. be *flexible*
2. have *high radial stength*
3. *minimise arterial injury*
4. *minimise hemodynamic alteration*
5. provide *adequate drug delivery*
6. facilitate *uniform drug distribution* (DD)

The first point of flexibility stems from ease of deliverability that is desired in an ideal stent. The process of delivering a stent to the stenotic site involves manoeuvring through arteries which can be highly curved and tortuous, thereby necessitating the need for high flexibility. High radial strength is analogous to provision of adequate arterial support. It is important to meet the very idea of stent invention i.e. to prevent arterial recoil. The rest of the four properties listed above relate to limiting the adverse responses of restenosis and thrombosis. This relation between these properties and adverse responses are briefly discussed here, but are presented in detail in future chapters. Arterial injury caused during the stenting procedure can be directly correlated with restenosis rates (see chapter 5). This implies that an ideal stent should minimise the injury caused during deployment. Several studies (see Chapter 4) have showed a link between altered haemodynamics in stented vessels and restenosis rates. Consequently, a good stent should alter the haemodynamics minimally. Since DES rely on an anti-proliferative drug to inhibit restenosis, a good stent should ensure that adequate drug is delivered in the tissue. Moreover, depending on the toxic-to-therapeutic ratio of the drug used, the drug distribution should be uniform across the tissue surrounding the coronary lesion.

As will be discussed in future chapters, the geometrical features of a stent design dictate all the aforementioned features. This leads to the conclusion that it should be possible to alter stent geometry to improve the aforementioned properties, and consequently minimise adverse responses. However, the consideration of numerous conflicting factors while designing a stent presents a major challenge. A change in stent geometry leading to an improvement in one of the desirable characteristics often leads to degradation in one or more of the other characteristics. As a result, the consideration of all the desirable characteristics, all originating from one stent design, lead to a very complex multi-objective and multi-disciplinary design problem.

A major part of this thesis deals with this design problem. In future chapters, the aforementioned desirable properties are quantified, so that given two geometrically different designs a judgement regarding the superiority of one over the other, in terms of a particular desired characteristic can be made. Thereafter, two studies, one based on a constrained optimisation formulation and one based on a multiobjective formulation, are presented as potential solutions to the stent design problem. Before concluding this chapter, various approaches that are adopted to create computer aided design (CAD) models of stents is briefly presented in the next section.

2.6 CAD geometry construction

The first step to evaluate any property of a stent computationally is the construction of its computer aided design (CAD) geometry. Rhinoceros 4.0 (Robert McNeel & Associates), a commercially available NURBS (Non-uniform rational basis splines) based modelling software, is used for this purpose. The following subsections describe various approaches that can be employed to construct full 3-D stent geometries.

2.6.1 Approaches to model a stent

Figure 2.9 shows a few contemporary stent designs. Most stent designs used today in clinical practice today are slotted-tube type. Three approaches can be used to model such stent geometries:

- **Approach 1:** The base geometry in this approach is a cylindrical shell with the required thickness of the stent. Cell patterns are then cut out (boolean difference) from this base geometry to obtain the stent geometry.

- **Approach 2:** In this approach the base geometry is a plane sheet of required thickness. The cell patterns, like in approach 1, are cut out from this plane sheet, and the resulting structure wrapped/mapped on to a cylindrical shell to obtain the final stent geometry.

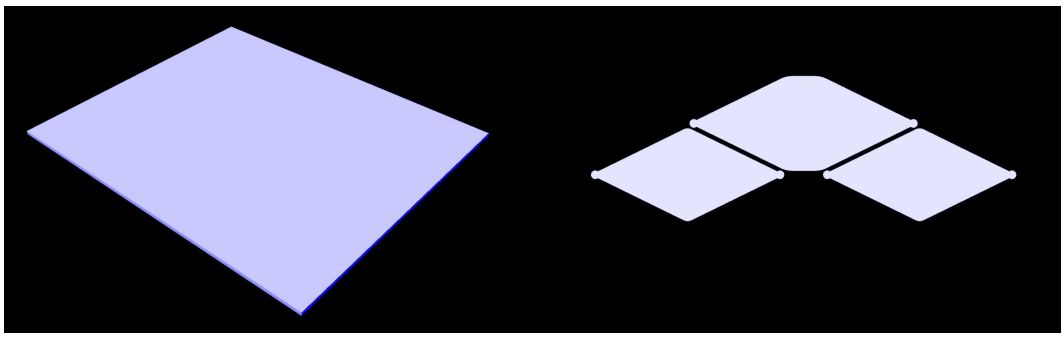
- **Approach 3:** This approach uses curves (splines or NURBS) to model the shape of the stent struts on a flat plane. The resulting network of curves is then converted into a closed surface, extruded to the required height, and mapped on to a cylindrical shell to obtain the final stent geometry.

Depending on the stent design, one of the above methods can be used. If the pattern is easy to construct and periodically repeating, then either of the first two methods can be used. However, if the patterns in the stent design are not so apparent, and the shape consists of non-periodic or complex shapes, the third option proves very powerful to construct 3-D models.

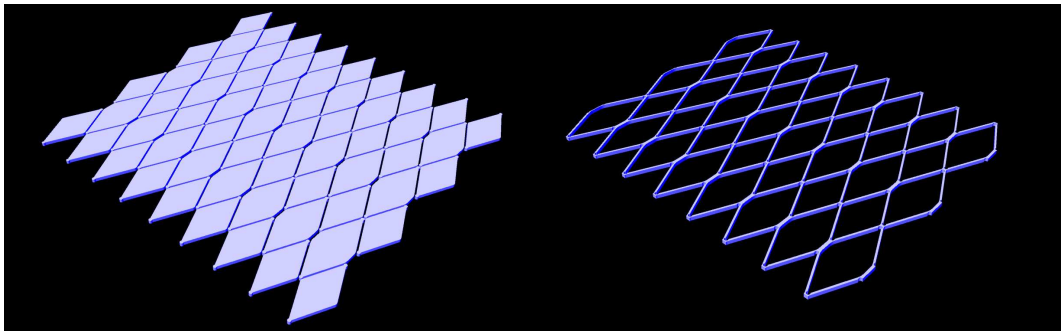
Approaches 1 and 2 are similar, but it is more difficult to cut the patterns out from a cylindrical shell than from a flat sheet/plate. Hence, either of approaches 2 and 3 are used for constructing stent geometries in future chapters. The following subsections describe approaches 2 and 3 in more detail.

2.6.1.1 Approach 2

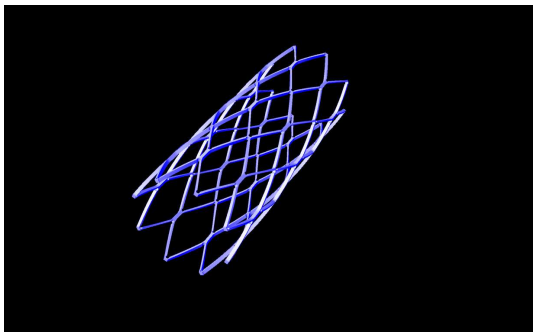
Figure 2.10a shows the starting geometry, a flat plate, used in this approach. A periodic pattern similar to the Palmaz-Schatz stent is created (Figure 2.10b). This pattern is a solid which can be made by outlining the pattern with closed curves (usually NURBS), converting them to a surface, and then extruding the surface to the required height. This pattern is then repeated periodically to fill in the base plate (see Figure 2.10c). After this a boolean subtraction operation is performed on the flat plate with the solid patterns, which results in the flat geometry of the stent (Figure 2.10d). This flat structure can then be rolled around a cylinder to obtain the final stent model, as shown in Figure 2.10e.



(a) The flat plate (b) Cutting Pattern



(c) Array of the cutting patterns (d) Stent in its flat configuration (a)-(c)



(e) Flat stent rolled to get a cylindrical shape

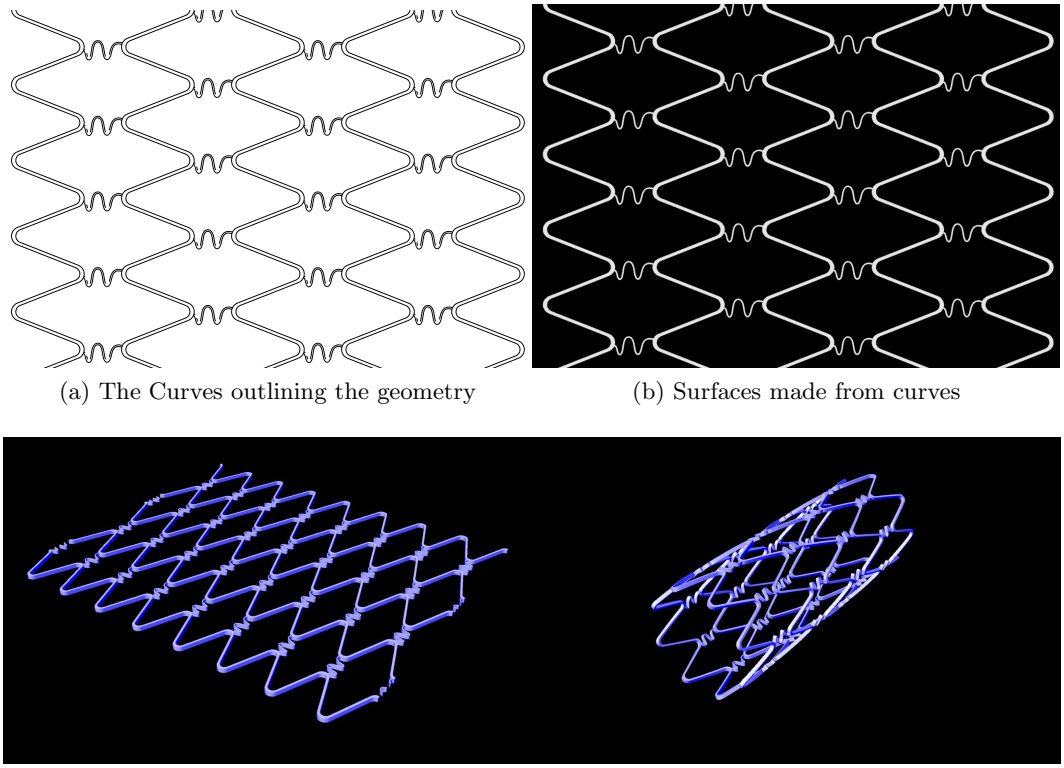
Figure 2.10: Approach 2 to construct stent geometry

2.6.1.2 Approach 3

This approach requires the stent struts to be modelled with the help of curves, splines or NURBS, to create a network of curves(Figure 2.11a). This network is then closed and converted into a surface or a collection of surfaces (Figure 2.11b). Thereafter these surfaces are extruded to a height equal to the required strut thickness (Figure 2.11c) to obtain the flat geometry of the stent. This flat geometry can then be wrapped around a cylinder to obtain the final stent (Figure 2.11d).
Figure 2.12 shows 3-D stent models constructed using the three approaches mentioned above.

2.6.2 Pre-crimped state models

The approaches discussed above show how the final expanded state geometries can be created. In reality stents are manufactured, for example laser cut, in a pre-crimped state, whose diameter is usually lower than the final diameter when expanded. These pre-crimped stents are then crimped, through crimping machines, to further decrease the diameter and enable mounting on a delivery system. Thus, to model the expansion process of the stents, pre-



(a) The Curves outlining the geometry (b) Surfaces made from curves

(c) Extrude srufaces to make flat solid (d) Flat stent rolled to get a cylindrical shape

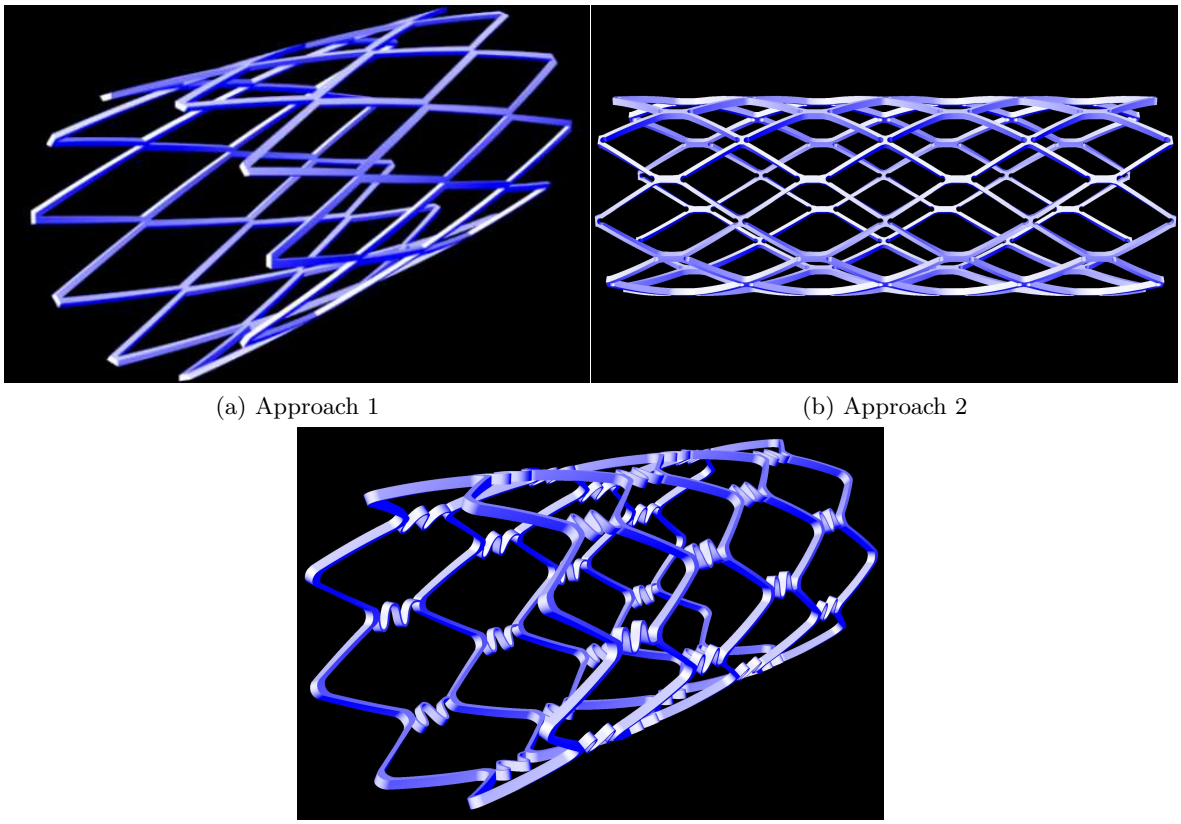
Figure 2.11: Approach 3 to construct stent geometry

crimped or crimped state geometrical models of the stents, depending on whether the crimping process is modelled or not, are needed. This does not present any difficulty in terms of CAD modelling of stents as the dimensions, for example the width of the flat plane and the radius of the cylindrical shell used for mapping, can be altered to create stent geometries in any required dimensions. Figure 2.13 shows the two crimped state representative models for the Bx VELOCITY and Xience V stents, both created using approach 3.

2.7 Conclusions

This chapter has introduced coronary artery disease, its treatment options, and issues related to its most common treatment, coronary stenting. Furthermore, a survey of a variety of coronary stents and their classification is presented. Thereafter, the properties that are desirable in an ideal coronary stent are briefly outlined. These properties are discussed individually in future chapters. Lastly, computer modelling approaches to create geometries of both expanded and pre-crimped state stents are presented.

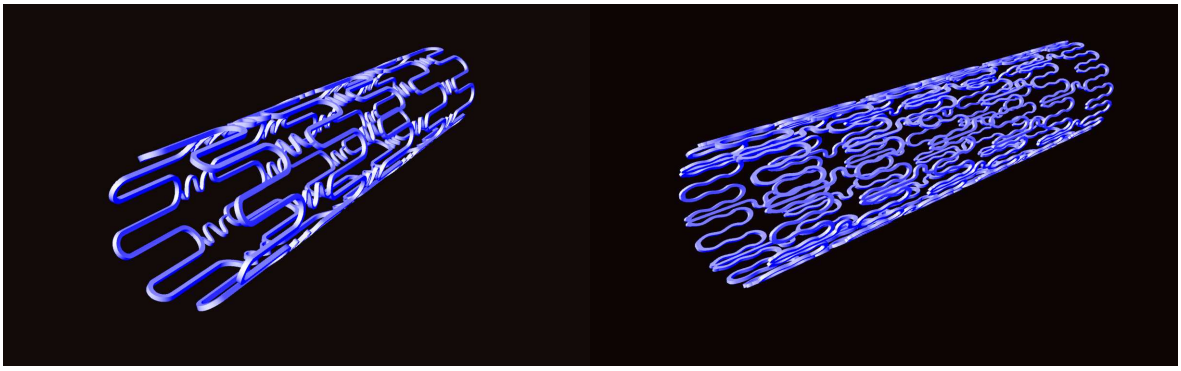
Before moving on to computational analysis of stents based on the desired properties, the next chapter presents an introduction to surrogate modelling and optimisation methodologies that are employed in future chapters.



(a) Approach 1 (b) Approach 2

(c) Approach 3

Figure 2.12: Stents modelled from the three approaches



(a) Crimped state model of the Bx VELOCITY stent (b) Crimped state model of the Xience V stent

Figure 2.13: Crimped state models created using approach 3

Chapter 3

Introduction to surrogate modelling and optimisation methodologies

This chapter presents an introduction of surrogate modelling and optimisation methodologies. First, a need for surrogate modelling and the basic formulation of the surrogate model used in this thesis, i.e. the Gaussian Process model, is presented. Thereafter, along with an introduction to various optimisation methodologies, the optimisation framework adopted in this thesis is presented.

3.1 What is optimisation?

In the most general sense optimisation can be defined as the process of "finding and comparing feasible solutions until no better solution can be found" [36]. Here, 'solutions' refer to different designs of the problem at hand, for example in the case of aerodynamic wing design solutions might refer to the different shapes of the wing; in the case of bridge design solutions might refer to the different structures which the bridge can take; and in the case of pharmaceutical drug design, it might refer to drugs produced by different combinations of individual drug components. Hence, one design can be thought of as a combination of several decision variables, where each decision variable can take multiple values. Decision variables in the case of wing design could be the location of NURBS control points that define the shape of the wing; for bridge design it could be the network (i.e. nodal locations and connectivity) of trusses and the lengths, cross-sections, materials etc. for each truss; and for drug design they could be the mole-fraction of each individual drug component. A 'better' solution in the definition refers to comparison with regards to a goal. For example, in the case of wing design the goal could be to achieve minimum drag; for bridge design the goal could be to achieve minimum weight; and for drug design the goal could be to kill maximum cancerous cells. In all of the tasks the goal would involve either maximisation or minimisation (collectively known as optimisation) [36] of a goal. Optimisation procedures that have only one goal are known as single-objective optimisation procedures, and those that have more than one goal are known as multi-objective optimisation procedures, where the goal is to simultaneously minimise or maximise two or more goals. For example, in wing design the two goals could be to minimise drag and maximise lift; for bridge design the goal

could be to minimise weight and minimise the cost of materials; for the problem of drug design the goal could be to maximise number of cancerous cells killed and minimise the cost of production. The final term in the definition is 'feasible solutions'. A feasible solution refers to solutions that do not violate any constraints. Constraints are requirements that are imposed irrespective of the goal. There are usually two type of constraints: those that are imposed on decision variables, commonly known as decision variable bounds; and those that are based on evaluation of a quantity based on the solution. An example of decision variable bounds for bridge design could be that each truss must not exceed a pre-determined length; and an example of an evaluated constraint could be that the maximum load that the bridge structure can withstand without failure must not be lower than a pre-determined threshold. Figure 3.1 shows an example of optimisation performed by National Aeronautics and Space Administration (NASA), USA, to design a space antenna for the Space Technology (ST5) satellites [9]. Using a systematic search and a parameterisation, i.e. the choice of design variables, that can represent a large design space, often non-intuitive designs, which are highly efficient can be obtained by the process of optimisation.

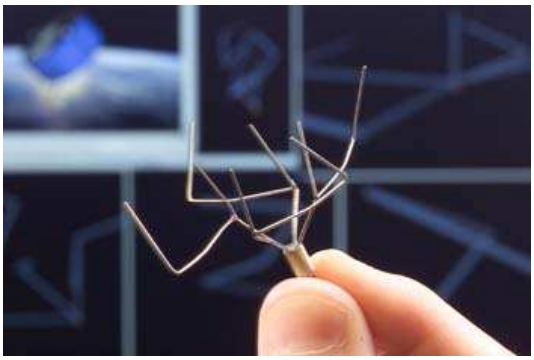


Figure 3.1: Space antenna designed by NASA using evolutionary optimisation and artificial intelligence [9]

Given the above background, a general single objective optimisation problem of minimising a particular goal, say $y(\mathbf{x})$ that depends on n decision variables $x_1, x_2, \ldots, x_n$, collectively written as the vector $\mathbf{x}$, subject to decision variable bounds, $x_i \in [x_i^L, x_i^U]$, l equality constraints, $h_j(\mathbf{x}) = 0$, $j = 1, 2, \ldots, l$, and m inequality constraints $g_k(\mathbf{x}) > 0$, $k = 1, 2, \ldots, m$, can we written as

$$\begin{aligned} &\text{Minimise} && y(\mathbf{x}) && \\ &\text{such that} && h_j(\mathbf{x}) = 0 && j = 1, 2, \ldots, l \\ &\text{and} && g_k(\mathbf{x}) > 0 && k = 1, 2, \ldots, m \\ &\text{subject to} && x_i \in [x_i^L, x_i^U] && i = 1, 2, \ldots, n. \end{aligned} \tag{3.1}$$

Similarly, a generic multiobjective optimisation problem, with q goals, $y_p(\mathbf{x})$, $p = 1, 2, \ldots, q$, can be written as

$$\begin{aligned} &\text{Minimise} && y_p(\mathbf{x}) && p = 1, 2, \ldots, q \\ &\text{such that} && h_j(\mathbf{x}) = 0 && j = 1, 2, \ldots, l \\ &\text{and} && g_k(\mathbf{x}) > 0 && k = 1, 2, \ldots, m \\ &\text{subject to} && x_i \in [x_i^L, x_i^U] && i = 1, 2, \ldots, n. \end{aligned} \tag{3.2}$$

In both the above formulations, many times, the variable bounds are included as inequality constraints.

Several algorithms, ranging from classical point by point methods to population based evolutionary algorithms have been developed to solve the above problems. Before discussing such algorithms in section 3.3, an overview of surrogate modelling is first presented in the following section.

3.2 Surrogate modelling

Any algorithm that is used to solve the optimisation problems formulated in equations 3.1 and 3.2 inevitably employs evaluation of the objective function, or goal(s), $y(\mathbf{x})$ or $y_p(\mathbf{x})$, and the constraints, $h_j(\mathbf{x})$ and $g_k(\mathbf{x})$, at multiple solutions, $\mathbf{x}$, in order to guide the search process towards the minimum. Usually, this does not present a problem if the objective function is either available in analytical form, or can be computed cheaply. Low cost here refers to the computational time taken to evaluate the objective function at one solution, and is a relative term. For example, for aerodynamic wing design, various solvers based on their respective complexity and assumptions, may take different amounts of computational time to evaluate the drag or lift on the wing. If the cost of evaluating the objective and the constraints at one solution is very high, cost being measured in terms of both analysis times and the computational budget available, then the application of almost any optimisation algorithm to such objective functions becomes impractical. This is primarily because it becomes difficult to evaluate the expensive objective function, also known as the high-fidelity solution, at so many solutions as required by any optimisation algorithm [18].

The aforementioned difficulty in using high-fidelity analysis for optimisation led to the development of approximation techniques. Pioneered by Schmit, Farshi, and Miura [37, 38], the concept that from the data obtained by the analysis of an initial design, an approximation of the objectives in the neighbourhood of the initial design can be constructed was developed. This later led to the development of more general approximation techniques [18]. The central notion behind such techniques is to represent the true functional relationship, $y = f(\mathbf{x})$, as an approximation, $\hat{y} = \hat{f}(\mathbf{x}, \alpha)$, where α is a vector of unknown parameters. The unknown parameters are either evaluated by a black-box based approach or a physics based approach [18]. In contrast to physics based approaches, where the form of governing equations is exploited to determine the functional form, $\hat{y} = \hat{f}(\mathbf{x}, \alpha)$, black-box approaches assume that the existing analysis codes can not be modified and hence for each solution the only information obtained from the analysis is the value of the objective function. Consequently, for black-box approaches, the high-fidelity analysis code is typically run at a number of pre-selected solutions (inputs) to find the corresponding objective function values (outputs). Thereafter, the input-output data obtained is used to train a surrogate model, by various approaches (most usually minimising a loss function, for example root mean squared error) [18]. This thesis is primarily concerned with this black-box surrogate modelling approach, owing to the practical advantage of no modification to the analysis codes, such as the computational fluid dynamics and finite element analysis codes that are employed to analyse the performance of stents.

The approximation models developed, also known as surrogate models, are computationally cheaper to evaluate and can be used in-lieu of the high fidelity analysis for all purposes, including optimisation, where repeated evaluation at several solutions is required for the task. As will be clear in future chapters, the high-fidelity analysis of each stent for evaluation of the desired characteristics, takes over seven days of computing time. This makes surrogate modelling indispensable for the design optimisation studies related to coronary stents.

In the following subsections, the three most popular approaches for surrogate modelling, viz. polynomial models, radial basis function models, and Gaussian Process models are presented. While the first two approaches are presented succinctly, the formulation for the

Gaussian Process predictor is presented in detail (as it forms the basis for the optimisation method developed in chapter 10).

3.2.1 Polynomial models

In a polynomial surrogate model, the functional form of the input-output relationship, $\hat{y} = \hat{f}(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^n$, is assumed to be of polynomial form [39, 40]. For example, a quadratic model can be written as

$$\hat{y} = c_0 + \sum_{1 \le j \le p} c_j x_j + \sum_{1 \le j \le p, k > j} c_{p-1+j+k}\, x_j\, x_k \ , \tag{3.3}$$

where $c_0, c_1, \ldots, c_{m-1}$, $m = (n+1)(n+2)/2$, are the m unknown coefficients. This model can be compactly written as

$$\hat{y}(\mathbf{x}) = \mathbf{c}^T \overline{\mathbf{x}} \tag{3.4}$$

where $\mathbf{c} = [c_0, c_1, \ldots, c_{m-1}]^T \in \mathbb{R}^{m \times 1}$, and $\overline{\mathbf{x}} = [1, x_1, x_2, \ldots, x_1^2, x_1x_2, x_1x_3, \ldots, x_n^2]^T$. Now, if p solutions are observed, i.e. the high-fidelity analysis is run over p points, $\mathbf{x^{(i)}}, i = 1, 2, \ldots, p$, to evaluate the corresponding function value, $y^{(i)}, i = 1, 2, \ldots, p$, the coefficients, $\mathbf{c}$ can be determining by using least squares regression, i.e. by solving the following system of equations:

$$\mathbf{Ac} = \mathbf{y}, \tag{3.5}$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \ldots & (x_n^{(1)})^2 \\ 1 & x_1^{(2)} & x_2^{(2)} & \ldots & (x_n^{(2)})^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(p)} & x_2^{(p)} & \ldots & (x_n^{(p)})^2 \end{bmatrix} \in \mathbb{R}^{p \times m}, \tag{3.6}$$

and

$$\mathbf{y} = [y^1, y^1, \ldots, y^p]^T \in \mathbb{R}^{p \times 1}. \tag{3.7}$$

The reader is referred to the texts by Box and Draper [39] and Myers and Montgomery [40] for further details of such models and methods of determining the accuracy of such models.

3.2.2 Radial Basis Function models

Radial Basis Function (RBF) models are similar to polynomial models except that they employ a different set of basis functions, the radial basis functions, as opposed to polynomials [18]. The general form of the surrogate model can be written as follows

$$\hat{y}(\mathbf{x}) = \sum_{i=1}^{p} \alpha_i K(\|\mathbf{x} - \mathbf{x^{(i)}}\|), \tag{3.8}$$

where $\|\cdot\|$ represents the norm, usually the Euclidean distance, $K(\|\mathbf{x} - \mathbf{x^{(i)}}\|)$ is a radial basis function and α_i, $i = 1, 2, \ldots, p$, are unknown weights for each radial basis function centered at the sampled point, $\mathbf{x^{(i)}}$. The many choices for the radial basis functions are listed in Table 3.1 [18]. It can be observed from this table that the Gaussian, multiquadratics, and inverse multiquadratics RBFs involve an additional parameter, θ, called the shape parameter. The shape parameter controls the shape of the RBF and hence the domain over which each radial basis function has an influence. For example, for the Gaussian RBF, as θ increases, each RBF gets wider, thereby influencing a larger domain.

Table 3.1: Typical choices for RBFs for radial basis function surrogate modelling [18]

RBF	form for $K(\|\mathbf{x}-\mathbf{x^{(i)}}\|)$ (Equation 3.8)
Linear splines	$\|\mathbf{x}-\mathbf{x^{(i)}}\|$
Thin plate splines	$\|\mathbf{x}-\mathbf{x^{(i)}}\|^k \ln \|\mathbf{x}-\mathbf{x^{(i)}}\|$; $k \in [2,4,\ldots]$
Cubic splines	$\|\mathbf{x}-\mathbf{x^{(i)}}\|^3$
Gaussian	$\exp\left(-\frac{\|\mathbf{x}-\mathbf{x^{(i)}}\|^2}{\theta}\right)$
Multiquadratics	$\left(1+\frac{\|\mathbf{x}-\mathbf{x^{(i)}}\|^2}{\theta}\right)^{1/2}$
Inverse multiquadratics	$\left(1+\frac{\|\mathbf{x}-\mathbf{x^{(i)}}\|^2}{\theta}\right)^{-1/2}$

In order to calculate the weight vector $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_p]^T \in \mathbb{R}^{p\times 1}$, the method of linear least squares is employed, and the following system of equations is solved

$$\mathbf{K}\boldsymbol{\alpha} = \mathbf{y}, \tag{3.9}$$

where $\mathbf{K}$ is the Gram matrix [18]

$$\mathbf{K} = \begin{bmatrix} K(\|\mathbf{x^{(1)}}-\mathbf{x^{(1)}}\|) & K(\|\mathbf{x^{(1)}}-\mathbf{x^{(2)}}\|) & \ldots & K(\|\mathbf{x^{(1)}}-\mathbf{x^{(p)}}\|) \\ K(\|\mathbf{x^{(2)}}-\mathbf{x^{(1)}}\|) & K(\|\mathbf{x^{(2)}}-\mathbf{x^{(2)}}\|) & \ldots & K(\|\mathbf{x^{(2)}}-\mathbf{x^{(p)}}\|) \\ \vdots & \vdots & \ddots & \vdots \\ K(\|\mathbf{x^{(p)}}-\mathbf{x^{(1)}}\|) & K(\|\mathbf{x^{(p)}}-\mathbf{x^{(2)}}\|) & \ldots & K(\|\mathbf{x^{(p)}}-\mathbf{x^{(p)}}\|) \end{bmatrix} \in \mathbb{R}^{p\times p}, \tag{3.10}$$

and

$$\mathbf{y} = [y^1, y^1, \ldots, y^p]^T \in \mathbb{R}^{p\times 1}. \tag{3.11}$$

For a discussion on singularities of the Gram matrix, that might lead to difficulties in solving equation 3.9, the reader is referred to the work of Micchelli [41] and Wang [42]. As mentioned before the Gaussian, multiquadratics, and the inverse multiquadratics RBFs have an additional shape parameter θ, which is selected by the user. This shape parameter can have a significant effect on the performance and smoothness of the approximation model [18]. Various methods are often employed to estimate the optimal value of θ from the observed data. The two primary methods to accomplish this are the 'leave-one-out' method and the 'maximum likelihood' method. In this section, the 'leave-one-out' method is described. The 'maximum likelihood' approach is presented in detail in the next section where the formulation of a Gaussian Process predictor is presented.

For the leave-one-out approach, first the prediction error needs to be defined. This error refers to the square of the difference between the actual value of the function and the prediction made by the surrogate model at a point $\mathbf{x}$. It is defined as

$$\mathbf{Q}(\mathbf{z}, \boldsymbol{\alpha}) = \left(y^{(i)} - \sum_{j=1}^{p} \alpha_j K(\|\mathbf{x}-\mathbf{x^{(j)}}\|)\right)^2, \tag{3.12}$$

where $\mathbf{z}$ denotes the training data set, $y^{(i)}$ denotes the real function value at a point $\mathbf{x}$, and $\mathbf{x^{(j)}}$ represents the training data points. If a validation data-set is available, the sum of the above errors for all the points in the validation data-set can be minimised to yield an optimum value of θ. However, this is not always practical, especially when the computational time to create

the training data-set itself are high. The leave-one-out procedure overcomes this challenge. The procedure involves calculating the weight vector, $\boldsymbol{\alpha}$, by using all but one training point, i.e. leaving one point out, and making a prediction from the trained model at the left-out point. Hence, the leave-one-out approximation error can be written as

$$\mathbf{Q}(\mathbf{z}, \boldsymbol{\alpha}_{-i}) = \left(y^{(i)} - \sum_{j=1, j\neq i}^{p} \alpha_j K(\|\mathbf{x^{(i)}} - \mathbf{x^{(j)}}\|) \right)^2, \tag{3.13}$$

where $\boldsymbol{\alpha}_{-i}$ represents the weight vector calculated by deleting the i^{th} point from the training data-set, and $y^{(i)}$ denotes the real function value at a point $\mathbf{x^{(i)}}$. The leave-one-out process is repeated p, the number of points in the training data set, times leaving out a different point each time to calculate the total approximation error

$$\mathbf{Q}(\mathbf{z}) = \frac{1}{p} \sum_{i=1}^{p} \mathbf{Q}(\mathbf{z}, \boldsymbol{\alpha}_{-i}), \tag{3.14}$$

where $\mathbf{Q(z)}$ is the mean of the total approximation error. A value of θ that minimises the above mean error is chosen as the optimal value of θ for the RBF model [18]. In the next section, a popular surrogate model, the Gaussian Process model, is presented.

3.2.3 Gaussian Process surrogate models

Gaussian Process (GP) modelling is a widely used and statistically rigorous method for constructing surrogate models [18]. Its origins lie in the work of Krige [43], who developed the method to predict mineral concentrations, in the area of geostatistics [44]. A detailed account of Gaussian Process modelling can be found in the works of Sacks et. al. [45], Santner et. al. [46], Mackay [47], and Rasmussen & Williams [48]. It must be noted that GP modelling is also referred to as Kriging or DACE (design and analysis of computer experiments) modelling [18]. Throughout this thesis the words Kriging and Gaussian Process modelling are used interchangeably to imply the same process. Similarly, the words Krigs and GP models are also used interchangeably.

In what follows, the formulation of a Gaussian Process predictor, and the corresponding uncertainty involved in the prediction, is presented. However, first it should be noted that a Gaussian Process model by definition relies on Gaussian Processes, which are a kind of stochastic processes. Hence, before moving on to the formulation of a GP predictor, the concept of stochastic processes is discussed. Parzen [49] defines stochastic processes as *"a family of random variables* $\{X(t), t \in \mathbf{T}\}$*, indexed by a parameter* t *varying in an index set* $\mathbf{T}$*"*. For example, if the radius, r, of a continuous one-dimensional wire of length L, is represented as a stochastic process, then this stochastic process is the family of random variables that represent the radius of the wire at each location t, where t can take any real value between 0 and length, L. Thus, the index set $\mathbf{T}$ to which each t belongs is $[0, L]$. Now, it seems that if one wants to completely describe the radius of this wire through this stochastic process, then an infinite number of random variables are needed in the process. However, it is possible [49], for all practical purposes, to adequately represent the stochastic process by a finite number of ordinates, t. Hence, one way to describe a stochastic process, $\{X(t), t \in \mathbf{T}\}$, is to specify the joint probability distribution for n random variables, $X(t_1), X(t_2), \ldots, X(t_n)$, i.e. to specify the form of

$$F_{X(t_1),\ldots,X(t_n)}(x_1, x_2, \ldots, x_n) = P(X(t_1) \leq x_1, X(t_2) \leq x_2, \ldots, X(t_n) \leq x_n), \tag{3.15}$$

where F is the cumulative probability distribution function and P represents probability. Alternatively, the probability density, $f(x_1, x_2, \ldots, x_n)$ function may be specified, i.e.

$$F_{X(t_1),\ldots,X(t_n)}(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\ldots\int_{-\infty}^{x_n} f(x_1, x_2, \ldots, x_n)\ dx_1 dx_2 \ldots dx_n. \tag{3.16}$$

It will be seen in the next section, that Gaussian process models, following the definition of stochastic processes above, are stochastic processes where the aforementioned joint probability density function is assumed to be a multivariate Gaussian probability distribution, i.e. $f(x_1, x_2, \ldots, x_n) = f(\mathbf{x})$, where $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T \in \mathbb{R}^{n\times 1}$ is specified to be

$$f(\mathbf{x}) = (2\pi)^{-n/2}|\Sigma|^{-1/2}\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right), \tag{3.17}$$

where $\boldsymbol{\mu} \in \mathbb{R}^{n\times 1}$ is the mean vector and $\Sigma \in \mathbb{R}^{n\times n}$ is the covariance matrix [48]. The above is written compactly as

$$\mathbf{x} \backsim \mathcal{N}(\boldsymbol{\mu}, \Sigma). \tag{3.18}$$

With the above background, the formulation of the Gaussian Process predictor is presented in the next subsection.

3.2.3.1 Formulation of a Gaussian Process predictor

Let y be an unknown function which depends on n variables $(x_1, x_2, \ldots, x_n)$. If the vector $\mathbf{x}$ is written as $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$, then the goal is to formulate a model structure which represents the relationship of y as a function of the vector $\mathbf{x}$, i.e.

$$y = f(\mathbf{x}). \tag{3.19}$$

In Gaussian process modelling, any finite number of function values, $y(\mathbf{x^{(1)}}), y(\mathbf{x^{(2)}}), \ldots, y(\mathbf{x^{(q)}})$, evaluated at the q points, $\mathbf{x^{(1)}}, \mathbf{x^{(2)}}, \ldots, \mathbf{x^{(q)}}$, where $\mathbf{x^{(i)}} = [x_1^{(i)}, x_2^{(i)}, \ldots, x_n^{(i)}]^T$, are assumed to be random variables which are a realisation of the Gaussian random field, i.e. they have a joint Gaussian (normal) distribution. If the mean of the Gaussian random field is β and its covariance is $\Gamma(\mathbf{x}, \mathbf{x}')$, then the model structure can be written as

$$Y(\mathbf{x}) = \beta + Z(\mathbf{x}), \tag{3.20}$$

where β, the mean of the random field, is to be estimated, and $Z(\mathbf{x})$ is a Gaussian process with zero mean and the following covariance:

$$Cov(\mathbf{x}, \mathbf{x}') = \Gamma(\mathbf{x}, \mathbf{x}') = \sigma_z^2 R(\mathbf{x}, \mathbf{x}'), \tag{3.21}$$

where σ_z^2 is the process variance and $R(\mathbf{x}, \mathbf{x}')$ is a parameterised correlation function, between two points $\mathbf{x}$ and $\mathbf{x}'$. A common choice for the correlation function is

$$R(\mathbf{x}, \mathbf{x}') = \exp\left[-\sum_{j=1}^{n}(\theta_j|x_j - x_j'|^{m_j})\right], \tag{3.22}$$

where $\theta_j \geq 0$ and $0 < m_j \leq 2$ are undetermined hyperparameters. Since any finite observations of the function are assumed to be realisations of the above field, the set of observed outputs $\{y(\mathbf{x^{(1)}}), y(\mathbf{x^{(2)}}), \ldots, y(\mathbf{x^{(p)}})\}$, compactly written as $\{y^{(1)}, y^{(2)}, \ldots, y^{(p)}\}$, also have a joint normal distribution. Thus the vector of outputs $\mathbf{y} = [y^{(1)}, y^{(2)}, \ldots, y^{(p)}]^T$, has the following Gaussian distribution with the assumed covariance structure as specified by equation 3.21, that is

$$\mathbf{y} \backsim \mathcal{N}(\mathbf{1}\beta, \mathbf{\Gamma}), \tag{3.23}$$

where $\mathbf{1}$ is $[1, 1, \ldots, 1]^T \in \mathbb{R}^p$, and $\mathbf{\Gamma} \in \mathbb{R}^{p \times p}$ is the variance-covariance matrix whose ij^{th} element is given by $\mathbf{\Gamma}(\mathbf{x^{(i)}}, \mathbf{x^{(j)}})$, i.e.

$$\mathbf{\Gamma} = \begin{pmatrix} \Gamma(\mathbf{x^{(1)}}, \mathbf{x^{(1)}}) & \Gamma(\mathbf{x^{(1)}}, \mathbf{x^{(2)}}) & \ldots & \Gamma(\mathbf{x^{(1)}}, \mathbf{x^{(p)}}) \\ \Gamma(\mathbf{x^{(2)}}, \mathbf{x^{(1)}}) & \Gamma(\mathbf{x^{(2)}}, \mathbf{x^{(2)}}) & \ldots & \Gamma(\mathbf{x^{(2)}}, \mathbf{x^{(p)}}) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma(\mathbf{x^{(p)}}, \mathbf{x^{(1)}}) & \Gamma(\mathbf{x^{(p)}}, \mathbf{x^{(2)}}) & \ldots & \Gamma(\mathbf{x^{(p)}}, \mathbf{x^{(p)}}) \end{pmatrix}. \tag{3.24}$$

Consequently, from equation 3.17, the likelihood function, i.e. the likelihood of the observed data being generated by the parameterised Gaussian random field, is

$$L(\boldsymbol{\theta}, \beta, \sigma_z^2) = (2\pi)^{-p/2} (\sigma_z^2)^{-p/2} |\mathbf{R}|^{-1/2} \exp\left(-\frac{1}{2\sigma_z^2} (\mathbf{y} - \mathbf{1}\beta)^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{1}\beta) \right), \tag{3.25}$$

where $\boldsymbol{\theta}$ is a vector of the θ_i, $i = 1$ to n, hyperparameters, $\mathbf{R} \in \mathbb{R}^{p \times p}$ is the correlation matrix whose ij^{th} element is given by $R(\mathbf{x^{(i)}}, \mathbf{x^{(j)}})$, i.e.

$$\mathbf{R} = \begin{pmatrix} R(\mathbf{x^{(1)}}, \mathbf{x^{(1)}}) & R(\mathbf{x^{(1)}}, \mathbf{x^{(2)}}) & \ldots & R(\mathbf{x^{(1)}}, \mathbf{x^{(p)}}) \\ R(\mathbf{x^{(2)}}, \mathbf{x^{(1)}}) & R(\mathbf{x^{(2)}}, \mathbf{x^{(2)}}) & \ldots & R(\mathbf{x^{(2)}}, \mathbf{x^{(p)}}) \\ \vdots & \vdots & \ddots & \vdots \\ R(\mathbf{x^{(p)}}, \mathbf{x^{(1)}}) & R(\mathbf{x^{(p)}}, \mathbf{x^{(2)}}) & \ldots & R(\mathbf{x^{(p)}}, \mathbf{x^{(p)}}) \end{pmatrix}, \tag{3.26}$$

and $|\mathbf{R}|^{-1/2}$ is the determinant of the matrix $\mathbf{R}$. In order to estimate the hyperparameters, $\boldsymbol{\theta}, \beta$, and σ_z^2, this likelihood function is maximised. This is equivalent to minimising the negative log of the likelihood function (in order to convert the products into sums). The negative log-likelihood function obtained from equation 3.25 is

$$\mathcal{L}(\boldsymbol{\theta}, \beta, \sigma_z^2) = \frac{1}{2} \left[p \ln(2\pi) + p \ln \sigma_z^2 + \ln|\mathbf{R}| + \frac{1}{\sigma_z^2} (\mathbf{y} - \mathbf{1}\beta)^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{1}\beta) \right]. \tag{3.27}$$

To minimise the above negative log-likelihood function, its derivatives with respect to the hyperparameters are set to zero. The derivative with respect to β is

$$\frac{\partial \mathcal{L}(\boldsymbol{\theta}, \beta, \sigma_z^2)}{\partial \beta} = \frac{1}{2\sigma_z^2} (\mathbf{1}^T \mathbf{R}^{-1} \mathbf{y} - \mathbf{1}^T \mathbf{R}^{-1} \mathbf{1}\beta) = 0 \tag{3.28}$$

$$\implies \hat{\beta} = \frac{\mathbf{1}^T \mathbf{R}^{-1} \mathbf{y}}{\mathbf{1}^T \mathbf{R}^{-1} \mathbf{1}}. \tag{3.29}$$

Similarly, differentiating the negative log-likelihood function with respect to σ_z^2 yields

$$\frac{\partial \mathcal{L}(\boldsymbol{\theta}, \beta, \sigma_z^2)}{\partial \sigma_z^2} = \frac{1}{2} \left[\frac{p}{\sigma_z^2} - \frac{1}{(\sigma_z^2)^2} (\mathbf{y} - \mathbf{1}\beta)^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{1}\beta) \right] = 0 \tag{3.30}$$

$$\implies \hat{\sigma}_z^2 = \frac{1}{p}(\mathbf{y} - \mathbf{1}\beta)^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\beta). \tag{3.31}$$

Equations 3.29 and 3.31 provide optimal estimates for β and σ_z^2, respectively, for a given $\boldsymbol{\theta}$. Differentiating the negative log-likelihood function with respect to θ_j yields

$$\frac{\partial \mathcal{L}(\boldsymbol{\theta}, \beta, \sigma_z^2)}{\partial \theta_j} = \frac{1}{2}\left[\mathrm{Tr}\left(\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \theta_j}\right) - (\mathbf{y} - \mathbf{1}\beta)^T \mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \theta_j}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\beta)\right] = 0, \tag{3.32}$$

where Tr denotes trace of a matrix, i.e. $\mathrm{Tr}(A) = \sum_{i=1}^{i=m} a_{ii}$, when $A \in \mathbb{R}^{m \times m}$. Unlike equations 3.28 and 3.30, equation 3.32 does not yield an analytical solution for the estimate of $\boldsymbol{\theta}$. Hence, an iterative scheme in $\boldsymbol{\theta}$ must be employed to minimise $\mathcal{L}$, where, for each $\boldsymbol{\theta}$ the optimal estimates for β and σ_z^2 are calculated using equations 3.29 and 3.31.

Having calculated the maximum likelihood estimates for $\boldsymbol{\theta}$, β, and σ_z^2, one can predict the function value at an unobserved point, $\mathbf{x}^*$. If y^* denotes the output at this unobserved point, $\mathbf{x}^*$, then according to the prior assumption that any finite number of outputs (function values calculated at different points) have a joint normal distribution, the following can be written

$$\begin{bmatrix} \mathbf{y} \\ y^* \end{bmatrix} \backsim \mathcal{N}\left(\begin{bmatrix} \mathbf{1}_{p+1}\beta \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Gamma} & \boldsymbol{\gamma}(\boldsymbol{x}^*) \\ \boldsymbol{\gamma}(\boldsymbol{x}^*)^T & \Gamma(\mathbf{x}^*, \mathbf{x}^*) \end{bmatrix}\right), \tag{3.33}$$

where $\mathbf{1}_{p+1}$ is $[1, 1, \ldots, 1]^T \in \mathbb{R}^{p+1}$ and $\boldsymbol{\gamma}(\boldsymbol{x^*}) = [\sigma_z^2 R(\mathbf{x}^*, \mathbf{x^{(1)}}), \sigma_z^2 R(\mathbf{x}^*, \mathbf{x^{(2)}}), \ldots, \sigma_z^2 R(\mathbf{x}^*, \mathbf{x^{(p)}})]^T \in \mathbb{R}^p$. Now, in order to get a prediction for y^* we take the conditional distribution of y^* given $\mathbf{y}$. This is called the *posterior* probability distribution. Since the *prior* distribution is Gaussian, the posterior distribution is also Gaussian (because conditional distribution of a Gaussian distribution is a Gaussian; see appendix A for details), and is given by

$$y^*|\mathbf{y} \backsim \mathcal{N}\Big(\beta + \boldsymbol{\gamma}(\boldsymbol{x}^*)^T \boldsymbol{\Gamma}^{-1}(\mathbf{y} - \mathbf{1}\beta)\ ,\ \Gamma(\mathbf{x}^*, \mathbf{x}^*) - \boldsymbol{\gamma}(\boldsymbol{x}^*)^T \boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}(\boldsymbol{x^*})\Big). \tag{3.34}$$

From the above, the posterior mean and posterior covariance can be written as

$$\hat{y}(\mathbf{x}) = \beta + \mathfrak{r}(\mathbf{x})^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\beta), \tag{3.35}$$

$$\text{and} \qquad C(\mathbf{x}, \mathbf{x}') = \sigma_z^2\Big(R(\mathbf{x}, \mathbf{x}') - \mathfrak{r}(\mathbf{x})^T \mathbf{R}^{-1}\mathfrak{r}(\mathbf{x}')\Big), \tag{3.36}$$

where $\mathfrak{r}(\mathbf{x}) = [R(\mathbf{x}, \mathbf{x^{(1)}}),\ R(\mathbf{x}, \mathbf{x^{(2)}}),\ \ldots,\ R(\mathbf{x}, \mathbf{x^{(p)}})]^T \in \mathbb{R}^p$, and represents the correlation of the new point at which prediction is being made with all the observed data points. The mean of the posterior, as predicted by equation 3.35, is seen as the prediction of the output at the new point, $\mathbf{x}$, and the posterior variance given by the following equation is seen as a measure of uncertainty associated with the prediction of the output

$$\sigma^2(\mathbf{x}) = C(\mathbf{x}, \mathbf{x}) = \sigma_z^2\Big(1 - \mathfrak{r}(\mathbf{x})^T \mathbf{R}^{-1}\mathfrak{r}(\mathbf{x})\Big). \tag{3.37}$$

For computational efficiency, it is advantageous to write $\mathbf{w} = \mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\beta)$, so that the posterior mean can be written as a dot product of two vectors, $\mathbf{w}$ and $\mathfrak{r}(\mathbf{x})^T$, as follows

$$\hat{y}(\mathbf{x}) = \beta + \mathfrak{r}(\mathbf{x})^T \mathbf{w} = \beta + \sum_{i=1}^{i=p} w_p R(\mathbf{x}, \mathbf{x}^{(i)}). \tag{3.38}$$

This concludes the formulation of the Kriging predictor, where the prediction at a new point and the associated uncertainty can be calculated by equations 3.38 (or 3.35) and 3.37, respectively.

3.2.3.2 Validation methods for GP models

Once a GP model has been constructed, the validity of the model, i.e. the accuracy of the model, needs to be ascertained. If testing data is available, i.e. the response of a function for a number of points other than the training points, then the validity of the GP model can be ascertained by evaluating the prediction error, i.e. the difference between the predicted value and the actual function value, at the testing points. However, availability of testing data is usually impractical owing to the computational cost of running the high-fidelity analysis at the testing points. To circumvent this difficulty, Jones et. al. [50] proposed the concept of "standardized cross-validated residual" (SCVR). It is based on both the leave-one-out method presented in section 3.2.2 and the posterior variance, as defined by equation 3.37. The basic idea is to create the GP model by leaving one point at a time, and calculating the posterior mean and variance at the left-out point. SCVR is defined as

$$\mathrm{SCVR}_i = \frac{y^{(i)} - \hat{y}_{-i}(\mathbf{x^{(i)}})}{\sigma_{-i}(\mathbf{x^{(i)}})}, \tag{3.39}$$

where $y^{(i)}$ is the observed value at the i^{th} point that is left-out in creating the GP model, $\hat{y}_{-i}$ is the prediction at the left-out point, and σ_{-i} is the square root of the posterior variance, σ^2_{-i}, of the prediction at the left-out point. If the SCVR for all the points lies in $[-3, +3]$ then the GP model constructed is appropriate [18]. This implies that the GP model is approximately 99.7% confident that the predictions lie between ± 3 times the square root of posterior variance [50].

This concludes the introduction to surrogate modelling. Before moving on to presenting an introduction to optimisation algorithms in the next section, useful tools of main effects and sensitivity indices are introduced in the next subsection. These tools are helpful in extracting relationships between the variables and the function that is modelled as a GP.

3.2.3.3 Extracting relationships between response and variables in GP models

It is apparent from equations 3.35 and 3.37 that the GP predictors do not present the relationship between the components of decision variable, $\mathbf{x}$, and the response, $y(\mathbf{x})$, in a readily interpretable manner [51, 45, 52]. Hence, in order to understand the functional relationship, the effect of one component of $\mathbf{x}$, say x_i, on $y(\mathbf{x})$ needs to be isolated from the effect of other components of $\mathbf{x}$. This is done by integrating out the effects of all variables but one from the response $y(\mathbf{x})$ as [51, 52]

$$\Upsilon(x_i) = \frac{1}{V} \int_V y(\mathbf{x}) \prod_{j \neq i} dx_j, \tag{3.40}$$

where $\Upsilon(x_i)$ is called the *main effect* of variable x_i, and V is the hypervolume created by all the variables excluding x_i. The above integral can be approximated numerically as a sum over a grid of points in the hypervolume [52]. It is often desirable to combine the main effects into global sensitivity indices, as suggested by Sobol [53], as

$$S_i = \frac{\int \Upsilon_i^2(x_i)\, dx_i}{\sum_{j=1}^{n} \int \Upsilon_j^2(x_j)\, dx_j}, \tag{3.41}$$

where n is the total number of variables influencing $y(\mathbf{x})$, and S_i are global sensitivity indices, which denote the relative effect/sensitivity of the variable x_i with respect to all other variables. These sensitivity indices are used in chapter 7 to study the relative effect of design variable specifying stent geometry on objective functions that measure stent performance. In the next section a brief overview of optimisation methods is presented.

3.3 Optimisation algorithms

As discussed before the process of optimisation involves locating and comparison of feasible solutions until better solutions can not be found [36]. Any numerical scheme that achieves this task is known as an optimisation algorithm.

For a minimisation problem, a local minimum and global minimum can be defined. A local minimiser of $f(\mathbf{x})$ is defined as *a point,* $\mathbf{x}^*$*, such that there is a neighbourhood,* $\mathcal{N}$*, of* $\mathbf{x}^*$ *such that* $f(\mathbf{x}^*) \leq f(\mathbf{x})$ *for all* $\mathbf{x} \in \mathcal{N}$ [54]. On the other hand a global minimiser of $f(\mathbf{x})$ is defined as *a point,* $\mathbf{x}^*$*, such that* $f(\mathbf{x}^*) \leq f(\mathbf{x})$ *for all* $\mathbf{x} \in \mathbb{R}^n$*, where* n *is the dimensionality of* $f(\mathbf{x})$*, or for all* $x \in \mathcal{D}$*, where* $\mathcal{D}$ *is the search domain* [54]. Hence, every global minimum is a local minimum but the converse is not true.

The necessary and sufficient conditions for a local minimiser can be easily defined. The first-order (implying that only first derivative information is used) necessary condition states that *"If* $\mathbf{x}^*$ *is a local minimiser and* $f(\mathbf{x})$ *is continuously differentiable in an open neighborhood of* $\mathbf{x}^*$*, then* $\nabla f(\mathbf{x}^*) = 0$*"* [54]. Intuitively, this can be rationalised by the argument that if $\nabla f(\mathbf{x}^*)$ is non-zero, then the function is decreasing in the direction of $-\nabla f(\mathbf{x}^*)$ and hence by moving in that direction the function value can be further decreased, thereby implying that $\mathbf{x}^*$ is not a local minimum. The necessary conditions however do not guarantee that the solution found is a local minimum: a point satisfying the necessary conditions is only a stationary point and it could be a local minimum, local maximum, or a saddle point. In order to be certain about the point being local minimum, the second-order sufficient conditions need to be satisfied. These are stated as *"Suppose that* $\nabla^2 f(\mathbf{x})$ *is continuous in an open neighbourhood of* $\mathbf{x}^*$ *and that* $\nabla f(\mathbf{x}^*) = 0$ *and* $\nabla^2 f(\mathbf{x}^*)$ *is positive definite. Then* $\mathbf{x}^*$ *is a strict local minimiser of* $f(\mathbf{x})$*"* [54].

Having laid out the necessary and sufficient conditions, in the following sections an overview of optimisation algorithms according to a broad classification is presented. Unless explicitly stated otherwise, a minimum in this chapter refers to a local minimum.

3.3.1 Gradient based methods

As the name suggests, gradient based methods use derivative information in order to guide the search process towards a minimum. In its most general sense, starting from a feasible or a random point in the search domain, the next point to be evaluated is decided by moving in a search direction that is calculated by using derivative information at the start point. This process is then repeated until no more improvement in the function value can be made. The general form of such methods can be written as

$$\mathbf{x}^{(\kappa+1)} = \mathbf{x}^{(\kappa)} + \alpha\, \mathbf{s}^{(\kappa)} \tag{3.42}$$

where $\mathbf{x}^{(\kappa+1)}$ is the next point to be evaluated, $\mathbf{x}^{(\kappa)}$ is the current point, α is the step length, and $\mathbf{s}^{(\kappa)}$ is the search direction at $\mathbf{x}^{(\kappa)}$. In essence, at every point $\mathbf{x}^{(\kappa)}$ a step of α is taken in a direction $\mathbf{s}^{(\kappa)}$, depending on the optimisation algorithm, to obtain the next point $\mathbf{x}^{(\kappa+1)}$. Common methods to determine the search direction are described in the next subsections.

3.3.2 Steepest descent

The formulation of steepest descent direction is easy to understand if one considers the first order Taylor's series expansion of the function $f(\mathbf{x})$ about the current point $\mathbf{x}^{(\kappa)}$

$$f(\mathbf{x}^{(\kappa+1)}) = f(\mathbf{x}^{(\kappa)} + \alpha\, \mathbf{s}^{(\kappa)}) \approx f(\mathbf{x}^{(\kappa)}) + \alpha\, (\mathbf{s}^{(\kappa)})^T\, \nabla f(\mathbf{x}^{(\kappa)}), \tag{3.43}$$

where $\nabla f(\mathbf{x}^{(\kappa)})$ represents the gradient vector at the current point $\mathbf{x}^{(\kappa)}$. Along a search direction $\mathbf{s}^{(\kappa)}$, the above equation represents a linear approximation for the function $f(\mathbf{x})$ about the current point $\mathbf{x}^{(\kappa)}$. It is clear from this that maximum decrease in the function will be achieved if $\mathbf{s}^{(\kappa)}$ is equal to $-\nabla f(\mathbf{x}^{(\kappa)})$ [55]. This direction is known as the direction of steepest descent and an algorithm that uses this direction at each iteration of equation 3.42 is called a steepest descent search algorithm.

3.3.2.1 Newton's method

Newton's method considers the second order Taylor series expansion of the function along a search direction $\mathbf{s}^{(\kappa)}$

$$f(\mathbf{x}^{(\kappa+1)}) = f(\mathbf{x}^{(\kappa)} + \alpha\, \mathbf{s}^{(\kappa)}) \approx f(\mathbf{x}^{(\kappa)}) + \alpha\, (\mathbf{s}^{(\kappa)})^T\, \nabla f(\mathbf{x}^{(\kappa)}) + \alpha^2\, \frac{1}{2} (\mathbf{s}^{(\kappa)})^T \mathbf{H}^{\kappa}\, \mathbf{s}^{(\kappa)} \tag{3.44}$$

or

$$\nabla f(\mathbf{x}^{(\kappa+1)}) \approx\ \nabla f(\mathbf{x}^{(\kappa)}) + \alpha \mathbf{H}^{\kappa}\, \mathbf{s}^{(\kappa)}, \tag{3.45}$$

where $\mathbf{H}^{\kappa}$ represents the Hessian matrix, at the current point $\mathbf{x}^{(\kappa)}$, i.e. the ij^{th} element of $\mathbf{H}^{\kappa}$ is given by

$$H_{ij}^{\kappa} = \frac{\partial^2 f(\mathbf{x}^{\kappa})}{\partial x_i\, \partial x_j}. \tag{3.46}$$

First order optimality condition for an optimum implies that at the optimal solution, $\mathbf{x}^*$, the first derivative is zero, i.e. $\nabla f(\mathbf{x}^*) = 0$. Newton's method ensures this at each iteration by setting $\nabla f(\mathbf{x}^{(\kappa+1)})$ to zero. Equation 3.45 then becomes

$$0 =\ \nabla f(\mathbf{x}^{(\kappa)}) + \alpha \mathbf{H}^{\kappa}\, \mathbf{s}^{(\kappa)} \tag{3.47}$$

or

$$\alpha\, \mathbf{s}^{(\kappa)} = -[\mathbf{H}^{\kappa}]^{-1} \nabla f(\mathbf{x}^{(\kappa)}). \tag{3.48}$$

The above direction $\mathbf{s}^{(\kappa)}$ is used by the Newton's method in Equation 3.42. Newton's method assumes that the Hessian matrix is available at each iteration. In many cases the Hessian matrix is not available directly or is expensive to calculate [18]. This led to the development of Quasi-Newton methods, which are discussed next.

3.3.2.2 Quasi-Newton or variable metric methods

Newton's method, as shown above, required the inverse of the matrix to be computed at each iteration. This might not be practical either due to the expense of calculating the inverse of the Hessian or due to the non-availability of the Hessian directly. In Quasi-Newton or variable metric methods, the inverse of the Hessian matrix is approximated in the region of the minimum by a sequence of iterative application of first derivative information. The search direction, instead of equation 3.48, is written as

$$\mathbf{s}^{(\kappa)} = -\mathbf{A}^{\kappa}\nabla f(\mathbf{x}^{(\kappa)}), \tag{3.49}$$

where $\mathbf{A}^{\kappa}$ is an approximation to the inverse of the Hessian.

The Davidon, Fletcher, and Powell (DFP) [56, 57] method, calculates $\mathbf{A}^{\kappa}$ starting from $\mathbf{A}^0 = \mathbf{I}$, where $\mathbf{I}$ is an identity matrix [54, 55] using

$$\mathbf{A}^{\kappa+1} = \mathbf{A}^{\kappa} - \frac{\mathbf{A}^{\kappa}\gamma^{\kappa}\gamma^{\kappa^T}\mathbf{A}^{\kappa}}{\gamma^{\kappa^T}\mathbf{A}^{\kappa}\gamma^{\kappa}} + \frac{\mathbf{s}^{(\kappa)}\mathbf{s}^{(k)^T}}{\gamma^{\kappa^T}\mathbf{s}^{\kappa}}, \tag{3.50}$$

where

$$\gamma^{\kappa} = \nabla f(\mathbf{x}^{(\kappa+1)}) - \nabla f(\mathbf{x}^{(\kappa)}). \tag{3.51}$$

The DFP method, although quite effective, was superseded by the BFGS method [54]. The BFGS updating method, named after the inventors Broyden, Fletcher, Goldfarb, and Shanno, constructs the approximations to the Hessian as

$$\mathbf{H}^{\kappa+1} = \mathbf{H}^{\kappa} - \frac{\mathbf{H}^{\kappa}\mathbf{s}^{(\kappa)}\mathbf{s}^{(k)^T}\mathbf{H}^{\kappa}}{\mathbf{s}^{(\kappa)}\mathbf{A}^{\kappa}\mathbf{s}^{(k)^T}} + \frac{\gamma^{\kappa}\gamma^{\kappa^T}}{\gamma^{\kappa^T}\mathbf{s}^{(\kappa)}}, \tag{3.52}$$

and approximations to the inverse of the Hessian as

$$\mathbf{A}^{\kappa+1} = \left(\mathbf{I} - \frac{\mathbf{s}^{(\kappa)}\gamma^{\kappa^T}}{\gamma^{\kappa^T}\mathbf{s}^{(\kappa)}}\right)\mathbf{A}^{\kappa}\left(\mathbf{I} - \frac{\gamma^{\kappa}\mathbf{s}^{(\kappa)^T}}{\gamma^{\kappa^T}\mathbf{s}^{(\kappa)}}\right) + \frac{\mathbf{s}^{(\kappa)}\mathbf{s}^{(\kappa)^T}}{\gamma^{\kappa^T}\mathbf{s}^{(\kappa)}}. \tag{3.53}$$

The BFGS method is currently considered to be most effective of all the Quasi-Newton methods [54]. Many other Quasi-Newton methods, such as the SR1 method, Broyden class updates, etc. have been proposed in the literature. These are not discussed in this thesis, but the reader is referred to the text by Nocedall [54] for details.

3.3.3 Conjugate Gradient methods

These methods originated with a view to solving the linear system of equations, $\mathbf{Ax} = \mathbf{b}$. The solution to this equation can be seen as the minimiser to the quadratic function, $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{Ax} - \mathbf{b}^T\mathbf{x}$ [54]. The underlying idea is that the solution can be found in exactly n searches, where n is the dimensionality of the problem, by sequentially searching along any n directions that are conjugate to each other with respect to the matrix $\mathbf{A}$. Two directions, $\mathbf{s}^{(i)}$ and $\mathbf{s}^{(j)}$ are conjugate with respect to $\mathbf{A}$ if $\mathbf{s}^{(i)^T}\mathbf{A}\mathbf{s}^{(j)} = 0$. The problem, however, is to find n directions that are conjugate to each other with respect to the matrix $\mathbf{A}$. Several techniques, such as eigen value decomposition, Gram-Schmidt orthogonalization, etc. can be used to find conjugate directions. However, this can be computationally expensive. To resolve this issue, the method of conjugate gradients (CG) was developed. The method of CG computes a new search direction, $\mathbf{s}^{(\kappa)}$, which is conjugate to all previous search directions, by using only the

last search direction, $\mathbf{s}^{(\kappa-1)}$. CG creates the new search direction, $\mathbf{s}^{(\kappa)}$, as a linear combination of the negative current residual, $-\nabla f(\mathbf{x}^{(\kappa)})$, and the previous search direction $\mathbf{s}^{(\kappa-1)}$ as follows

$$\mathbf{s}^{(\kappa)} = -\nabla f(\mathbf{x}^{(\kappa)}) + \frac{\nabla f(\mathbf{x}^{(\kappa)})^T \mathbf{A}\, \mathbf{s}^{(\kappa-1)}}{\mathbf{s}^{(\kappa-1)^T} \mathbf{A}\, \mathbf{s}^{(\kappa-1)}}. \tag{3.54}$$

Hence, in the above method, choosing the first direction to be the steepest descent direction and calculating subsequent directions as defined above, the CG method would find the minimum of a quadratic function in exactly n searches. Fletcher and Reeves [58] proposed a modification to calculate search directions so that non-quadratic functions could be optimised efficiently by the CG method. Their modification calculates successive search directions as follows

$$\mathbf{s}^{(\kappa)} = -\nabla f(\mathbf{x}^{(\kappa)}) + \frac{\nabla f(\mathbf{x}^{(\kappa)}))^T\ \nabla f(\mathbf{x}^{(\kappa)})}{\nabla f(\mathbf{x}^{(\kappa-1)}))^T\ \nabla f(\mathbf{x}^{(\kappa-1)})} \mathbf{s}^{(\kappa-1)}. \tag{3.55}$$

3.3.4 Non-gradient methods

Non-gradient methods are those which do not use the derivative information to guide the search process towards the minimum. These can be classified into pattern/direct search methods and evolutionary methods.

3.3.4.1 Pattern or Direct Search methods

These methods rely only on function evaluations around the current point, i.e. the function is evaluated in a neighbourhood and a decision, without evaluation of the gradient, is made whether the current point should be updated by a surrounding point or not. The two well known such methods are the pattern search by Hooke and Jeeves [59] and the simplex search by Nelder and Mead [60]. The Hooke and Jeeves pattern search works with search directions that span the search space, for example in an n-dimensional space n linearly independent directions are required. The search works by a series of exploratory moves that are changed heuristically [55]. In each iteration, the current point is perturbed in each of the search directions sequentially (called the exploratory move). If a better point is found, the current variable is updated to the best point; otherwise the perturbation distance is decreased by a specified factor. When a better point is found after the exploratory move, a jump along the direction of the movement in the exploratory move is made, in hope of a further decrease in the function (the pattern move). Thereafter an exploratory search is carried out from the new point obtained by the jump: if a better point is found, the current point is updated; otherwise the perturbation distance is further reduced. This procedure is repeated until the perturbation distance becomes smaller than a pre-specified convergence limit.

The Nelder and Mead [60] method works with a simplex. This simplex is a geometrical shape with non-zero hypervolume and $n+1$ vertices in an n-dimensional space. The algorithm begins by evaluating the function at all the vertices of the simplex and calculating the worst point, best point, and the second best point. Thereafter, the simplex is changed by rules to guide the search process away from the current worst point and towards the better points. The four rules are reflection, expansion and reflection, contraction, and multiple contraction. In reflection, the worst point is reflected through the opposite face of the simplex such that the hypervolume of the simplex remains the same. In expansion and reflection, the worst point is reflected but is placed further away than the previous case, thereby resulting in an increase

of the hypervolume. In contraction one vertex is displaced towards the opposite face, and in multiple contraction, a face is displaced towards the opposite vertex. The process is repeated until the hypervolume of the simplex reduces to a size less than a pre-specified convergence limit.

3.3.4.2 Evolutionary methods

Evolutionary Algorithms (EA) are algorithms that mimic evolutionary processes found in nature for optimisation [36]. Keane and Nair [18] identify two features common to all evolutionary algorithms: first, that these algorithms employ random numbers, and hence the repeated application of the same algorithm with identical parameters and starting points yields different search trajectories to locate optimal solutions; and second, that all such algorithms are global optimisation algorithms, i.e. locally optimal solutions might be discarded during the search process. The most well-known EAs are genetic algorithms, simulated annealing, evolution strategies, evolutionary programming, and particle swarm optimisation. Here, one of these methods, genetic algorithms, is discussed.

3.3.4.3 Genetic algorithms

Genetic algorithms (GAs), due to their wide applicability and documented success, have become a popular choice for optimisation procedures [36, 61]. As opposed to classical algorithms for optimisation that work with a single point at a time, GAs work with a population of points. An initial population is first created randomly in a search domain; then, at each iteration the members of the this population undergo *selection, crossover, and mutation* to create the next generation. As this process is repeated, the average *fitness* (representing the goodness of solution(s)) of the population increases, thereby moving towards the global optimum of the problem at hand. To represent solutions in a population, a representation of candidate solutions is needed. An important distinction among GAs is made in this regard. GAs that represent candidate solutions (phenotypes) as strings of binary bits (genotypes) are known as *Binary coded GAs*. In such GAs each binary string is mapped on to the real space in order to calculate fitness values. On the other hand GAs that work directly on real variables are called *Real coded GAs*. The working principles of a GA can be succinctly described by the following steps [36]:

Representaion: As mentioned above a candidate solution can either be represented as a string of binary bits (sections of which correspond to different variables), or real variables (no mapping needed).

Fitness assignment: Each candidate of the population is assigned a fitness value based on the objective function value at that point. Many schemes exist for such assignment, for example, fitness of a candidate can be equal to the objective function, an appropriately scaled value of the objective function, rank of the candidate in the population, etc.

Selection: The selection operator is the first genetic operator that is applied to the population in order to select candidates for a mating pool (i.e. for future genetic operators). Its goal is to create more copies of strong (having relatively high fitness value) candidates while diminishing the number of weak candidates (relatively low fitness value).

The simplest, but powerful, of selection operators is tournament selection. In this procedure, any finite number of solutions are picked from the population, and the winner (i.e. the candidate with best fitness among the picked candidates), is declared winner and copied to the mating pool. This procedure is repeated until the mating pool is full. Many other schemes exist for selection, such as roulette-wheel selection, stochastic remainder roulette-wheel selection etc.

Crossover: This is the next genetic operator that is applied to candidates in the mating pool. The goal of crossover is to create new candidate solutions (called the *child* solutions) from the members of the mating pool. One such scheme for binary coded GAs is the single-point crossover. In this scheme, two candidates are picked at random from the mating pool, and a random crossover site (between one and length of the binary string representation) is generated. Thereafter, with a probability (p_c) (crossover probability) the bits before the crossover site are exchanged between the picked candidates to create two new child solutions. These children are added to the new population and this procedure is repeated until the entire new population is created. The working principles of a single-point crossover were implemented by Deb et. al. [62] for real coded GAs. This operator, called the simulated binary crossover (SBX) operator, has capabilities to restrict child solutions within the bounds of the search variables and has been successfully applied in many test problems [15]. A number of other crossover operators have been proposed and analysed in the GA literature; see [63] for details.

Mutation: The mutation operator is next applied to the post-crossover population. The goal of a mutation operator is similar to that of the crossover operator, i.e. to create new solutions. For binary coded GAs the mutation operator involves flipping the bits of a candidate genotype with a, usually very low, mutation probability (p_m). For real coded GAs, various equivalents, have been proposed [36]. One such widely used operator is the Polynomial mutation operator proposed by Deb and Goyal [64].

To summarise, a GA starts with an initial population which is a random set of solutions in the search domain. This population then undergoes selection, crossover, and mutation, to create the next generation of the population, and this process is repeated. At each generation, the selection operator increases the number of relatively better solutions, the crossover operator combines features of these relatively better solutions to create child solutions, and the mutation operator changes the solution locally in the hope of creating better solutions. Although not explicitly guaranteed, the combined effect of these operators is to increase the fitness of solutions over generations, and consequently lead towards the optimum result [36].

3.4 Constrained optimisation

So far in this chapter, methods that solve only unconstrained problems have been discussed. In this section an overview of optimisation methodology in the presence of constraints is discussed (cf. equation 3.1). At this point, a few definitions are necessary. The Lagrangian function ($\mathcal{L}$) of the general problem of equation 3.1 is defined as

$$\mathcal{L}(\mathbf{x}, \mathbf{u}, \mathbf{v}) = f(\mathbf{x}) - \sum_{k=1}^{m} u_k \, g_k(\mathbf{x}) - \sum_{j=1}^{l} v_j \, h_j(\mathbf{x}) \tag{3.56}$$

or

$$\mathcal{L}(\mathbf{x}, \mathbf{u}, \mathbf{v}) = f(\mathbf{x}) - \mathbf{u}^T\mathbf{g}(\mathbf{x}) - \mathbf{v}^T\mathbf{h}(\mathbf{x}), \tag{3.57}$$

where u_k and v_k are the Lagrange multipliers, and the vectors $\mathbf{u} = [u_1, u_2, \ldots, u_m]^T$ and $\mathbf{v} = [v_1, v_2, \ldots, v_l]^T$ are the vectors of Lagrange multipliers for the inequality and equality constraints, respectively. The addition of Lagrange multipliers combines the constraints to yield an unconstrained optimisation problem. One also needs to differentiate between *active* and *inactive* constraints: an inequality constraint, say $g_i(\mathbf{x})$, is said to be active at a point $\mathbf{x}^*$ if the point falls on the constraint surface, i.e. $g_i(\mathbf{x}^*) = 0$; otherwise the constraint is called inactive. With this background, the first-order optimality conditions, also known as the Karush-Kuhn-Tucker (KKT) conditions, for a point $\mathbf{x}^*$ to be a local minimiser of equation 3.1 can be stated as follows [54]

Theorem 3.4.1. *If $\mathbf{x}^*$ is a local solution to equation 3.1, the functions $f(\mathbf{x})$, $g_i(\mathbf{x})$, and $h_i(\mathbf{x})$ are continuosly differentiable, and the set of active constraints at the point $\mathbf{x}^*$ are linearly independent, then there exist Lagrange multiplier vectors $\mathbf{u}$ and $\mathbf{v}$ with components u_k, $k = 1, 2, \ldots, m$ and v_j, $j = 1, 2, \ldots, l$ such that the following conditions are satisfied*

$$\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}^*, \mathbf{u}, \mathbf{v}) = 0; \tag{3.58}$$

$$g_k \geq 0 \qquad k = 1, 2, \ldots, m; \tag{3.59}$$

$$h_j = 0 \qquad j = 1, 2, \ldots, l; \tag{3.60}$$

$$u_k g_k = 0 \qquad k = 1, 2, \ldots, m; \tag{3.61}$$

$$u_k \geq 0 \qquad k = 1, 2, \ldots, m. \tag{3.62}$$

Equation 3.61 represents the complementary conditions. They imply that either u_k is equal to zero or the constraint g_k is active [54, 55]. The above stated KKT conditions form the basis of many constrained optimisation algorithms.

A wide variety of optimisation problems are formulated as linear programming (LP) problems. A linear program is an optimisation problem with a linear objective function and linear constraints [54]. The two types of methods used to solve linear programming problems are *active set methods* and *interior point methods*. The simplex method developed by Dantzig [65] (not to be confused with the simplex method by Nelder and Mead discussed earlier) is the most famous of the active set methods to solve LP problems. Among interior point methods Karmarkar's projective algorithm [66] is very popular. These methods are not discussed in detail in this thesis; the reader is referred to standard texts [54, 65, 66] for further details. In what follows an overview of non-linear constrained optimisation algorithms is presented. These algorithms can be grouped as

3.4.1 Quadratic Programming

Quadratic Programming (QP) methods aim to solve the quadratic problem, i.e. an optimisation problem where the objective function is quadratic and the constraints are linear. These are important in non-linear constrained optimisation as many optimisation algorithms, such as sequential programming methods, solve a sequence of quadratic sub-problems iteratively [54]. The methods to solve quadratic programming problems can be divided into three groups: *active-set, gradient projection, and interior-point* methods. Here these methods are discussed

briefly. A general QP problem can be written as

$$\begin{aligned} \text{Minimise} \quad & q(\mathbf{x}) = \tfrac{1}{2}\mathbf{x}^T\mathbf{G}\mathbf{x} + \mathbf{x}^T\mathbf{p} \\ \text{subject to} \quad & \mathbf{a}_j^T\mathbf{x} = b_j \qquad j = 1, 2, \ldots, l \\ & \mathbf{c}_k^T\mathbf{x} \geq d_k \qquad k = 1, 2, \ldots, m. \end{aligned} \tag{3.63}$$

If there are no inequality constraints, then by theorem 3.4.1 the set of KKT conditions can be written as

$$\begin{bmatrix} \mathbf{G} & -\mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} -\mathbf{p} \\ \mathbf{b} \end{bmatrix}, \tag{3.64}$$

where $\mathbf{A}$ is a matrix whose rows are given by $\mathbf{a}_j^T$, $j = 1, 2, \ldots, l$, and $\mathbf{u}$ is a vector of Lagrange multipliers for the equality constraints, and $\mathbf{b}$ is a vector containing b_j, $j = 1, 2, \ldots, l$. This system of linear equations, can be solved by factorisation methods (such as LU factorisation), or iterative CG methods.

In the presence of inequality constraints, many methods use the form discussed above as subproblems to be solved in each iteration. Active set methods are those which maintain an estimate of the inequality constraints that are active in each iteration. If the contents of the active set, i.e. indices of inequality constraints that are active, are known, then the inequality constraints that are not active can be ignored, the active inequality constraints can be posed as equality constraints, and the resulting problem can be posed as the above discussed form to be solved. However, the contents of the active set are not known *a priori*. Thus active set methods move from one iterate to other by solving a subproblem where some of the inequality constraints, called the working set, are treated as equalities [54], i.e. it is assumed that only these inequality constraints are active. After each iteration the contents of the working set are systematically modified to identify a solution for the QP.

In active set methods, the working set changes in each iteration usually changes only by one index. This makes the method require a large number of iterations to converge to the optimum point. Gradient projection methods remedy this problem by offering a rapid way of changing the working set, and are most effective when the only constraints in the QP are variable bounds [54]. In this method, a steepest descent search is made from the current point until a constraint is encountered. When a constraint is encountered, the search direction is 'bent' to remain in the feasible region. Thus, the search is made along a piecewise linear path, and a local minimiser, called the *Cauchy Point*, is located. The working set is updated to the inequality constraints that are active the Cauchy point. Thereafter, a QP subproblem is solved with this new working set at the Cauchy point to obtain the next iterate [54].

Interior point methods also work with the KKT optimality conditions (theorem 3.4.1). They convert the inequality conditions, equations 3.59 and 3.62, to equality conditions by introducing slack variables. Furthermore a log barrier term is added to the objective function to result in the following problem

$$\begin{aligned} \text{Minimise} \quad & \tfrac{1}{2}\mathbf{x}^T\mathbf{G}\mathbf{x} + \mathbf{x}^T\mathbf{p} - \mu \textstyle\sum_{k=1}^{m} \log s_k \\ \text{subject to} \quad & \mathbf{a}_j^T\mathbf{x} = b_j \qquad j = 1, 2, \ldots, l \\ & \mathbf{c}_k^T\mathbf{x} - d_k - s_k = 0 \qquad k = 1, 2, \ldots, m, \end{aligned} \tag{3.65}$$

where s_k are the slack variables corresponding to the inequality constraints, and μ is a barrier parameter. In the limiting case of $\mu \to 0$ the solution of the above problem is the solution to

the QP. Most interior point methods employ Newton-Raphson's iterations to solve the system of KKT optimality conditions for the above problem and update the value of μ to obtain the solution at the limiting case of $\mu \to 0$ [54].

The reader is referred to the text by Nocedal [54] for details of the aforementioned algorithms.

3.4.2 Penalty and augmented Lagrangian methods

In these methods, the objective function and constraints are combined together, in each iteration of the algorithm, to form an unconstrained optimisation problem [54, 55]. In penalty methods, the general form of the combined function can be written as

$$C(\mathbf{x}, R) = f(\mathbf{x}) + \Omega(R, \mathbf{g}(\mathbf{x}), \mathbf{h}(\mathbf{x})), \tag{3.66}$$

where R is a set of penalty parameters, and Ω is the penalty function [55]. Typical choices for Ω include, $R_i h_i(\mathbf{x})^2$ for equality constraints and $R_i \langle g_i(\mathbf{x}) \rangle^2$ for inequality constraints (the bracket operator $\langle a \rangle$ is equal to a if a is negative and zero otherwise) [55]. Typically, the penalty parameter R_i is changed in successive iterations.

The augmented Lagrangian method, also known as the method of multipliers, uses the following combined function for unconstrained optimisation at the κ^{th} iteration [55, 54]

$$\begin{aligned} C(\mathbf{x}, R, \sigma^{(\kappa)}, \tau^{(\kappa)}) = f(\mathbf{x}) &+ R \sum_{k=1}^{m} \left[\left(\langle g_k(\mathbf{x}) \rangle + \sigma_k^{(\kappa)} \right)^2 - \left(\sigma_k^{(\kappa)} \right)^2 \right] \\ &+ R \sum_{j=1}^{l} \left[\left(h_j(\mathbf{x}) + \tau_j^{(\kappa)} \right)^2 - \left(\tau_j^{(\kappa)} \right)^2 \right], \end{aligned} \tag{3.67}$$

where the bracket operator is as defined previously, and R is a penalty constant. The parameters σ_k and τ_j are varied in successive iterations as follows

$$\sigma_k^{(\kappa+1)} = \langle g_k(\mathbf{x}^{(\kappa)}) \rangle + \sigma_k^{(\kappa)}, \tag{3.68}$$

$$\tau_j^{(\kappa+1)} = h_j(\mathbf{x}^{(\kappa)}) + \tau_j^{(\kappa)}. \tag{3.69}$$

It can be shown that the solution obtained by the above procedure satisfies the first-order KKT conditions [55]. Moreover, this method yields the Lagrange multiplier values

$$u_k = -2R\sigma_k, \tag{3.70}$$

$$v_j = -2R\tau_j. \tag{3.71}$$

Many modifications of the above method exist [54], for example a linearised or quadratic form for the constraints and objective function at the current point can be used in lieu of the original forms, in order to simplify the problem of unconstrained optimisation.

3.4.3 Linearised search techniques

Linearised search techniques work by linearising both the objective and constraints at the current point, $\mathbf{x}^{\kappa}$. A non linear function, $f(\mathbf{x})$ can be linearised at a point,$\mathbf{x}^{\kappa}$, by the following [55]

$$f(\mathbf{x}) \approx f(\mathbf{x}^{\kappa}) + \left[\nabla f(\mathbf{x}^{\kappa}) \right]^T (\mathbf{x} - \mathbf{x}^{\kappa}). \tag{3.72}$$

The linearisation of both objective function and constraints results in a linear programming problem, which can be efficiently solved by the simplex method proposed by Dantzig [65]. One such algorithm is the Frank-Wolfe method [55], where to form the new iterate, $\mathbf{x}^{\kappa+1}$, a line search is performed in the direction from the old point to the point found by the simplex search of the linear programming problem. This procedure is repeated until convergence criterion are met.

3.4.4 Cutting plane method

In the cutting plane method the objective function is minimised by introducing a new variable y which is minimised in lieu of the original function [55]. To account for the objective function, $f(\mathbf{x})$, an additional constraint, $y - f(\mathbf{x}) \geq 0$ is artificially introduced. Thus the objective function of the modified problem is y, which is a linear function. The cutting plane method begins by ignoring all the constraints of the modified problem except the bounds. This is a linear-programming problem which is solved using the simplex method. Next, at this solution, the most violated constraint is identified, and linearised. Now the linear programming problem is solved again with the bound constraints and the linearised form of only the most violated constraint. This solution forms the next iterate. Next, the most violated constraint at the solution is found again and linearised. A new linear programming problem with the bound constraints, previous constraint, and the currently linearised constraint, is solved, to yield the next iterate. This process is repeated until the maximum constraint violation is below the specified tolerance. This method works only when the feasible search space is convex.

3.4.5 Feasible directions method

The method of feasible direction [55, 18, 67] also works with linearised forms of the objective function and inequality constraints, and is only applicable in the absence of equality constraints. The underlying idea of this method is that if the current iterate has no active constraints then a steepest descent direction should be used for search, but if there are active constraints then the search direction should also account for feasibility. A direction of descent at a point $\mathbf{x}^{\kappa}$ is a direction, $\mathbf{s}^{\kappa}$, for which $\nabla f(\mathbf{x}^{\kappa}) \cdot \mathbf{s}^{\kappa} \leq 0$. Similarly a direction, $\mathbf{s}^{\kappa}$, is feasible for the k^{th} inequality constraint if $\nabla g_k(\mathbf{x}^{\kappa}) \cdot \mathbf{s}^{\kappa} \geq 0$. The method of feasible direction achieves a balance between these by solving the following linear-programming problem at each iterate, $\mathbf{x}^{\kappa}$

$$\begin{aligned} &\text{Maximise} && \theta \\ &\text{subject to} && \nabla f(\mathbf{x}^{\kappa}) \cdot \mathbf{s}^{\kappa} \leq -\theta \\ &&& \nabla g_k(\mathbf{x}^{\kappa}) \cdot \mathbf{s}^{\kappa} \geq \theta \quad k \in \text{set of active constraints} \\ &&& -1 \leq s_i \leq 1. \end{aligned} \tag{3.73}$$

The solution of the above problem is used to perform a unidirectional search along the found direction to find the next iterate $\mathbf{x}^{\kappa+1}$. This process is repeated until no feasible direction can be found, i.e. when the solution to the above linear programming problem is $\theta \leq 0$.

3.4.6 Sequential Quadratic Programming (SQP) methods

Sequential programming methods (SQP) are one of the most effective methods to solve non-linear constrained optimisation problems [54, 68]. SQP methods solve a quadratic sub-problem in each iteration. A quadratic sub-problem of equation 3.1, at the current point $\mathbf{x}^\kappa$ has the following form

$$\begin{aligned} \text{Minimise} \quad & (\mathbf{r}^\kappa)^T \mathbf{d}_\mathbf{x} + \tfrac{1}{2} \mathbf{d}_\mathbf{x}{}^T \mathbf{B}_\kappa \mathbf{d}_\mathbf{x} \\ \text{subject to} \quad & \nabla \mathbf{h}(\mathbf{x}^\kappa)^T \mathbf{d}_\mathbf{x} + \mathbf{h}(\mathbf{x}^\kappa) = 0, \\ & \nabla \mathbf{g}(\mathbf{x}^\kappa)^T \mathbf{d}_\mathbf{x} + \mathbf{g}(\mathbf{x}^\kappa) \geq 0, \end{aligned} \tag{3.74}$$

where $\mathbf{d}_\mathbf{x} = \mathbf{x} - \mathbf{x}^\kappa$. One choice for vector $\mathbf{r}^\kappa$ can be the gradient of $f(\mathbf{x})$ at $\mathbf{x}^\kappa$. Similarly, the symmetric matrix $\mathbf{B}_\kappa$ can be chosen as the Hessian matrix of $f(\mathbf{x})$ at $\mathbf{x}^\kappa$. This, however, is not appropriate when the constraint functions are non-linear, as the objective function does not include constraint non-linearity information and all the constraints have been linearised. To account for the non-linearity of the constraints, while using the linearised form in the quadratic sub-problem, SQP methods use a quadratic form of the Lagrangian as the objective function [68]. Hence, $\mathbf{r}^\kappa$ is chosen to be the gradient of the Lagrangian function

$$\mathbf{r}^\kappa = \nabla_\mathbf{x} \mathcal{L}(\mathbf{x}^\kappa, \mathbf{u}^\kappa, \mathbf{v}^\kappa), \tag{3.75}$$

where $\mathbf{u}^\kappa$ and $\mathbf{v}^\kappa$ are current estimates of the Lagrange multipliers. Similarly, $\mathbf{B}_\kappa$ is usually chosen as an approximation to the Hessian of the Lagrangian rather than the actual Hessian. Methods like BFGS updating, discussed in the previous section can be used for that purpose.

The solution of the quadratic sub-problem is then solved to yield a direction $\mathbf{d}_\mathbf{x}$, which is used to construct the new iterate $\mathbf{x}^{\kappa+1}$ from the current iterate $\mathbf{x}^\kappa$ by taking a step from $\mathbf{x}^\kappa$ in the direction of $\mathbf{d}_\mathbf{x}$. To determine the length of the step, a *merit* function, $\phi(\mathbf{x})$, is needed. Typical choices for merit functions include penalty functions and augmented Lagrangian functions as discussed in section 3.4.2. A decrease in the merit function, $\phi(\mathbf{x})$, ensures that an improvement in the original non-linear problem has taken place. With this background, the basic SQP method can be outlined as follows [68]

Step 1 Start with approximations of $(\mathbf{x}^0, \mathbf{u}^0, \mathbf{v}^0)$ and $\mathbf{B}_0$. Set $\kappa = 0$. Choose a merit function, $\phi(\mathbf{x})$; set $\kappa = 0$.

Step 2 Formulate the quadratic programming problem, equation 3.74, at the current iterate $\mathbf{x}^\kappa$. Solve the quadratic programming problem using any method from section 3.4.1. Obtain $(\mathbf{d}_\mathbf{x}, \mathbf{d}_\mathbf{u}, \mathbf{d}_\mathbf{v})$ from the solution. $\mathbf{d}_\mathbf{u}$ and $\mathbf{d}_\mathbf{v}$ denote the difference between $\mathbf{u}$ and $\mathbf{v}$ found by the solution of quadratic sub-problem and the current $\mathbf{u}$ and $\mathbf{v}$.

Step 3 Choose step length α by solving the unconstrained optimisation problem of minimising the merit function along $\mathbf{d}_\mathbf{x}$, i.e. unconstrained minimisation of $\phi(\mathbf{x}^\kappa + \alpha \mathbf{d}_\mathbf{x})$.

Step 4 Calculate the new iterate by the following.

$$\begin{aligned} \mathbf{x}^{\kappa+1} &= \mathbf{x}^\kappa + \alpha \mathbf{d}_\mathbf{x}, \\ \mathbf{u}^{\kappa+1} &= \mathbf{u}^\kappa + \alpha \mathbf{d}_\mathbf{u}, \\ \mathbf{v}^{\kappa+1} &= \mathbf{v}^\kappa + \alpha \mathbf{d}_\mathbf{v}. \end{aligned}$$

Step 5 If convergence achieved, stop.

Step 6 Update $\mathbf{B}_\kappa$ using an update algorithm like BFGS to obtain $\mathbf{B}_{\kappa+1}$

Step 7 Increment $\kappa = \kappa + 1$; go to **Step 2**.

The SQP algorithm is used in chapter 7 for constrained optimisation of coronary stents.

3.5 Multiobjective optimisation

Multiobjective optimisation, as mentioned in the beginning of this chapter, is an optimisation procedure where the goal is to minimise more than one objective. Similar to the single objective optimisation procedures, both classical, i.e. point by point search, methods and evolutionary, i.e. population based, methods have been proposed to solve multiobjective problems. However, evolutionary approaches, owing to the inherent advantage of working with a population, and the fact the optimal solution is not one point but a set of points, have gained popularity for such problems. This section reviews some of the algorithms proposed for multiobjective optimisation.

In order to understand multiobjective optimisation, the concept of dominance and optimality needs to be defined. In a problem with q objectives, as in equation 3.2, a solution $\mathbf{x}^i$ is said to dominate $\mathbf{x}^j$ if both the following are true [36]

1. The solution $\mathbf{x}^i$ is no worse than $\mathbf{x}^j$ in all the objectives, $f_p(\mathbf{x})$, $p = 1, 2, \ldots, q$
2. The solution $\mathbf{x}^i$ is better than $\mathbf{x}^j$ in at least one objective

With the above definition of dominance, a non-dominated set of solutions from a set of solutions can be defined. The *non-dominated* set of solutions, in a set of solutions P, is the set of solutions that are not dominated by any member of the set P. If the set P consists of all the members in the search domain, then the non-dominated set of P is called the *Pareto-optimal* set [36]. In a non-dominated set of solutions, no solution can be deemed better than the other. Hence, the goal of a multiobjective optimisation algorithm is to find non-dominated sets of solutions, while being as close to the Pareto-optimal set as possible.

A range of algorithms – from classical gradient based algorithms to evolutionary algorithms – have been proposed to obtain non-dominated solutions. The easiest extension of single objective optimisation algorithms to multiobjective problems is the weighted metric approach. In this approach the multiple objectives are combined into a single objective by taking a weighted-metric of the objectives, for example weighted-sum or weighted deviations from an ideal solution (eg. Tchebycheff metric) [36]. By choosing different weights for objectives, different single objective optimisation problems can be formulated, and solved to obtain a set of non-dominated solutions. Haimes et. al. [69] proposed the ϵ-constraint method, and suggested minimising one objective while specifying other objectives as constraints to lie in small bounds. Das and Dennis proposed the normal boundary intersection method (NBI) method to find a uniform spread of solutions on the Pareto front [70]. In this method first, a simplex of the convex hull of all the individual minima (of all the objective functions) is found. Thereafter, starting from a uniformly distributed set of points on this simplex, the distance from each point towards the origin is maximised to find solutions that could potentially be non-dominated. Ismail-Yahaya and Messac [71] proposed the normal constraint

(NC) method, which was later improved by Messac et. al. [17, 72]. In this method, the utopia line/plane is first found (for a biobjective problem this is the line joining the individual minima in the objective space). Thereafter a set of uniformly distributed points are generated on the utopia line/plane and a constrained optimisation problem from each of these points is solved. The constrained optimisation problem minimises one objective, while all other objectives are formulated as constraints, normal plane/lines from the point on utopia plane, to restrict the search space.

Compared to the aforementioned classical algorithms, evolutionary algorithms perform significantly better on most problems, especially as the dimensionality of the search space increases [36, 73]. Of the many evolutionary algorithms proposed for multiobjective optimisation (see [36] for details), the most popular are the modified non-dominated sorting genetic algorithm (NSGA-II) [74] and the improved strength Pareto evolutionary algorithm (SPEA2) [75]. Zitzler et. al. [75] showed that SPEA2 and NSGA-II show best overall performance when compared to the then contemporary multiobjective algorithms (namely Pareto-envelope based selection algorithm (PESA) [76] and SPEA [77], the predecessor of SPEA2). Here, SPEA2 and NSGA-II are described.

3.5.1 Improved strength Pareto evolutionary algorithm (SPEA2)

The improved strength Pareto evolutionary algorithm (SPEA2) [75] is an improvement over its predecessor SPEA [77] in three aspects: improved fitness assignment, density estimation, and archive truncation. SPEA2 works with two sets: the GA population and an external archive set. Fitness is assigned to every member of the population, P, and archive set, $\bar{P}$, as follows. Every individual, i, in the union of these sets ($P \cup \bar{P}$) is assigned a strength value, $S(i)$, equal to the number of individuals i dominates. The raw fitness of every individual, i, is then determined by summing the strength values of all members in $P \cup \bar{P}$ which dominate the member i. Hence, a raw fitness of zero implies a non-dominated individual in the combined set. Thereafter, density estimation for each individual is performed, to differentiate individuals with identical raw fitness values. For this, the distances of each individual to all other individuals in $P \cup \bar{P}$ are calculated, and sorted in a list. The inverse of the k^{th} element of this list (σ_i^k), k being chosen by the user (analogous to the $k-th$ nearest neighbour method), is chosen as the density estimate, $D(i)$, i.e. $D(i) = 1/(\sigma_i^k + 2)$. The fitness of an individual is then assigned to be the sum of its raw fitness score and density estimate.

After calculating the fitness values for each member of the population and the archive, an archive update operation is performed. In this operation the non-dominated individuals in the set $P \cup \bar{P}$ are copied to the archive of the next generation. If the number of non-dominated individuals is less than the size of the archive size allocated, then, after filling the archive set with the non-dominated individuals, the remaining places are filled with the best dominated members from the set $P \cup \bar{P}$. On the other hand, if the number of non-dominated individuals is higher than the allocated archive size, a truncation procedure to remove non-dominated individuals iteratively is performed. This ensures that the number of non-dominated solutions copied to the archive is equal to the allocated archive size. With the fitness assignment and archive update defined as above, population size N, archive size $\bar{N}$, the SPEA2 algorithm can be defined as follows [75]

Step 1 Create a random parent population P_0 of size N; set the archive population $\bar{P}_0 = \phi$,

where ϕ is the null set; set generation counter $\kappa = 0$.

Step 2 Calculate fitness for all individuals in P_κ and $\bar{P}_\kappa$.

Step 3 Perform archive update to generate $\bar{P}_{\kappa+1}$ of size $\bar{N}$.

Step 4 Stop is $\kappa > \kappa_{max}$, i.e. if maximum generations is reached. Return the non-dominated individuals in $\bar{P}_{\kappa+1}$.

Step 5 Perform selection on $\bar{P}_{\kappa+1}$, binary tournament selection with replacement, to fill the mating pool.

Step 6 Apply crossover and mutation operators to the mating pool obtained from **Step 5**, to generate $P_{\kappa+1}$; go to **Step 2**.

3.5.2 Non-dominated sorting genetic algorithm

The non-dominated sorting genetic algorithm-II (NSGA-II) is one of the most widely used evolutionary multiobjective optimisation algorithm. NSGA-II, proposed by Deb et. al. [74], is an improvement the earlier proposed NSGA by Srinivas and Deb [78] over three areas: computational complexity, elitism, and need of sharing parameter specification. NSGA-II uses a fast non-dominated sorting algorithm to identify Pareto sets and a crowding comparison operator to maintain diversity in the population. The latter aspect is driven by the motivation to get a uniform distribution of the solutions in the Pareto set. Before describing the algorithm, the fast non-dominated sorting algorithm and the crowding distance operator are first described.

NSGA-II works by classifying the population into non-dominated sets. The first non-dominated set consists of all the non-dominated members of the population. The second set consists of all the non-dominated members of the population after removing the members of the first set. In this way the entire population is classified into different Pareto sets by assigning a non-domination rank, i_{rank}: a non-domination rank of 1 is assigned to all members of the first non-dominated set, non-domination rank of 2 to the next non-dominated set, and so on. NSGA-II uses a fast method for this classification. Initially, for each solution, p, in the popoulation two quantities: a) the domination count, n_p, which is the number of solutions that dominate the solution p; and b) the set, S_p, containing all the members of the population which are dominated by the member p, are calculated. Now, the first non-dominated set has already been obtained, i.e. all the members who have their domination count equal to zero. Thereafter, for each solution p with $n_p = 0$, the members of the set S_p are visited, and their domination count is reduced by 1. During this process all members q in the set, S_p, for which the domination count becomes zero, are stored in a separate list. This separate list constitutes the second non-dominated set, Q. This procedure of visiting each element of the set S_p, belonging to each member of the list Q, and reducing its domination count by 1 is now repeated. This process leads to classification of all the members of the population into non-dominated sets.

For preservation of diversity in the population, NSGA-II uses crowding distance. For each member, p, in a non-dominated set, identified by the aforementioned procedure, the crowding distance, $i_{distance}$, is the average distance between the two points on either side of the member, p, in its non-dominated set. For the corner points, $i_{distance}$ is assigned to be infinity. Based on this the crowding distance operator ($\prec_n$) is defined as follows [74]

$$\text{solution } i \prec_n \text{solution } j \text{ if}$$
$$i_{rank} < j_{rank} \text{ or}$$
$$i_{rank} = j_{rank} \text{ and } (i_{distance} > j_{distance})$$

Thus, between two solutions belonging to two different non-dominated sets, the solution with a lower non-domination rank is preferred; if the solutions belong to the same non-dominated set, then the lesser crowded solution is preferred.

With this background, the NSGA-II algorithm can now be defined

Step 1 Create a random parent population P_0 of size N;

Step 2 Sort P_0 into non-dominated sets, i.e. assign i_{rank} to each member i of N;

Step 3 Use the standard tournament, crossover, and mutation operators to form an offspring population Q_0 (see section 3.3.4.3) of size N;

Step 4 Set generation count, $\kappa = 0$;

Step 5 Generate combined population, R_κ by joining populations P_κ and Q_κ;

Step 6 Sort R_κ according to non-domination. Let the set $\mathcal{F}_i$ refer to all members with non-domination rank i;

Step 7 Create $P_{\kappa+1}$ by filling members from sets $\mathcal{F}_i$, startin with $i = 1$ and incrementing i by one. Identify the set $\mathcal{F}_l$, the last non-dominated set which, if included, will make the size of $P_{\kappa+1}$ will exceed N;

Step 8 Sort the last non-dominated set $\mathcal{F}_l$ based on crowding-distance operator, and fill the remaining places in the population, $P_{\kappa+1}$, with the least crowded members of the last front $\mathcal{F}_l$;

Step 9 Create offspring population $Q_{\kappa+1}$ by using the crowded-distance based tournament operator, and standard crossover and mutation operators;

Step 10 If κ is equal to the maximum generation count, stop; otherwise set $\kappa = \kappa + 1$ and go to **Step 5**.

NSGA-II is used in chapter 9 for a multiobjective optimisation study on coronary stents.

3.6 Surrogate assisted optimisation methodology

Having presented an overview of both surrogate modelling and optimisation algorithms in this chapter, in this section the general optimisation methodology with the help of surrogates is outlined. As mentioned in section 3.2, when the analysis times for the high-fidelity simulations (real response) are high, it is practically advantageous to construct a surrogate for the response. This surrogate can then be used by the optimisation algorithms in lieu of the high-fidelity simulations for approximating the response at all the points which need evaluation by the chosen optimisation algorithm. The general methodology for optimisation in such cases is [18]

Step 1: Choose a sampling plan to generate locations, $\mathbf{x^{(i)}},\ i = 1, 2, \ldots, p$ where the high-fidelity simulations will be run to evaluate the real response of the function at these points;

Step 2: Run high-fidelity simulations at $\mathbf{x^{(i)}},\ i = 1, 2, \ldots, p$, to calculate the corresponding function response, $y^{(1)}, y^{(2)}, \ldots, y^{(p)}$. Represent the inputs and outputs with the dataset, $\mathcal{D}^0 \equiv \{\mathbf{x^{(i)}}, y^(i)\},\ i = 1, 2, \ldots, p$; set update counter, $\kappa = 0$;

Step 3: Construct a GP surrogate model using the dataset, $\mathcal{D}^\kappa$ by methods outlined in section 3.2.3;

Step 4: Apply an optimisation algorithm to search the predictor, $\hat{y}(\mathbf{x})$;

Step 5: Verify the results of the **Step 4** by running the high-fidelity simulation on the predicted optimum;

Step 6: If stopping criterion are met stop; otherwise update[1] the dataset $\mathcal{D}^\kappa$ with additional points generated either in **Step 4** or other update methods [18, 50, 79] such as expected improvement criterion, maximum error, etc., to generate the appended dataset $\mathcal{D}^{\kappa+1}$; set update counter $\kappa = \kappa + 1$; go to **Step 4**.

3.7 Conclusions

In this chapter an overview of surrogate modelling and optimisation is presented. It is clear that in order to perform optimisation studies – single objective or multiobjective – on coronary stents, figures or merit, i.e. objective functions, need to be formulated which can be used to compare different stent designs. A major part of this thesis deals with such formulations. For example, in the next chapter a measure of haemodynamic alteration in a stented coronary artery is formulated. Similarly, in chapters 5, 6, and 8, objective functions from the point of view of structural analysis, flexibility, and drug-distribution are formulated, respectively.

[1]The update process of surrogate models is referred as *active learning* in the field of machine learning [18]

Chapter 4

Haemodynamics in stented vessels

This chapter[1] primarily deals with haemodynamic evaluation of coronary stents, i.e. to study how the flow features in stented vessels are related to the design features of a stent. This relation is made within the currently accepted paradigm for the causes of in-stent restenosis. In subsequent chapters, other factors including arterial injury, drug-distribution, and flexibility are evaluated. The aims of this chapter are

1. to set up a computational fluid dynamics (CFD) model to evaluate blood flow in a stented segment of a coronary artery,

2. to compare flow features relevant to restenosis (cf. section 2.3.1) in five different coronary stents, and

3. to formulate an objective function that quantifies haemodynamic alteration, relevant to restenosis, that can be used in optimisation studies.

This chapter begins by a review of studies – *in vivo*, computational, and experimental – that identify a potential link between altered haemodynamics and restenosis. Thereafter, the details of the CFD model are presented, and the results of the comparison of flow features between five different stents are discussed. Finally, a computationally measurable objective function that can be used to compare stents based on haemodynamic alteration is formulated.

4.1 Introduction

Though coronary stent designs have evolved significantly over the past two decades, they still face the problem of in-stent restenosis, formation of neo-intima within six-months of the implant. Studies show that biological response post stent implantation depends on various factors including the stent design and how it alters the haemodynamics. This chapter takes five different stent designs, representing different coronary stents used in clinical practice,

[1]The contents of this chapter are published in the below mentioned article. Dr. A.I.J Forrester, as the author's then co-supervisor, contributed to this article through discussions, and Dr. N. Curzen is a cardiac surgeon who actively advises the author regarding the clinical aspects of restenosis.
Pant, S., Bressloff, N. W., Forrester, A. I. J. and Curzen, N. The influence of strut-connectors in stented vessels : A comparison of pulsatile flow through five different coronary stents. *Annals of Biomedical Engineering*, 38:1893–1907, 2010.

and explores the haemodynamic differences arising due to the differences in their design. Of particular interest is the design of the segments that connect two struts. Pulsatile blood flow analysis is performed for each stent, using 3-D CFD, and various flow features viz. recirculation zones, velocity profiles, wall shear stress (WSS) patterns, and oscillatory shear indices are extracted for comparison. Vessel walls with abnormal flow features, particularly low, reverse & oscillating WSS, are usually more susceptible to restenosis. Unlike previous studies, which have considered the effect of design parameters such as strut-thickness and strut-spacing on haemodynamics, this chapter investigates the differences in the flow arising purely due to differences in stent-geometry, other parameters being similar.

The following subsections review the studies which show the effect of flow disturbance on restenosis.

4.1.1 Effect of stent design: *in vivo* studies

Kastrati et al. [80] analysed 4,510 patients with stent implantations and showed that vessel size and stent design were first and second most important factors affecting restenosis, respectively. In their analysis, which predominantly used eight different types of stents, it was found that, depending on the stent design, the incidence of restenosis varied between 20.0% to 50.3%. Rogers and Edelman [81] studied denuded rabbit iliac arteries after implanting steel stents for 14 days. They reported that stent material and configuration were critical factors in determining intimal hyperplasia and thrombosis. By comparing two stents, with the same surface area and mass but different strut-strut intersections, they reported that reducing strut-strut intersections significantly reduced vascular injury, thrombosis, and neointimal hyperplasia. Kastrati et. al. [82] performed an analysis over 651 patients, in which coronary stents of similar designs but different strut-thickness were implanted, and reported that reduction in strut-thickness resulted in significant reduction in angiographic and clinical restenosis.

From the point of view of altered haemodynamics, a significant body of evidence suggests that sites with low mean shear stress, oscillatory shear stress, high particle residence times, and non-laminar flow are the sites where most intimal-thickening occurs. Ku et. al. [83, 84] reported a strong correlation between low mean wall shear stress (less than 0.5 Pa) and atherosclerotic intimal thickening in human carotid bifurcations. They also reported that regions experiencing oscillating shear stress may enhance atherosclerosis. Moore et. al. [85] studied intimal thickening in 15 post-mortem aortas and reported that in the region of the infrarenal aorta, vessel regions exposed to low mean and oscillating WSS are more inclined to the development of plaque when compared to regions exposed to high WSS. Wentzel et. al. [86] studied neointimal thickness in 14 patients after 6-months of Wallstent implantation. They used a 3-D reconstruction of arteries to determine neointimal thickness and computational flow analysis to calculate shear stress on the surface of the stent. For 9 out of 14 implantations they observed that neointimal thickening and in-stent shear stress were inversely correlated. The aforementioned studies suggest that stent design, which leads to specific WSS patterns, is a key determinant of restenosis rates.

4.1.2 Effect of stent design: studies on altered haemodynamics

This section reviews computational and experimental studies that have identified a link between stent design and haemodynamic alteration.

4.1.2.1 Computational Studies

Computational fluid dynamics provides an excellent tool for studying micro features of the flow and has been widely used for flow analysis through stented vessels. Berry et. al. [87] performed experimental and 2-D computational flow analysis using custom-made models of a braided wire stent, Schneider Wallstent®, to reveal flow separation and formation of stagnation zones between wires. They studied the effect of wire spacing and diameter on the stagnation zones and reported that stent geometry had a significant effect on arterial haemodynamics. In particular, their findings concluded that the fluid stagnation zones are continuous if wire spacings in the stent design is less than six wire diameters. Ladisa et. al. [88] performed steady state 3-D CFD simulations in a Palmatz-Schatz slotted-tube stent using data from *in vivo* measurements of canine left anterior descending coronary artery diameter and blood flow velocity. They reported that regions of low wall shear stress are localized around stent struts. They also suggested that angled struts that are aligned in the direction of the flow could minimise flow disturbances. In another study they [89] reported that while reducing the number of struts and strut-thickness reduced the percentage of arterial wall area exposed to low wall shear stress, the opposite was observed if strut-width was decreased. Rajamohan et. al. [90] studied pulsatile & non-Newtonian blood flow through a stent with a helical strut matrix and identified recirculation zones immediately upstream and downstream of each strut intersection. Their investigation suggested that such areas could be more susceptible to restenosis. Similar other studies [91, 92, 93, 94] have shown that stents, depending on their design, cause significant alterations in haemodynamics leading to particular zones which could be susceptible to smooth muscle cell proliferation and restenosis.

Seo et. al. [91] studied haemodynamic disturbances induced by stents in straight and curved segments of vessels. They suggested that in curved segments the difference in WSS patterns between the outer and inner walls depends on the vessel curvature and the flow Reynolds number in a complex manner. They studied the flow for two different designs, one with spiral structure and one with an intertwined ring structure, and revealed that the stent-design had a major effect on the flow disturbances. Balossino et. al. [95] modelled expansion of four different stents against plaque and artery using finite element analysis and used the expanded geometries to evaluate the haemodynamics. They compared the WSS distribution for these stent models and also studied the effect of strut-thickness on vessel haemodynamics. In particular, they reported high wall shear stress values at the centre of the stent cells and a decreasing trend in wall shear stress as one moved away from the centre towards the stent struts. The aforementioned computational studies re-affirm the role of stent design in determining WSS in stented artery segments, and hence restenosis rates.

4.1.2.2 Experimental Studies

Some experimental studies have been performed which employ various methods like dye injection flow visualization and particle image velocimetry. Peacock et. al. [96] used an *in vitro* pulse duplicator system to investigate flow disturbances in the downstream region

of stented segments and used alumina particles for flow visualization. They detected flow instabilities 1 cm downstream, and no instabilities were found 5 cm downstream, for mild-exercise conditions. Berry et. al. [87] used dye injection flow visualization and reported accumulation of dye at each strut-strut intersection. Benard et. al. [97] used a programmable pump and particle image velocimetry to investigate laminar flow over stents and identified zones around the stent struts that had low WSS. Such experimental studies imply a correlation between the stent design and flow instabilities in stented vessels, particularly in and around the stent-struts.

4.1.3 Studies on endothelial cell response

Relatively few studies exist which investigate the response of endothelial cells to haemodynamics in stented vessels. DePaola et. al. [98] showed by in vitro experiments that the vascular endothelium responds to shear stress gradients. They reported that endothelial cells migrate from areas where shear stress is low but the shear stress gradient is large, and that cells remaining in such regions divide at a faster rate compared to the cells exposed to uniform shear. Nagel et. al. [99] reported that alteration of blood flow leads to differential WSS gradients that modulate endothelial gene expression at atherosclerotic prone sites. Yeh et. al. [100] compared the growth of endothelial cells on different stent materials. They concluded that endothelial cell growth and protein expression level varies widely depending on the metal sheet used. In particular, for all seeding levels, they found that endothelial cell growth was high for TiN and TiO_2 when compared to 316L stainless steel and nitinol.
Although many studies have tried to understand the effect of stent geometry on altered haemodynamics, most have focussed on the effect of strut spacing and strut thickness. This chapter deals with the effect of stent shape on haemodynamics. Connectors (mostly flex) are an essential component of a stent-design as their presence makes the stent flexible, which in turn improves stent deployment. With the new stent designs now used in clinical practice, especially drug eluting stents, there comes a need to study the effect of these connectors, along with the overall geometry, on haemodynamics. In order study these effects a CFD model is developed. This model forms the content of next section.

4.2 Methodology

4.2.1 Geometry

The five stents used in this chapter resemble the ART stent [8], Bx VELOCITY stent [34], NIR stent [34], the MULTI-LINK Zeta stent [34], and the Biomatrix stent [101]. These stents are chosen to represent a wide range of stent geometric shapes: the ART stent has straight connectors, the NIR, Bx VELOCITY, and MULTI-LINK ZETA stents have differnt complexity of the curved connectors, and the Biomatrix stent does not follow quarter-symmetry like the other stents. The details of each stent are listed in table 4.1.

In order to make a comparison between the stents, representative geometries for each stent are constructed with the same diameter (3 mm), length (8 mm), strut width (0.05 mm), and strut thickness (0.10 mm). Strut thickness represents the radial dimension of the struts and strut width represents the circumferential dimension. It should be noted that the strut spacings used in this chapter are purely representative and they are likely to differ, to some

Table 4.1: Stents: details

Stent	Manufacturer	Referred as
ART stent	Arterial Remodelling Technologies	Stent A
BxVELOCITY stent	Johnson & Johnson	Stent B
NIR stent	Boston Scientific	Stent C
MULTI-LINK Zeta	Abbott Vascular	Stent D
Biomatrix stent	Biosensors	Stent E

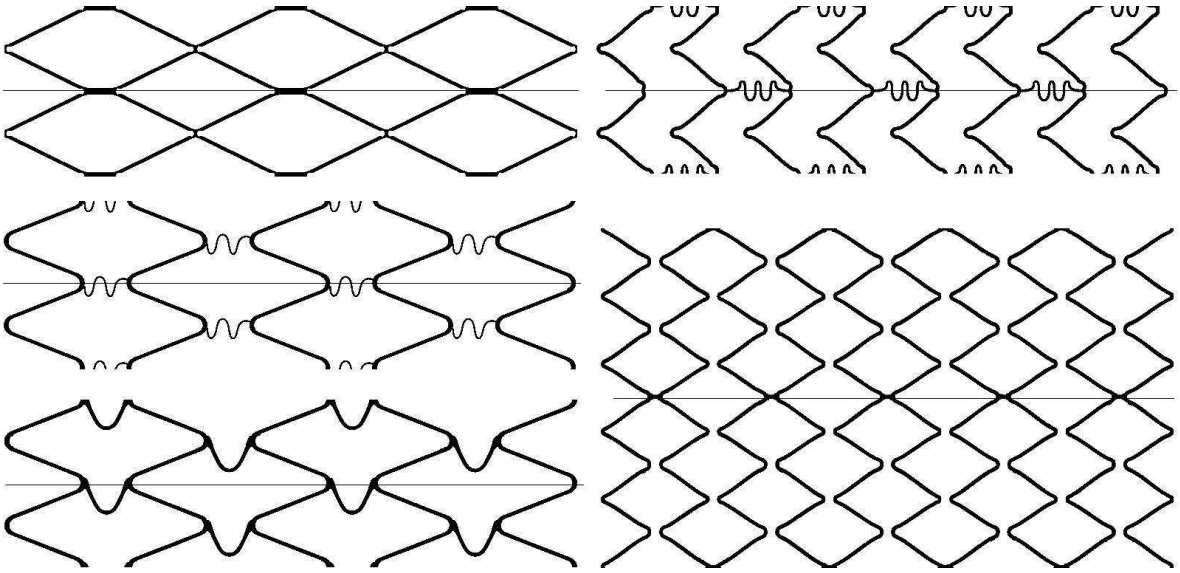


Figure 4.1: Flat geometries for the five stents – left top: Stent A, left mid: Stent B, left bottom: Stent C, right top: Stent D, right bottom: Stent E

extent, from the actual spacings for each of the stent designs. Figure 4.1 shows the flattened out geometries for one quarter of each stent except for Stent E, for which a half section is shown. The straight lines drawn in each model define the line on the artery wall along which wall shear stress and modified oscillatory shear index (MOSI) values are calculated and compared in subsequent figures. For all flow simulations the stent is placed at the centre of the artery with an axial distance of two times the artery diameter on both the proximal and distal ends of the stent. The artery wall is assumed to be straight with a constant diameter. Figure 4.2 shows the stent-artery assembly for Stent B. Numerical simulations are performed over a quarter of the stented segment for all stents, except for Stent E for which a half segment (owing to the quadrature links which do not allow quarter symmetry) is used, to exploit the periodic symmetry of the stent-artery assembly. All the geometries are constructed in Rhinoceros 4.0, a NURBS-based CAD modelling tool (1993 – 2008, Robert McNeel & Associates). Approach 2 (section 2.6.1.1) is used to construct Stent A while all the other stent models are created using approach 3 (section 2.6.1.2). Figure 4.3 shows the full 3-D models of the five stents.

4.2.2 Governing Equations

The following mass conservation (equation 4.1) and momentum conservation (equation 4.2) equations are solved over the computational flow domain of the stent-artery assembly:

$$\nabla.(\mathbf{v}) = 0, \tag{4.1}$$

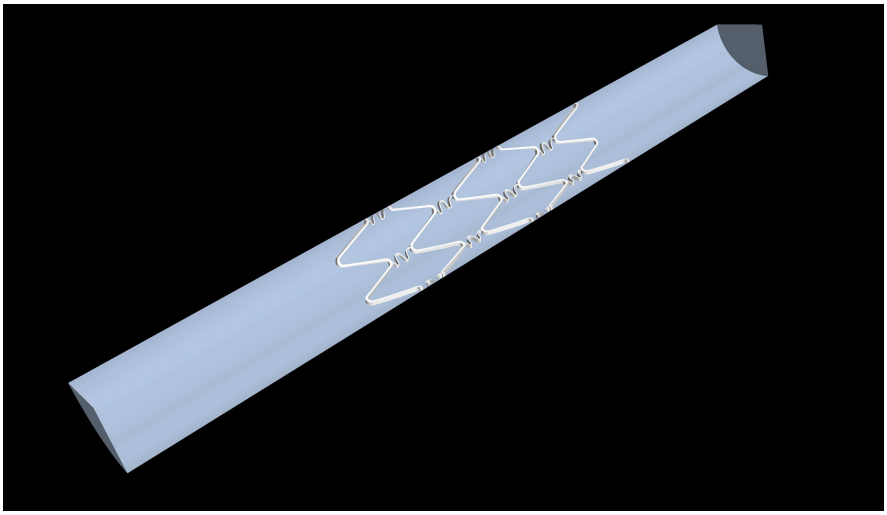
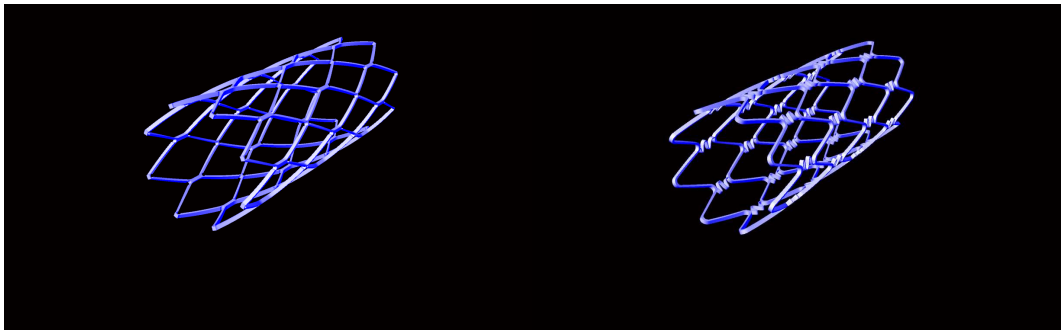
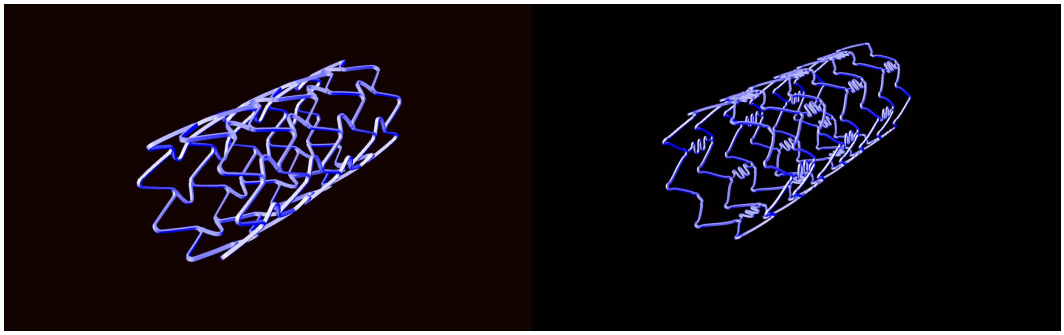


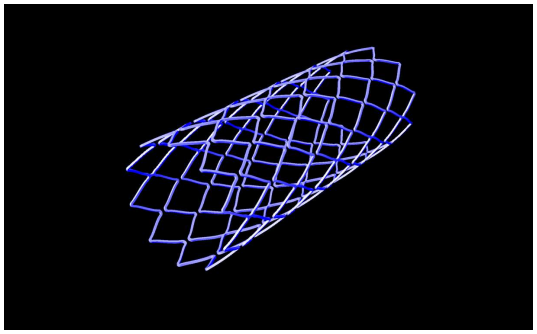
Figure 4.2: Stent-artery assembly for Stent B



(a) Stent A (b) Stent B



(c) Stent C (d) Stent D



(e) Stent E

Figure 4.3: 3-D models of the stents used in this chapter

$$\rho\frac{\partial \mathbf{v}}{\partial t} + \rho\mathbf{v}.\nabla\mathbf{v} = -\nabla P + \mu\nabla^2\mathbf{v}. \tag{4.2}$$

In the above equations $\mathbf{v}$, ρ, μ and P represent blood velocity, density, dynamic viscosity, and pressure respectively. Blood flow is assumed to be pulsatile, incompressible, laminar, and Newtonian with a density of 1.060×10^{-3} kg/m^{-3} and dynamic viscosity [88] of 3.7×10^{-3} Pa-s. Reynolds number (Re) and Womersley parameter (α) are defined as follows:

$$Re = \frac{\rho|\mathbf{v}|D}{\mu}, \qquad \text{and} \qquad \alpha = D\sqrt{\frac{\pi\rho}{2\mu T}}.$$

where D is the internal diameter of the artery and T denotes the time period of the cardiac pulse.

At every time, t, in the cardiac pulse the following two parameters are defined

$$p_l(t) = \frac{\iint_{wall} \tau^m_{w0.5}\, dA}{\iint_{wall} dA}, \tag{4.3}$$

$$p_r(t) = \frac{\iint_{wall} \tau^x_{w0.0}\, dA}{\iint_{wall} dA}, \tag{4.4}$$

where

$$\tau^{(\cdot)}_{wx} = \begin{cases} 1, & \text{if } \tau^{(\cdot)}_w <= x; \\ 0, & \text{otherwise.} \end{cases} \tag{4.5}$$

where τ^m_w and τ^x_w represent wall shear stress magnitude and axial wall shear stress, respectively. Hence, $p_l(t)$ and $p_r(t)$ denote the percentages of artery wall area exposed to WSS magnitude less than 0.5 Pa and axial WSS less than 0.0 Pa respectively. While $p_l(t)$ is a measure of low WSS in the artery wall, $p_r(t)$ measures the artery wall area exposed to reverse flow.

To compare the oscillatory nature of WSS, the modified oscillatory shear index [90], is calculated using the following equation:

$$MOSI = \frac{\int_0^T \tau^x_w dt}{\int_0^T |\tau^x_w| dt}, \tag{4.6}$$

where both integrals are calculated over one cardiac pulse.

4.2.3 Boundary Conditions

The outer wall of the stent is assumed to conform to the inner artery wall with no gaps. Both the stent and the artery wall are assumed to be rigid with a no-slip flow boundary condition imposed on each. A physiologically realistic coronary artery waveform is applied as the velocity inlet condition and the outlet is set to a zero pressure boundary. The inlet velocity profile is based on laser doppler velocimeter (LDV) measurements carried out in a replica of human LAD coronary artery [102]. These velocities are applied as plug flow at the inlet. Figure 4.4 shows the inlet velocity waveform where the eight points of interest are marked. Table 4.2 summarizes the key features of the inlet waveform.

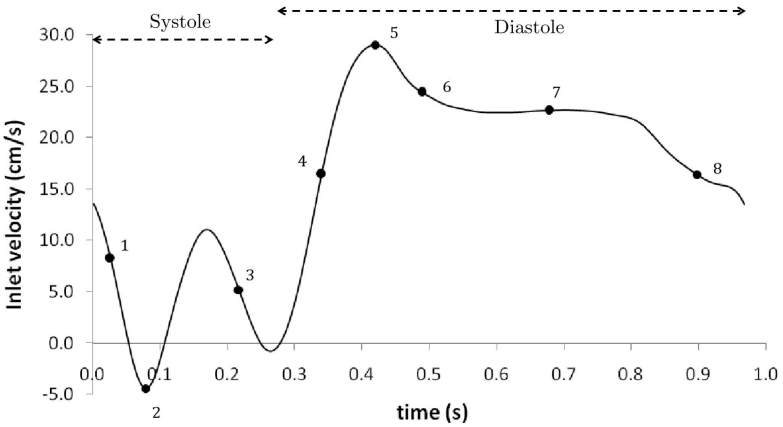


Figure 4.4: Inlet velocity profile showing the points of interest in one pulse

Table 4.2: Inlet velocity: key features

Feature	Value
Time period (T)	0.967 s
Mean Velocity	16.29 cm/s
Peak velocity	29.0 cm/s
Mean Reynolds number	140
Peak Reynolds number	249
Womersley parameter (α)	1.97
Eight points of interest	0.026s, 0.078s, 0.217s, 0.340s, 0.419s, 0.489s, 0.677s, and 0.897s

4.2.4 Computational Fluid Dynamics

Star CCM+ 3.06.006, CD-adapco, Melville, NY USA, a commercially available flow solver, is used for generating finite volume meshes and for numerically solving the governing equations. An implicit unsteady solver and segregated laminar flow solver (which uses the semi-implicit method for pressure-linked equations, SIMPLE, algorithm for pressure-velocity coupling), is used [103]. A second order temporal discretisation scheme is used for the transient term and a second order upwind scheme is used to discretise momentum. Under-relaxation factors of 0.7 and 0.3 are used for velocity and pressure, respectively.

Mesh, time-step, and pulse dependence studies are carried out for Stent C. Three different meshes are used: base, mesh-1, and mesh-2 (mesh-1 and mesh -2 have 1.5 and 2.5 times the number of cells relative to the base mesh, respectively). The WSS magnitude results for mesh-1 and mesh-2 vary by less than 1% as shown in Figure 4.5a.

Four different times steps viz. 10^{-2}s, 10^{-3}s, $5\text{x}10^{-4}$s, and 10^{-4}s are used for time-step dependence study on mesh-1, see Figure 4.5b. The maximum difference in WSS magnitude between time steps of 10^{-2}s and 10^{-3}s is nearly 30%. However differences in WSS magnitudes for time steps $5\text{x}10^{-4}$s and 10^{-4}s when compared to time-step of 10^{-3}s are less than 1%.

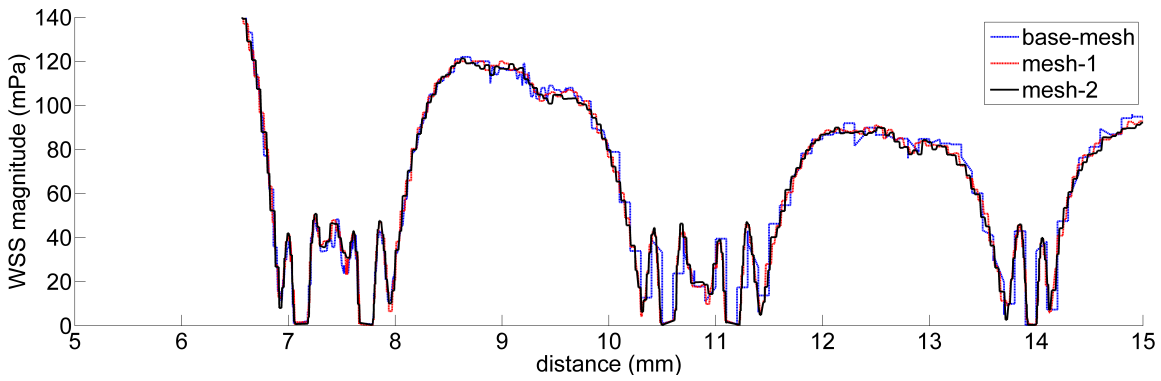
Simulations for five pulses are carried out for mesh-1 and the results show little variation after the second pulse as shown in Figure 4.5c. While the difference in WSS magnitude values for pulse 1 and pulse 2 is quite large, the difference in WSS magnitude for the 2^{nd} pulse onwards is less than 0.02%.

Based on the mesh, time-step and pulse dependence studies, all final simulations are run for two pulses for a time step of 10^{-3}s and mesh sizes as shown in table 4.3. For each time-step

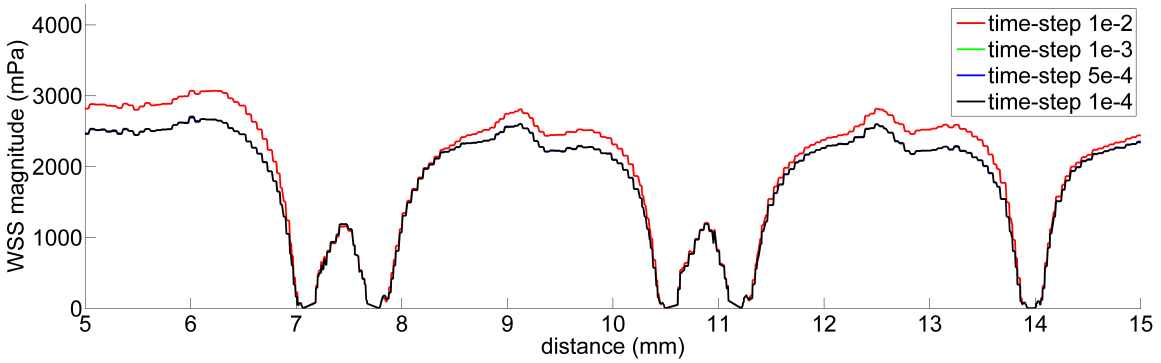
Table 4.3: Mesh statistics

Stent	Base size (mm)	Cell size in stent	No. of cells
Stent A	0.075	50% of base	1,076,793
Stent B	0.075	50% of base	1,097,951
Stent C	0.075	50% of base	1,097,788
Stent D	0.075	50% of base	1,031,211
Stent E	0.075	30% of base	1,855,559

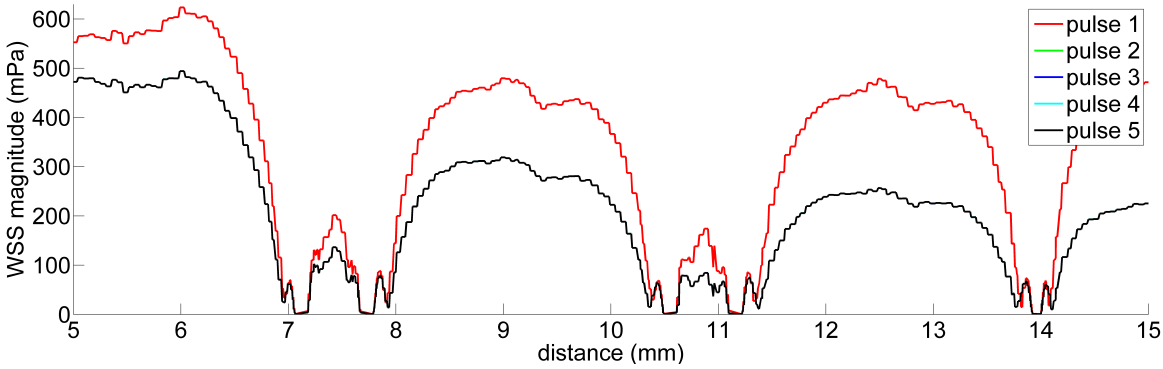
50 inner iterations of the SIMPLE algorithm [103] are carried out to achieve convergence.



(a) Mesh dependence study: WSS magnitude along central line on artery wall for Stent C; second point in Figure 4.4; second pulse; time-step $10^{-3}s$



(b) Time-step dependence study: WSS magnitude along central line on artery wall for Stent C; fifth point in Figure 4.4; mesh-2; second pulse



(c) Pulse dependence study: WSS magnitude along central line on artery wall for Stent C; first point in Figure 4.4; mesh-2; time-step $10^{-3}s$

Figure 4.5: Mesh, time-step, and pulse dependence studies for Stent C

4.3 Results

The flow features in the stented vessels are reported both qualitatively and quantitatively. In particular, differences in wall shear stress patterns, recirculation zones, and oscillatory shear indices are reported, thereby confirming the effect of stent design, especially the connectors, on haemodynamics of stented vessels. Furthermore, the connector design in Stent C is varied to study the effect of connector length, in the cross flow direction, on flow features.

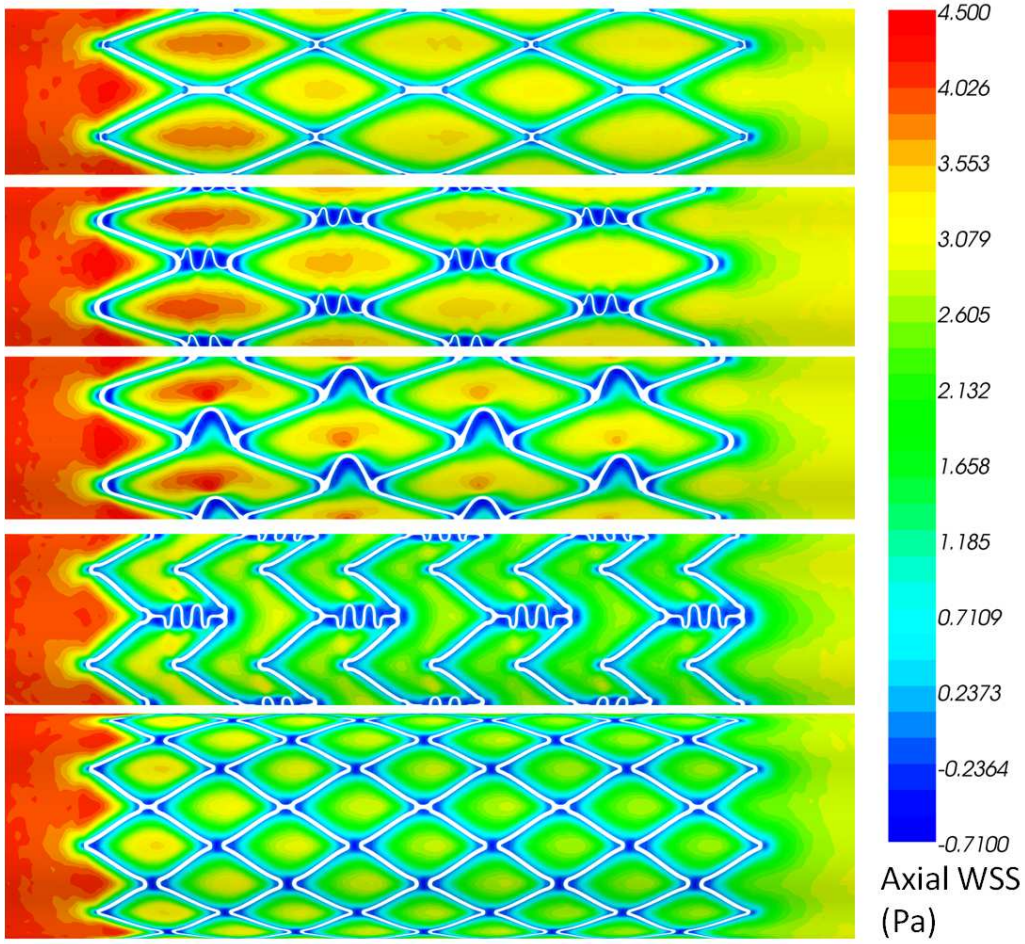


Figure 4.6: Axial WSS at point 5: Stents A-E from top to bottom

4.3.1 Wall Shear Stress

Wall shear stress follows a general trend for all the stents except for the regions between the connectors. Figure 4.6 shows the general axial WSS patterns for all stents at point 5, the point of maximum inlet velocity on the cardiac pulse. For all five stents, axial WSS has a high value proximal to the stent and a relatively lower value in the area occupied by the stent. Artery wall area distal to the stent experiences a higher axial WSS again as the flow disruptions minimize due to absence of stent struts. For Stents D and E, a larger artery area is exposed to relatively low WSS (green area in Figure 4.6 after the stent ends) , when compared to Stents A, B, and C, at the distal end of the stents. For all stents, and more notably for Stents A, B, and C in Figure 4.6, axial WSS at the centre of the struts decreases for consecutive struts in the direction of the flow (transition from red to yellow in consecutive struts). The artery wall region around the first strut experiences a relatively high WSS as

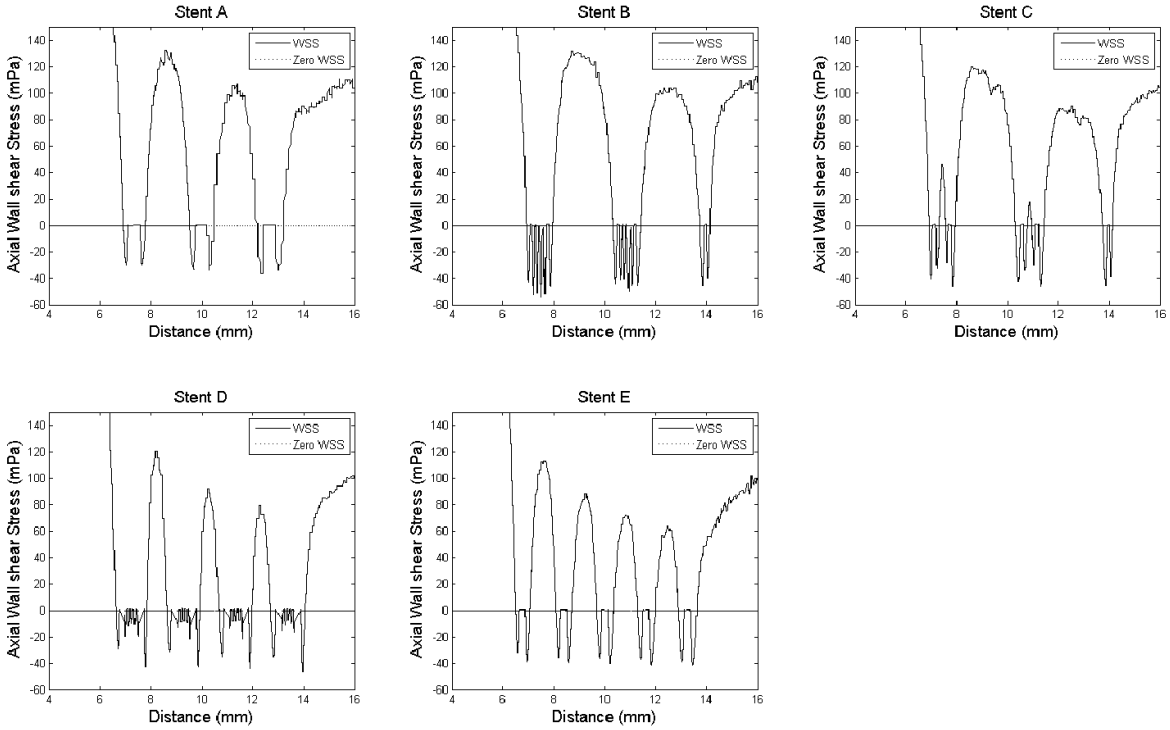


Figure 4.7: Axial WSS for all stents along the central line at Point 3

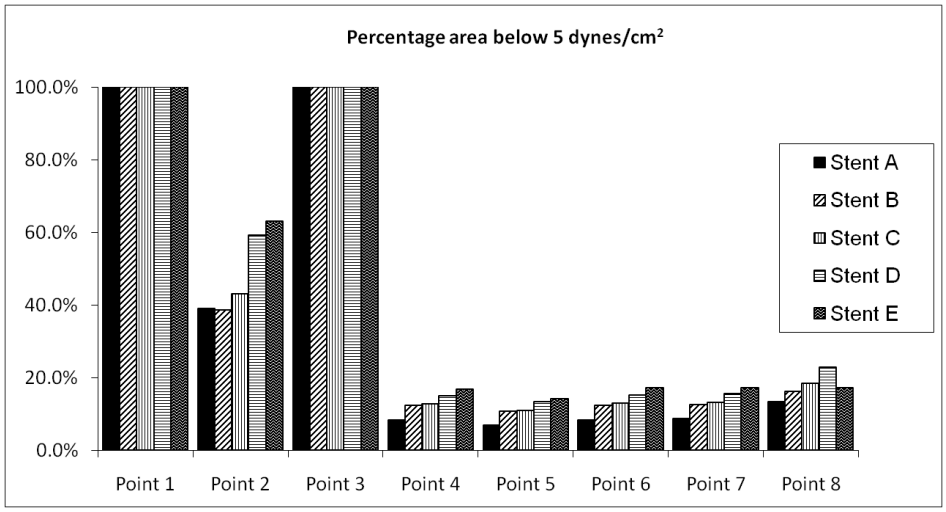
compared to the area around other struts. The areas of low WSS are found to be localised around the stent struts. This is in agreement with earlier findings of Ladisa et. al. [88] and Rajamohan et. al. [90].

Recirculation zones are formed at the proximal and distal end of each strut/strut-connector intersection, which cause the WSS to change sign before and after each strut/strut-connector intersection. In Figure 4.6 the blue regions show the artery wall area with negative axial WSS implying formation of recirculation zones. The phenomenon of recirculating flow is particularly significant in the decelerating phase of systole (point three, cf. Figure 4.4) as the recirculation zones are largest during this phase. Figure 4.7 shows the axial-WSS variation, along a central line on the arterial wall (as shown in Figure 4.1), for the five stents at point 3 of the cardiac pulse. For all stents the WSS values proximal and distal to the ends are the same. In between the struts WSS recovers from zero to a peak value which decreases for consecutive struts in the direction of flow. This peak value is different for all the stents and depends on the overall stent design. Depending on the design of the strut connectors WSS oscillates spatially in the connector region between zero, negative, and a positive value. In contrast to other stents, Stent C connectors allow the WSS to recover to a positive value in between the struts (as apparent in Figure 4.7). This can be attributed to the fact that Stent C connectors have more open space between the connectors.

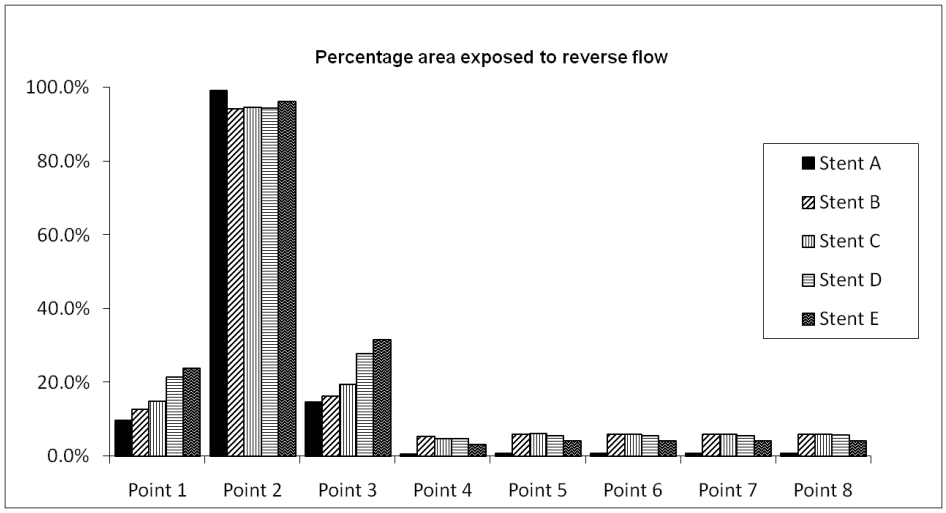
Previous studies [83, 104] suggest that areas exposed to WSS magnitude of less than 0.5 Pa correlate to areas that show most intimal thickening. Balossino et al. [95] have used this 0.5 Pa limit as a benchmark to compare the performance of stents. Figure 4.8a shows a histogram of the percentage of vessel wall area, over the axial length occupied by the stent, exposed to WSS less than 0.5 Pa, at the eight points listed in table 4.2. At points 1 and 3 this area is 100% irrespective of the stent as the flow is decelerating in the systole phase of the cardiac cycle. Point 2, also in the systole phase, shows unexpected behaviour of less than

100% area exposed to low WSS. However, the reason for this becomes clear when considering the negative WSS in Figure 4.8b. Other points show considerable difference in the area exposed to low WSS which can be used to compare their performance. Stent A outperforms the other stents at all points except Point 2. Stents D and E have a significantly higher percentage of low WSS area as compared to the other stents. While the difference between Stents B and C is not very large, Stent C has a slightly higher area exposed to low WSS.

Another factor that could promote restenosis is negative WSS caused by reverse flow. Figure 4.8b shows a histogram of the percentage vessel area exposed to reverse flow at the eight points for all stents. Point 2, owing to the negative inlet velocity and a hence strong reverse flow, has the highest percentage area exposed to reverse flow. While points 1 and 3 show no difference in terms of the 0.5 Pa WSS benchmark, these points show very significant differences in the area exposed to reverse flow. Stent A, although outperforming other stents at most points shows a near 100% area exposed to reverse flow at point 2.



(a) Percentage vessel wall area below 0.5 Pa WSS magnitude



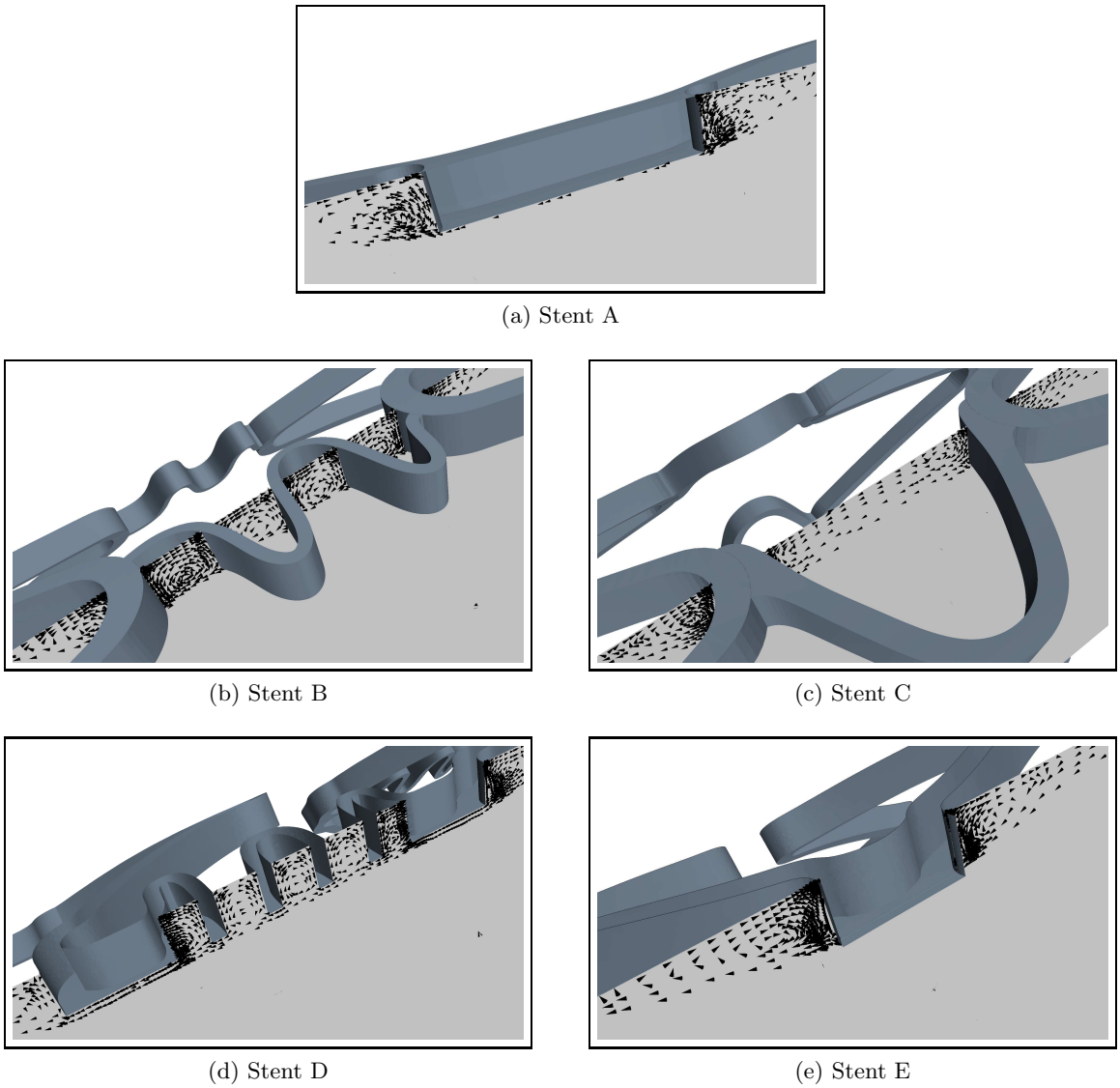
(b) Percentage vessel wall area exposed to reverse flow

Figure 4.8: Percentage vessel wall area exposed to low WSS & reverse flow

4.3.2 Recirculation zones

The presence of a stent inside the vessel gives rise to the formation of recirculation zones. Figure 4.9 shows the recirculation zones formed between the struts and the connectors of all

the stents at point 3 of the cardiac pulse. Each segment in the connector design gives rise to one recirculation zone. For instance Stent B has four recirculation zones in the connector region while Stent D has five.



(a) Stent A

(b) Stent B

(c) Stent C

(d) Stent D

(e) Stent E

Figure 4.9: Recirculation zones at point 3

Figure 4.10, shows the velocity profile adjacent to the artery wall for Stents B, C, and D at point 3 of the cardiac pulse. Recirculation zones in the cross flow direction are observed for these designs close to the artery wall. This can be attributed to the fact that the connectors in these designs, owing to their wavy nature, protrude into the space between the struts and hence cause more alteration in the flow.

4.3.3 Modified oscillatory shear index

For all the five stents modified oscillatory shear index is calculated using equation (4.6) along the central line in the artery wall as shown in Figure 4.1. MOSI is important as this index gives a time average value and hence is a measure of axial WSS over the entire pulse as opposed to single points in time. MOSI values of '1' or '-1' indicate that the axial WSS is positive or negative over the entire cardiac pulse respectively. Figure 4.11 shows a plot of MOSI values along the central line mentioned above. Each plot shows the MOSI values for

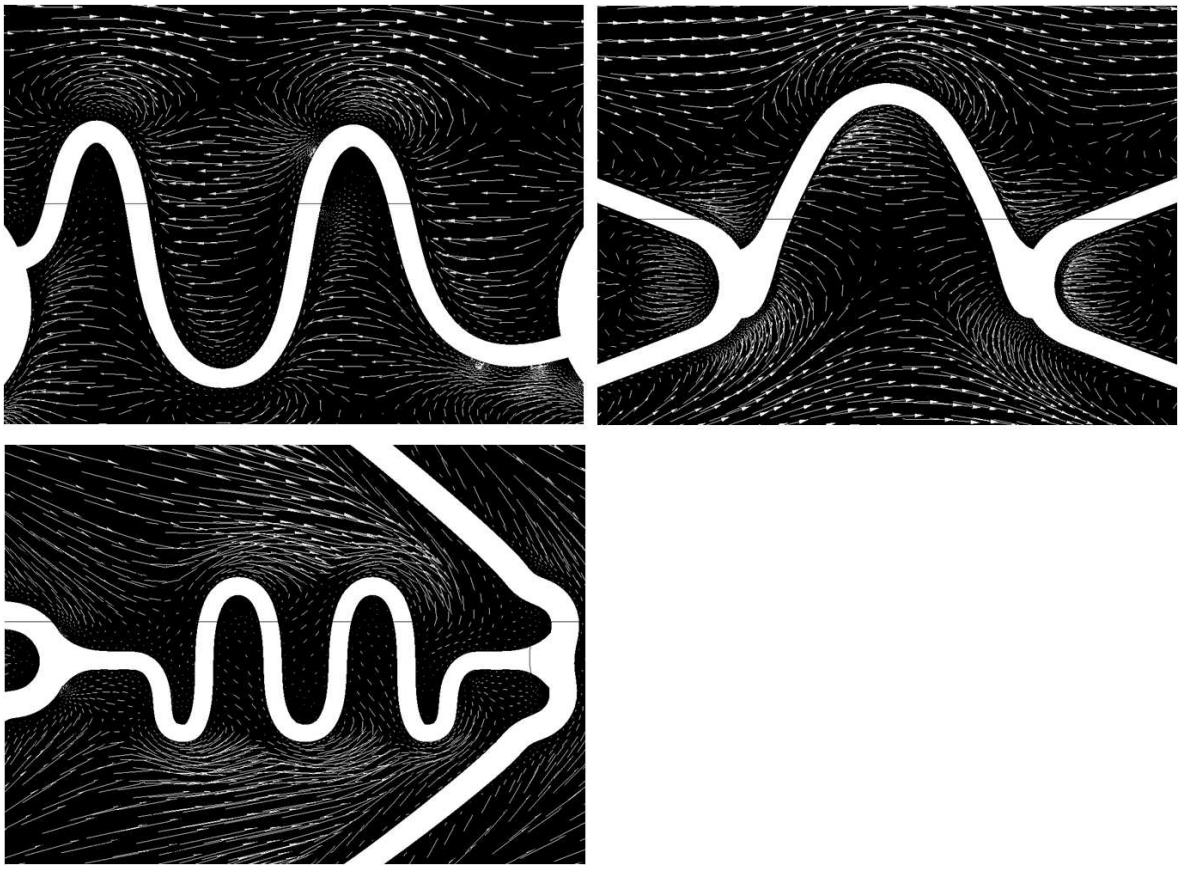


Figure 4.10: Stents B, C, and D : Secondary recirculation zones

the one connector-strut-connector segment from the assembly.

4.3.4 Variation in Stent C

In order to further investigate the effect of connector shape on haemodynamics, the design of the connector in Stent C is varied. Keeping the thickness constant, its length in the cross-flow direction is varied. This changes the area between the struts that is covered by the connector. Figure 4.12 shows the two altered designs – one with a shorter cross-flow length and one with a longer cross-flow length. These are referred as Stent C-SC and Stent C-LC, respectively. Simulations are carried out for these designs and the results are compared with Stent C. Figure 4.13 shows a comparison of percentage vessel area below 0.5 Pa and Figure 4.14 shows a comparison of percentage vessel area exposed to reverse flow for Stent C and its variations.

4.4 Discussion and formulation of objective function

WSS, recirculation zones, MOSI, and results for all stent designs are reported above. While the general qualitative features of WSS, such as localisation of low WSS regions around struts, match those described in earlier studies [88, 89, 90, 87, 92], this chapter brings forth finer differences at different parts of the cardiac pulse by comparing the factors that could have an effect on restenosis rates. Such differences, when compiled over the entire cardiac pulse, can be used to compare the relative haemodynamic performance of various stents.

Areas of low WSS (less than 0.5 Pa) and reverse flow are found for all the stents during the entire cardiac cycle. For points 1 and 3, 100% of wall area is exposed to low WSS. This

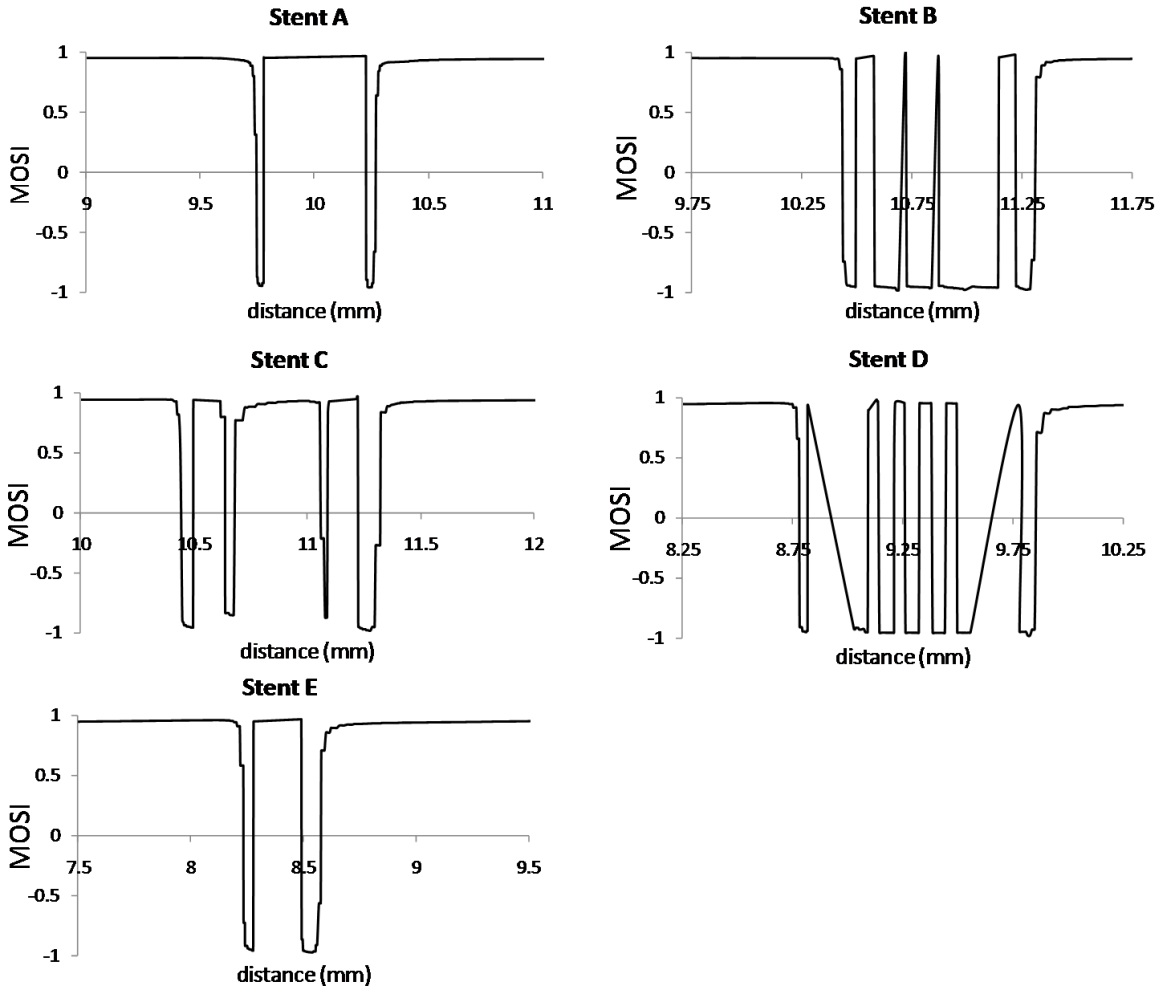


Figure 4.11: MOSI values for a strut with connectors in each side

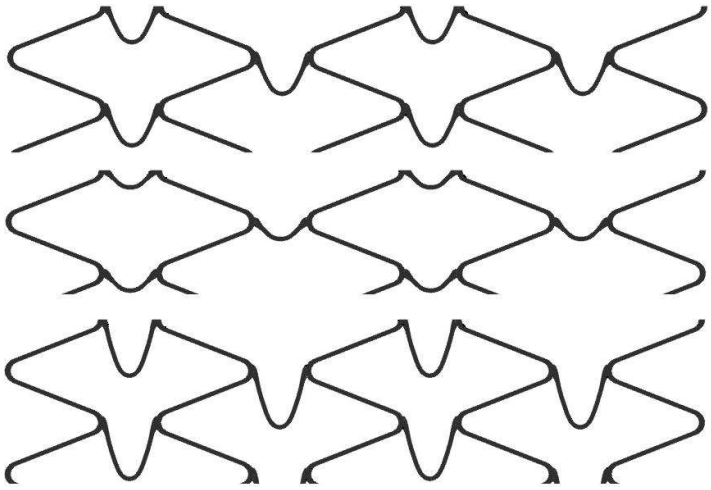


Figure 4.12: Variations in Stent C : top: normal, mid: shorter, and bottom: longer connectors

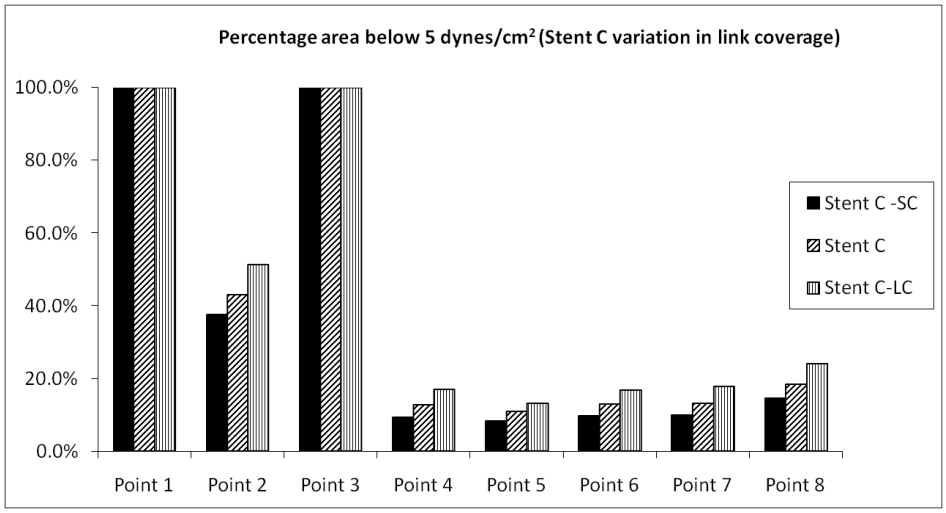


Figure 4.13: Area exposed to WSS magnitude below 0.5 Pa for Stent C variations

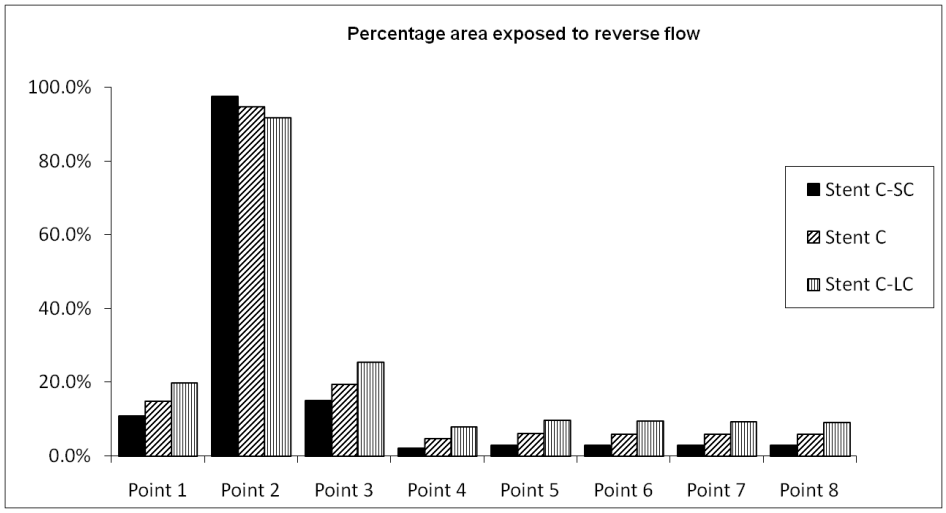


Figure 4.14: Area exposed to reverse flow for Stent C variations

can be attributed to the decelerating nature of flow at these points and that the inlet velocity is low. The peculiar behaviour of point 2, which too is in the decelerating phase and yet has a less than 100% area exposed to low WSS, can be explained by the fact that in this phase the inlet velocity gradients are high which cause a strong reverse flow thereby causing the axial WSS to be negative but higher than 0.5 Pa in magnitude. This is confirmed in Figure 4.15 for the following index, AWI, as defined below for one cardiac pulse–

$$AWI(t) = \frac{\iint_{wall} \tau_w^x(t) dA}{\iint_{wall} dA}. \tag{4.7}$$

The horizontal dashed lines bound the region with axial WSS magnitude less than 0.5 Pa. Since other components of WSS are very small when compared to axial WSS, the major factor determining the WSS magnitude is its axial component. Zones 1, 2, and 3 in Figure 4.15 mark the time zones in the systole phase of the cardiac pulse where the axial WSS magnitude exceeds ±0.5 Pa. Since Point 2 lies in zone 2, the WSS magnitude at some regions in the artery wall is greater than 0.5 Pa.

Returning to Figure 4.8a, points 4, 5 ,6, 7, and 8 also show considerable differences in the percentages of areas exposed to low WSS. Stents D and E stand out, both for low and reverse

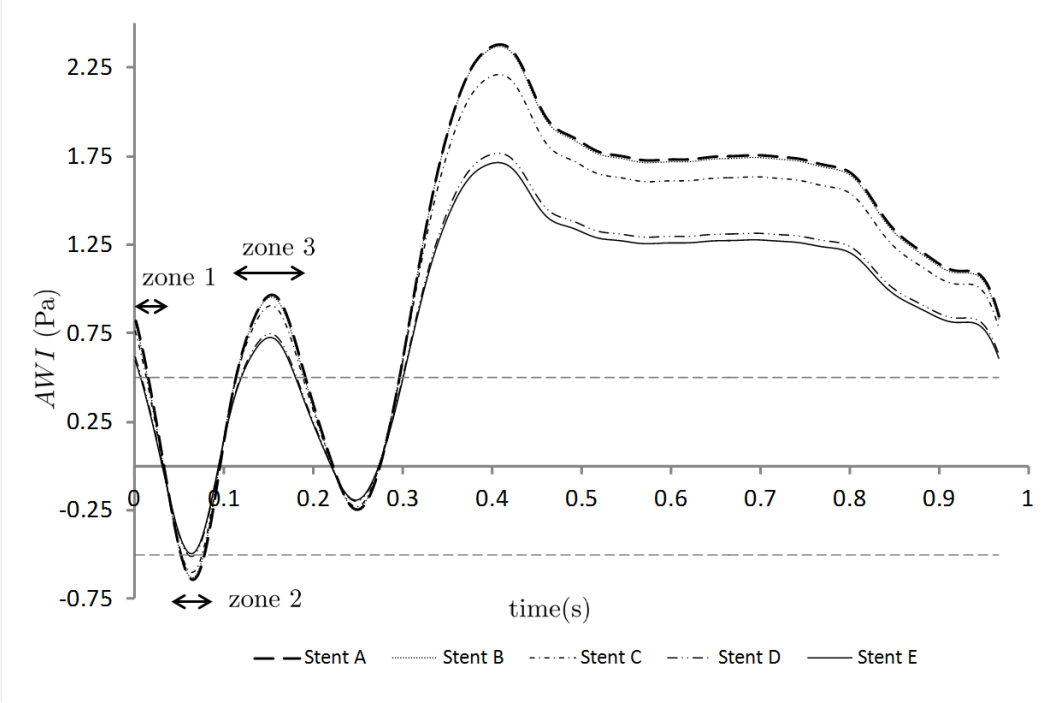


Figure 4.15: AWI for all stents for one cardiac pulse

WSS – because of the relatively lower strut spacing. However, even though Stents A, B, and C, have the same strut spacing, percentage areas exposed to low WSS differ significantly. Similarly, for all points there are significant differences in the percentages of area exposed to reverse flow between stents A, B, and C. This can be attributed to the difference in the design of the connector.

The connectors in Stents B,C, and D have a finite length in the cross flow direction – this cross-flow area coverage being largest for Stent C. Consequently, the struts tend to project into the central part of the space between struts. This causes a further disruption of the flow in that area - illustrated by Figure 4.10 which shows the velocity profile adjacent to the artery wall at point 3 in the cardiac pulse. Recirculation in the top ends of the connectors is clear in these designs. Such a phenomenon is absent in Stents A and E as the connectors are a straight segment joining the struts. The difference of such a protruding connector design is further confirmed when Stent C is altered to make the connector shorter and longer in the cross-flow direction (Stent C-SC and Stent C-LC). It can be seen in Figures 4.13 and 4.14 that areas exposed to low WSS and reverse flow are proportional to the connector length in the cross-flow direction.

Traditionally [90], MOSI has been used to quantify the oscillatory nature of WSS. In Figure 4.11 we see that MOSI takes a value close to '-1' at each strut-connector intersection and between the connectors. This implies incessant reverse flow or formation of recirculation zones over a large part of the cardiac cycle at such points. In Stents B and D, due to the presence of multiple gaps in the connector design multiple areas of persistent reverse flow are formed. This is consistent with the results of dye injection flow visualisation studies [87] where more dye accumulation was observed at each strut-strut intersection. The number of recirculation zones formed is directly related to the design, specifically the number of gaps either between struts or between the connector; see Figure 4.9 which illustrates this point. However, the recirculation lengths depend on the overall strut-connector-strut configuration.

Another factor which affects the extent of recirculation zones is the cross section of the struts. This was shown in a study [105] where stents with cross sections of a circular arc shape were compared with those having a rectangular shape. Streamlining of the strut cross section would reduce the size of the recirculation zones, and consequently reduce the areas exposed to low and reverse WSS.

An *in vitro* experimental study [98] showed that vascular endothelium responds to shear stress gradients. It was reported that endothelial cells migrate from areas where shear stress is low but the shear stress gradient is large, and that cells remaining in such regions divide at a faster rate compared to the cells exposed to uniform shear. Hence, the endothelial response to different WSS patterns created by different stents could be important in the process of re-endothelialisation. Furthermore, if the tangential components on a plane perpendicular to the flow direction are considered, additional recirculation of flow is observed; see Figure 4.16 which shows the in-plane velocity components, at the centre of the stent, at point 3. This additional recirculation of flow, although with velocity magnitudes of roughly $1/100^{th}$ of the inlet velocity, induces transverse WSS and WSS gradients which could have an effect on endothelial cell response. Qualitatively, Stents A and E, have minimum recirculation (when the tangential components are considered) on planes perpendicular to the flow direction.

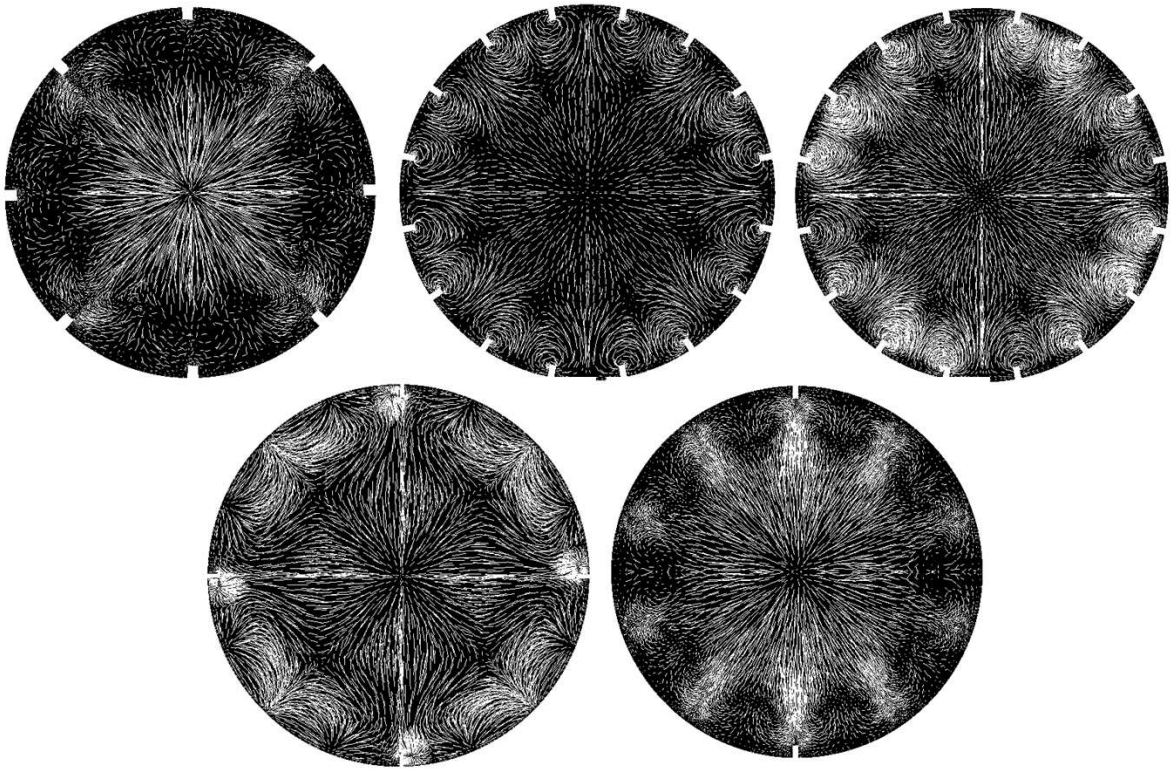


Figure 4.16: Recirculation zones on a cross section perpendicular to the flow direction: top left:Stent A; top mid: Stent B; top right: Stent C; bottom left: Stent D; bottom right: Stent E

From Figures 4.8a and 4.8b it can be concluded that differences in stent designs are apparent for areas of both low WSS and reversed flow. Keeping both the factors in mind and assuming that both the phenomena are equally unwanted, an index can be proposed which takes a weighted average of these percentages at the relevant points. Thus the haemodynamic low and reverse flow index ($HLRFI$) is defined as –

$$HLRFI = \frac{\sum_{i=1}^{n}(w_i p_{il} + w_i p_{ir})}{2\sum_{i=1}^{n} w_i}, \tag{4.8}$$

where w_i are the weights for each of the n points in the cycle, and p_{il} & p_{ir} denote percentages of areas exposed to low and reverse flow, respectively.

It is expected that the higher the value of n, the better will be the efficacy of the index in determining the haemodynamic alteration due to stents. The reason for taking a weighted average is that some specific points, such as the point of negative inlet velocity (point 2), could be clinically more relevant than others, and may require (a higher) differential weighting. The peculiar nature of such points on the cardiac pulse can be seen for Stent A, for which percentage area, exposed to both low WSS and reverse flow, at point 2, is abnormally high in reference to its relative performance at other points.

In order to capture the entire pulse, *HLRFI* can be modified as follows:

$$HLRFI = \frac{\int_0^T (w(t)p_l(t) + w(t)p_r(t)))dt}{2\int_0^T w(t)dt}, \tag{4.9}$$

where $w(t)$ is the weight function for the cardiac pulse and $p_l(t)$ & $p_r(t)$ is the percentages of artery area, A, exposed to low and reverse WSS at time t. $p_l(t)$ & $p_r(t)$ can be defined as in equations 4.3 and 4.4 respectively.

It should be noted that the wall area over which the surface integrals for $p_l(t)$ & $p_r(t)$ are calculated includes only the area exposed to flow, i.e. it exludes the wall area covered by the stent. Assuming the weight function to be unity, *HLRFI* for all the stents and Stent C variations, as calculated using equation 4.9, are shown in Figures 4.17 and 4.18 respectively. *HLRFI*, being one single number evaluated over the entire pulse, can be used to rank stents based on their haemodynamic performance. Lower *HLRFI* values indicate lesser alteration of haemodynamics, and hence better resistance of a stent towards restenosis.

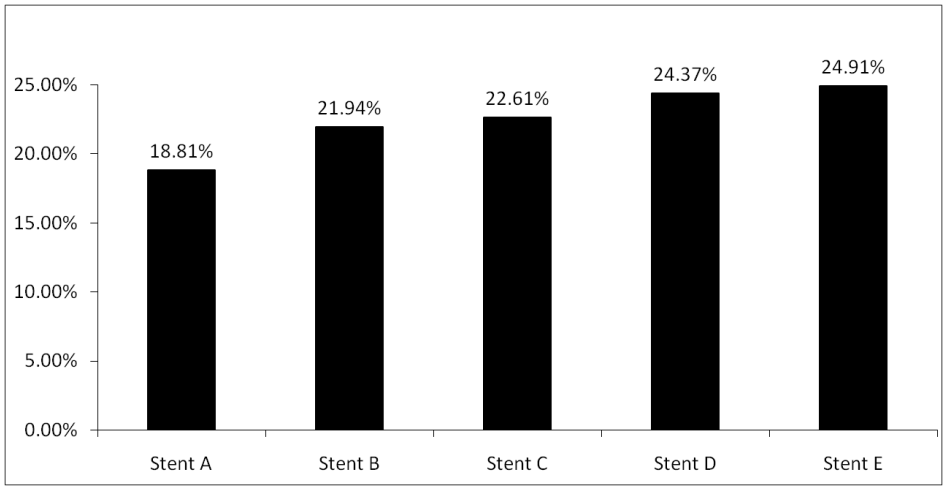


Figure 4.17: *HLRFI* for all the stents

As the length of Stent C connector is lowered in the cross-flow direction the haemodyanamic alteration decreases. This is reflected in decreasing *HLRFI* values for Stent C variations from 25.95% to 22.61% to 19.91%. This decreasing trend tends towards a value of 18.81% for Stent

A, which can be seen as a Stent C variation with minimal connector length in cross-flow direction. It is also interesting to note that Stent C-LC has the largest *HLRFI* value, even higher than Stents D and E which have a shorter strut spacing. This further emphasizes the effect design of strut connectors can have on stented vessel haemodynamics.

It is clear from the above findings that stent design dictates haemodyanamic alteration. Although strut thickness and spacing are the most important factors, blood flow depends strongly on the shape of the struts and the connectors. Strut thickness is governed mostly by material properties of the stent to minimise post-expansion recoil and manufacturing processes. Strut spacing is governed by the constraints of structural strength and flexibility. Thus the shape of the struts and connectors can be varied to improve the haemodynamic performance. It is important to be conscious of the fact that changing the stent design impacts other properties too, especially drug distribution. For instance, haemodynamic results for Stent C, C-SC, and C-LC, show that Stent C-LC has poor haemodynamic performance – however, it is likely to have better drug distribution potential as the links cover a larger wall area in the cross-flow direction.

Significant differences exist between the stents with regards to the number and extent of recirculation zones in the directions of both axial and cross-flow. Although it is not currently very clear how endothelial cells respond to complex flow phenomenon, it is possible that restenosis rates could be affected by them. It is notable that Stent A produces minimal alteration of flow both in the axial direction and the direction perpendicular to the main flow. This is reflected in its lowest *HLRFI* value (Figure 4.17) and minimal recirculation in the direction perpendicular to the main flow (Figure 4.16). This behaviour can be attributed to the fact that Stent A has straight segments as connectors between the struts. These straight segments, being aligned in the direction of the flow, disturb the flow to a lesser extent when compared to other connectors which, owing to their wavy nature, do not align completely with the direction of the flow.

In order to rank stents, an objective function (figure of merit) is needed which quantifies the flow features, and hence determines the patency of stents. In the past, relatively few metrics have been defined to quantify the distribution of WSS in arterial flow. One such metric is defined by Bressloff [106] to quantify relevant WSS information in a human carotid bifurcation. Along similar lines, the proposed *HLRFI* index captures and quantifies the two

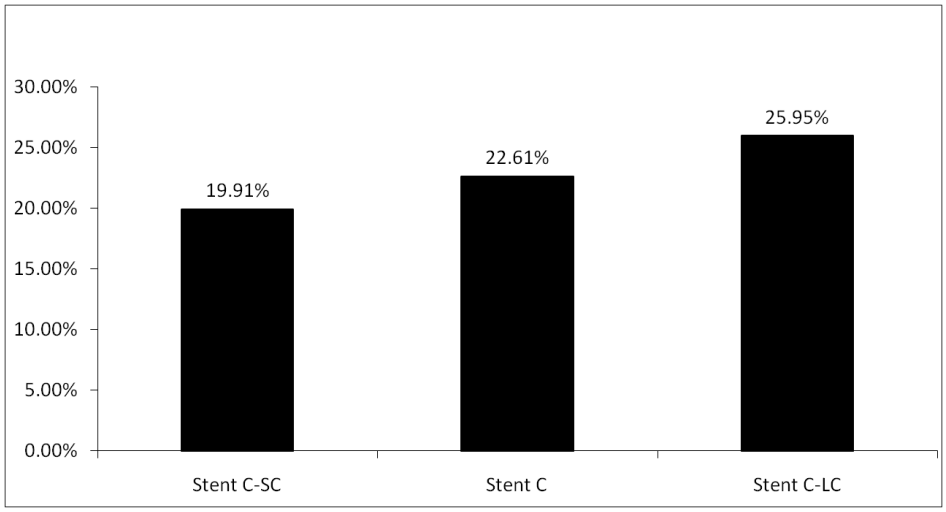


Figure 4.18: *HLRFI* for all the Stent C variations

phenomena of low and negative WSS which are detrimental to the resistance of a stent against restenosis. *HLRFI*, defined as in equation (4.9), can be used as an objective function to compare a family of related stent designs solely on their haemodynamic performance. For instance, Stent E, with an *HLRFI* value of 24.91%, has almost 33% worse haemodynamic performance when compared to Stent A (with an *HLRFI* value of 18.81%). Similarly Stent C-LC has 30% and 38% worse performance when compared with Stent C-SC and Stent A respectively.

4.5 Conclusions

Different points in the cardiac pulse produce different responses to the stent when measured by artery wall areas exposed to low WSS and reverse flow. Substantial differences in the flow features exist when both these factors are considered, simultaneously. Even for similar strut spacings, the design of the connector, especially its length in the cross-flow direction, significantly influences blood flow. Particularly for Stent C, it can be concluded that the haemodynamic alteration, measured by percentages of areas exposed to low and reverse WSS, is proportional to the length of the connector in the cross-flow direction. The relatively better performance of Stent A can be attributed to its connector's minimal cross-flow length and better alignment with the flow. Furthermore, the number of recirculation zones formed, and hence the oscillations in the MOSI values along any axial line on the arterial wall, is equal to the gaps between the stent struts and connectors. The differences in *HLRFI* values, which may be indicative of a stent's resistance to restenosis, reinforce the effect of stent design on alteration of haemodynamics. In essence, overall stent efficacy can be improved by improving the connector designs (in particular, their cross-flow length and alignment with flow) in the stent for minimal alteration of blood flow or as a tradeoff to improve other features such as drug distribution or flexibility.

To conclude, a CFD model to evaluate haemodynamic alteration is developed in this chapter. Furthermore, with a view of performing design optimisation studies in future chapters, an objective function, namely *HLRFI*, has been formulated. This measure of *HLRFI* combines the flow features in a stented artery that relevant to restenosis. In future chapters *HLRFI* is used as a figure of merit to compare different stents based on their haemodynamic efficacy.

Having evaluated the haemodynamic response, the next chapter deals with balloon-expansion of stents in order to quantify the mechanical stress-strain environment in a stented artery. One of the limitations of the models in this chapter is that the artery segments are assumed to be straight, cylindrical, and they contain no plaque. In reality, the stent is expanded in the diseased artery using a balloon. Hence, it is important that the geometry used for flow simulations, i.e. for the CFD models developed in this chapter, is the post balloon expansion geometry of the stented artery. The balloon expansion models developed in the next chapter serve this purpose.

Chapter 5

Balloon expansion of stents

This chapter deals with the numerical modelling of the expansion process of balloon-expandable stents inside a stenosed artery. Such modelling is essential for the following three reasons –

- to investigate the acute recoil after the stenting procedure,
- to investigate the stress-strain distribution in the stent and the arterial tissue, and
- to obtain the post-stenting geometry of the stent, plaque, and the artery, which can be used in haemodynamic models developed in chapter 4.

As discussed in section 2.5, an ideal stent should have high radial strength. This radial strength can be measured in term of the acute recoil after expansion of the stent. Furthermore, in order to minimise arterial injury due to the stenting procedure, an evaluation of the stress-strain environment in the stented segment of the artery is needed. Lastly, the post stenting geometry of the stented vessel is needed for accurate evaluation of haemodynamics as discussed in chapter 4. In this chapter finite element analysis (FEA) models are developed to meet all these objectives. In particular this chapter has the following aims

1. to develop a balloon-expansion model of a stent against the plaque in a representative model of a diseased artery,
2. to quantify structural strength of a stent in terms of acute recoil, and
3. to quantify the change in the mechanical stress-strain environment after the stenting procedure.

This chapter is organised as follows. In the first section, a review of a) studies which show a correlation between arterial injury and restenosis, and b) various approaches employed by researchers to model expansion of stents is presented. Thereafter, the expansion modelling approach adopted in this thesis is presented. Two sets of expansion are modelled. First, a free expansion of a stent, i.e. without the presence of plaque and artery, using a balloon is presented. This is primarily done to validate the obtained numerical results with manufacturer data (*in-vitro* testing without the resistance of a vessel). Then, the expansion model which includes a representative plaque and the artery is presented. Finally, the objective functions (figures of merit), representing radial strength of a stent and injury caused by the stenting

procedure, are formulated. These objective functions are used for optimisation studies in future chapters.

5.1 Introduction

Even though the primary aim of stenting is to achieve lumen gain in stenosed arteries, other factors such as the post-procedural injury and the stress-strain environment in the arterial tissue can not be neglected while designing a good stent. Several studies, Farb et. al [107], Schwartz et. al. [108], Carter et. al. [109] and König et. al. [110], have shown that neointimal formation is correlated to the vessel injury caused during the stenting procedure. Rogers and Edelman [81] showed that stent configuration was a key determinant of vessel injury caused due to stent implantation.

Vessel injury comprises endothelial exposure, intima laceration, and rupture of the medial layer [110]. Usually, it is quantified by the extent of strut penetrations into the arterial tissue. Schwartz et. al. and Carter et. al. in their respective studies reported correlations of $r = 0.84$ and $r = 0.72$ between mean neointimal hyperplasia and injury. Hoffman et. al. [111] performed a study with Palmaz-Schatz stents and showed that intimal hyperplasia moderately correlated with the aggressiveness of the balloon implantation technique. Kornowski et. al. [112] reported that neointimal formation after coronary stenting was dependent on both arterial injury and inflammatory reaction. Other studies, Kastrati et. al. [82], Pache et. al. [113], have concluded that stent geometry, specifically the strut thickness, is a key determinant of restenosis rates. A review article by Morton et. al. [114] showed the effect on restenosis of various geometrical stent parameters, such as type of stent (coil, tube, slotted, etc.), length, percentage metal coverage, number of struts, strut thickness, cross-section, surface finish, symmetry, and material. This article, supported by a series of randomised trials, suggests that stent geometry plays an important role in determining a stent's resistance to restenosis.

In engineering terms, injury can be translated to the post-procedural stress-strain environment in the arterial tissue. In order to pass an engineering judgement on the performance of a group of stents, we must be able to evaluate the stresses, quantify them, and develop a figure of merit (objective function) based on the evaluated stresses. Finite element analysis provides an efficient technique for such evaluations, and has been used in the past for numerous numerical studies on stents.

A number of different techniques, with different levels of complexity, have been used in the past to model stent expansion. Such strategies can be divided into the following classes –

- Disregard of balloon
  - application of uniform radial pressure to the inner surface of the stent.
  - application of uniform displacements to the inner surface of the stent.
- Using a cylindrical balloon
  - application of uniform radial pressure to the inner surface of the cylinder.
  - application of uniform displacements to the inner surface of the cylinder.
- Using a folded balloon
  - uniform radial pressure inside a tri-folded balloon.

– application of a negative pressure to a fully inflated balloon to get deflated shape (multi-folded), and then inflating the balloon by application of uniform radial pressure.

Before 2000, most studies on determining mechanical properties of stents were based on self-expanding stents, Perry et. al. [115], Auricchio et. al. [116], and Trouchu et. al. [117, 118]. Rogers et. al. [119] investigated the effect of balloon-artery interactions by means of *in vivo* studies and 2-D finite element models. They demonstrated that inflation pressure, strut spacing, and balloon compliance, all affect the surface contact stresses on the artery wall.

Dumoulin et. al. [120] were the first to perform a 3D analysis of stent expansion. They investigated the free expansion (without including the plaque and the artery) of a 3-D model of a Palmaz P308 (Johnson & Johnson company) stent. They evaluated mechanical characteristics such as recoil, shortening, resistance to crushing, buckling, and fatigue. Their *in vitro* observations of an expanding stent led them to conclude that the stent expanded uniformly everywhere except at the ends. Hence, they disregarded the balloon by assuming an infinite length prosthesis, and drove the finite element analysis by applying a uniform radial pressure to the inner surface of the stent. Auricchio et. al. [121] did a similar study with a Palmaz-Schatz (Johnson & Johnson) like stent for both free expansion and by including the plaque and the artery in the analysis. Apart from reporting several mechanical parameters such as elastic recoil, foreshortening, metal-artery ratio, and residual stenosis, they suggested a modification in the stent geometry to facilitate uniform expansion of the stent. This study too neglected the presence of the balloon. Etave et. al. [122] used FEA to study two different types of stents – slotted tube and coil stents. They drove the analysis by specifying radial displacements to the stent nodes. Apart from reporting the mechanical properties like other aforementioned studies, they also evaluated flexibility. Migliavacca et. al. [123] studied the effect of stent geometry (thickness, metal-to-artery ratio, etc.) on its mechanical performance. Despite neglecting the balloon, their analysis, driven by uniform internal pressure, went a step further by analyzing next generation stents (Multi-Link Tetra and Carbostent) besides the Palmaz-Schatz like stents. Gu et. al. [124] performed a uniform pressure driven FEA on a covered microstent (a stent with thin coatings) to evaluate the effect of coatings on the mechanical performance and compared them with bare metal stents. Holzapfel et. al. [125] proposed a parameterisation technique to enable generation of new stent designs, and performed displacement driven FEA to analyze these designs. While most studies mentioned so far had assumed the artery to be a straight cylindrical segment, this study used a full 3-D geometrical model of a post-mortem specimen of a stenosed human iliac artery. Moreover, they considered eight different types of tissue components for the artery with different material properties. This study evaluated the effect of stent geometry on the total mechanical environment post stenting, and quantified this stress-strain environment by formulating scalar indicators which can be used to compare stents. Migliavacca et. al. [126] used FEA to study the mechanical properties of the BX Velocity (Cordis corp., Johnson & Johnson) stent, and compared the pressure-diameter (p-d) relationship during expansion to those provided by the manufacturer. The comparison of the two p-d relationships led to the conclusion that the absence of a balloon in numerical modelling was the cause of the observed discrepancies between numerical and experimental data.

Holzapfel et. al. [127] introduced the balloon in their analysis for both angioplasty and

stenting. They modelled the balloon as a rigid cylinder, and the analysis was driven by specifying displacements. Their stenosed artery geometry was obtained by MRI scans of a human external iliac artery and the constitutive models for each part of the tissue obtained by mapping MRI data with histological tissue type. Liang et. al. [128] modelled the balloon as a cylindrical hyperelastic shell to mimic the unfolding process of a folded balloon. Marrey et. al. [129] modelled the balloon as a semi-rigid shell, although what the authors meant by semi-rigid is not detailed. In their analysis the stent is first crimped to simulate the stent-catheter assembly. This is done by putting a semi-rigid crimp cylinder outside the stent and specifying the radial contraction of the crimp cylinder. Contact modelling between the crimp cylinder and stent's outer surface ensures cripming of the stent. Stent expansion into the artery is modelled in a similar way. After the expansion step, the authors apply a physiologically realistic cyclic pressure loading to the inner surface of the artery to simulate systolic and diastolic blood pressure. The response of the stent is then used to investigate fatigue resistance of the stents. Wang et. al. [130] studied the transient expansion of coronary stents by modelling the balloon as an incompressible isotropic linear elastic material. They evaluated the effect of varying balloon lengths on different stent geometries. Takashima et. al. [131] used both experimental and numerical techniques to study the contact area between the stent and the artery/plaque wall. They modelled the balloon as a linear elastic material with a very high Young's modulus, and drove the analysis by specifying radial displacements to the balloon. Wu et. al. [132] studied the expansion of stents in both straight and curved vessels and compared the results. In their displacement driven analysis they modelled the balloon as a rigid body, while the artery and tissue were modelled as hyperelastic materials. Gijsen et. al. [133] studied the deployment of a stent in a realistic coronary artery. They constructed the geometry of a right coronary artery of a 57-year old male using biplane angiography and intravascular ultrasound (IVUS).

All the aforementioned studies have either neglected the balloon or treated it as a rigid cylindrical shell. It was De Beule et. al. [134] who, in 2008, modelled a realistic balloon expansion of the CYPHER stent (Cordis corporation, Johnson & Johnson). They modelled the balloon as a linear elastic material, whose properties were extracted from the manufacturer's compliance chart, and took into account the folded shape of the balloon. They validated their results, specifically the *p-d* relationship with those provided by the manufacturer, and their results showed a maximum error of 4.1%. In the same year, Gervaso et. al. [19] published a study outlining different strategies for modelling balloon expansion of stents. They also proposed a new technique to take into account the folded shape of the balloon. They started with a fully expanded shape of balloon, with its nominal diameter, and applied a negative pressure to deflate it. This resulted in a shape that could then fit inside the crimped stent for expansion. The balloon was modelled as a linear elastic material, and the analysis was driven by increasing the inflation pressure of the balloon. Their numerical results for the Bx VELOCITY stent, which is similar to the CYPHER stent used by De Beule et. al. except for a thin drug coating, corresponded well with those provided by the manufacturer. Zunino et. al. [20] extended this technique to take the expanded geometry of a unit model of the Bx VELOCITY stent and performed flow and drug diffusion simulations on the deformed shape. Although the application of such expansion techniques to realistic artery models, such as those used by Holzapfel et. al. [127, 125], remains to be seen, to the best of the author's knowledge, these (De Beule et. al. [134] and Gervaso et. al. [19]) are the only two realistic

models for studying expansion of stents.

As mentioned earlier, in order to perform a design optimization study on stents, it is necessary to model the expansion of stents for three reasons – first, to evaluate recoil, second, to evaluate the mechanical stress-strain environment in the artery, and third, to obtain the expanded geometry for evaluation of flow and drug distribution. In the following sections the methodology adopted in this thesis for these purposes is presented.

5.2 Methodology

Of the various balloon expansion methodologies listed in section 5.1 the one proposed by Gervaso et. al. [19] is adopted in this study. While Gersavo et. al. presented the results only for free expansion, results for both free expansion and expansion into a representative plaque and artery model are presented in this chapter. The following sub-sections detail the steps of the analysis procedure.

5.2.1 Geometry

5.2.1.1 Stent

The CYPHER stent is chosen for developing the expansion methodology. This is because the CYPHER stent –

- was a widely used stent in clinical practice until its discontinuation in 2011 [35].
- is representative of most contemporary stent designs – circumferential rings connected by flex connector segments.
- has been studied in the past in various FEA studies.
- has easily available pressure-diameter relationship chart provided by the manufacturer after *in vitro* testing.

Figure 5.1 shows an electron microscope picture [34] of the Bx VELOCITY stent. Two important features of the CYPHER/Bx VELOCITY stent geometry can be easily located. First, it contains 4 circumferential rings placed in the longitudinal direction, and second, the circumferential rings are connected by 'n' shaped connectors/links [34, 135].

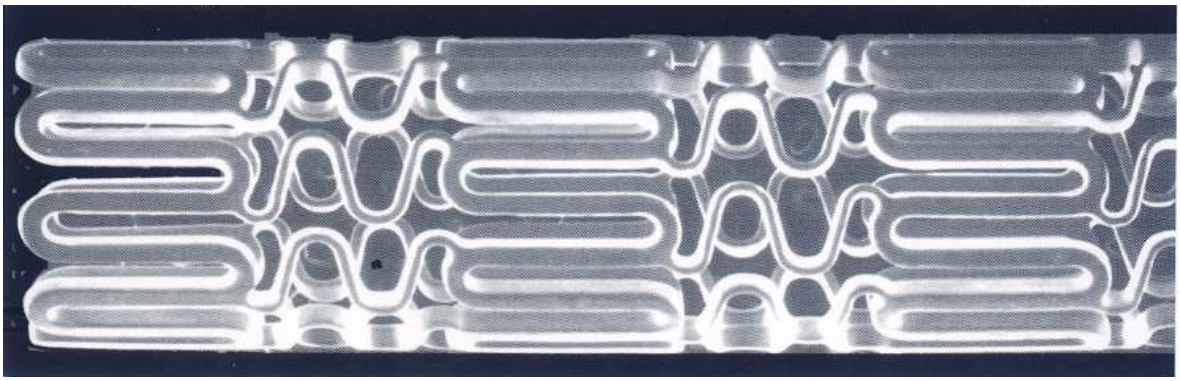


Figure 5.1: Electron microscope image of Bx VELOCITY stent

The important characteristics of the stent geometry used in this chapter are obtained from the 'instructions for use' manual provided by the manufacturer [135] and 'the handbook of coronary stents' [34]. Table 5.1 lists these properties

Table 5.1: Geometric information for the CYPHER stent

Dimension	**Value**	**Note**
Target artery diameter	3.00 mm	a range of diameters are available; 3.0 mm is chosen for this study
Length	8.00 mm	a range of lengths are available; 8.0 mm is chosen for this study
No. of circumferential cells	6	for 2.25-3.00 mm diameter stents, the number of rings is 6
Strut thickness (radial)	0.14 mm	converted from inches to mm (0.0055 inch)
Strut width (circumferential)	0.13 mm	converted from inches to mm (0.0052 inch)

With the help of these geometric parameters and visual insepction of the crimped state geometry, a representative model of the CYPHER stent is generated in Rhinoceros 4.0. The semi-crimped state internal diameter is chosen to be 1.22 mm. An image of the semi-crimped state CAD model is shown in Figure 5.2.

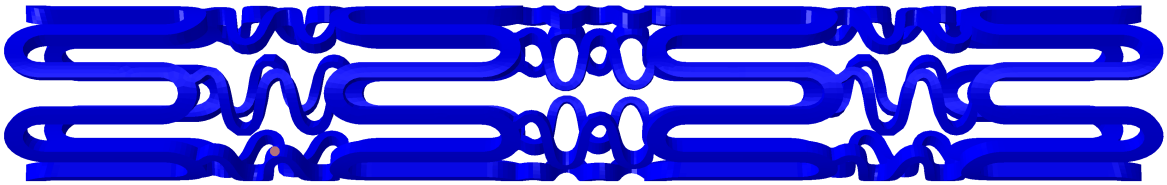


Figure 5.2: CAD model of the CYPHER stent

5.2.1.2 Balloon

To construct the geometry for the balloon two important dimensions are needed – the nominal diameter and the length. The nominal balloon diameter is chosen to be 2.85 mm, which was obtained by De Beule et. al. [134] after extrapolating the balloon's compliance chart (provided by the manufacturer) to a zero stress state. The nominal length of the balloon is specified to be 10.0 mm according to the data in [135] which specifies that the stent delivery balloon's length is nominally 2.0 mm longer than the stent being used. Based on this data the nominal geometry of the balloon is created in Rhinoceros 4.0, see Figure 5.3.

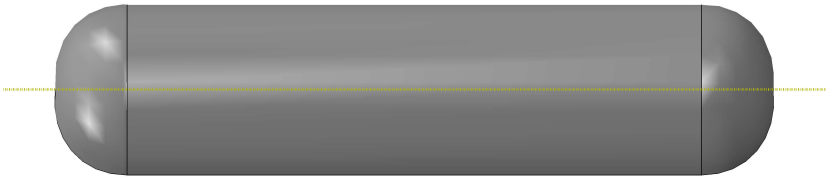


Figure 5.3: CAD model of the nominal balloon

5.2.1.3 Artery and Plaque

The artery is assumed to be a single layer straight cylindrical tube with an internal diameter of 3.00 mm, an outer diameter of 4.40 mm (thickness of 0.70 mm), and a length of 15.00 mm. As far as the geometry of the plaque/stenosis is concerned no statistical data could be found in the literature. Most studies have assumed it to be cylindrical with round ends. The stenosis shape, however, can have a significant effect on the performance of the stent, and a stent that performs well for one particular stenosis shape might not perform as well with a different stenosis shape. A different methodology to represent stenosis shapes which can easily be altered by changing a few parameters is proposed. Hicks-Henne bump functions [136] are a class of functions which allow easy manipulation of a shape representing a smooth bump by altering three parameters. These functions, commonly used in aircraft wing design and used by Kolachalama et. al. [137] to represent a carotid artery stenosis, seem to appropriately represent stenosis shapes. A Hicks-Henne bump function is represented by the following formula –

$$y = A[\sin(\pi x^{(-\ln 2/\ln x_p)})]^t \ ; \text{for } 0 < x < 1, \tag{5.1}$$

where A is the height of the bump (analogous to the peak stenosis), x_p is the location of the peak, and t controls the width/sharpness of the peak. A larger value of t gives sharper peaks representing a focal stenosis, while smaller values of t produce a more diffused stenosis shape. To include any length of stenosis and add a base thickness to the plaque, the function is modified as follows –

$$y = t_b + (A - t_b)[\sin \pi \left(\frac{x}{L}\right)^{(-\ln 2/\ln x_p)}]^t \ ; \text{for } 0 < x < L, \tag{5.2}$$

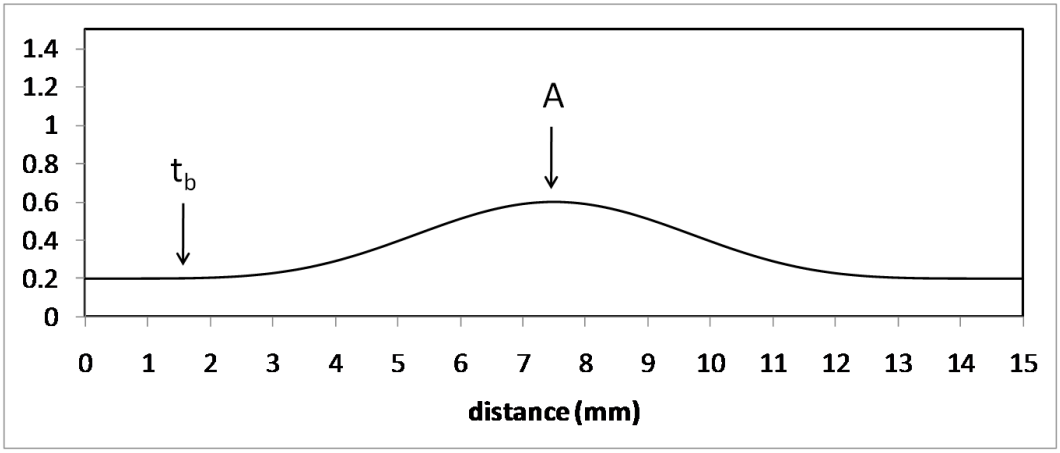
where t_b is the base thickness and L is the length of the stenosis. Figure 5.4 shows the Hicks-Henne bump function and the corresponding stenosis shape for $A = 0.6$, $x_p = 0.5$, $t = 5$, $t_b = 0.2$, and $L = 15$. Figure 5.5 shows the assembly of the stent, plaque and the artery.

5.2.2 Materials

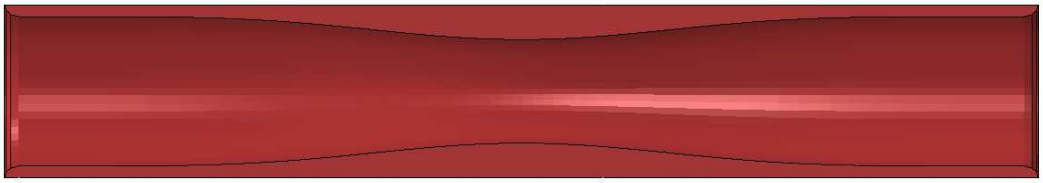
The following subsections describe the material properties used for each of the stent, plaque, and the artery.

5.2.2.1 Stent

Most of the bare metal stents, including the CYPHER stent [135] are made up of medical grade 316L Stainless Steel. The constitutive behaviour of a stent material depends on a number of pre-treatments such as hot rolling, annealing, cold finishing, electropolishing etc. [134, 138]. In the past a number of different material properties have been used for finite element analysis of 316L Stainless Steel stent expansion processes. Table 5.2 presents an extension to the list presented in De Beule's PhD thesis [11] where the different material properties used in different studies are listed. Murphy et. al. [10] performed uniaxial tensile tests on 316L stainless steel strut specimens and presented the stress-strain curves for different size specimens. Figure 5.6 shows the engineering stress strain curves from their experiments. These stress-strain curves, which have also been used by De Beule et. al. [134], are deemed most appropriate for numerical studies in this chapter as these have been determined experimentally for the 316L



(a) Hicks-henne bump function for $A = 0.6$, $x_p = 0.5$, $t = 5$, $t_b = 0.2$, and $L = 15$



(b) Stenosis shape by revolving the hicks-henne bump function

Figure 5.4: Stenosis shape

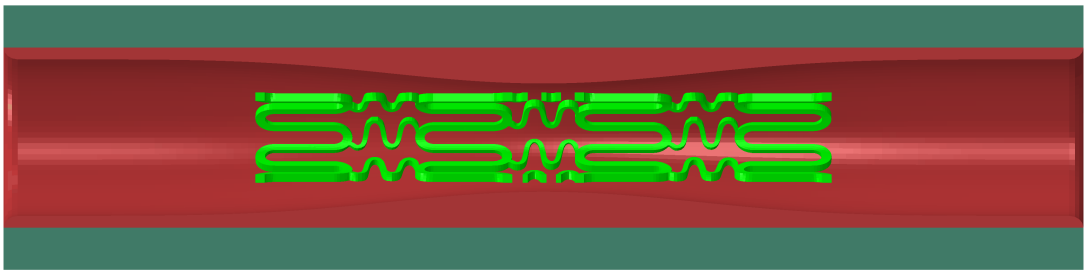


Figure 5.5: Assembly of the stent, plaque and artery (No balloon present)

stainless steel strut specimens of sizes similar to those used in this study. A Young's modulus (E) of 196,000 N/mm^2, a yield stress (σ_y) of 375 N/mm^2, and an isotropic piece-wise linear hardening is used to mimic the plastic region. It must be noted that the geometric model of the stent used in this study is a semi-crimped model, i.e. the crimping process of the stent is ignored. Consequently, if kinematic hardening is taken into account, then it is likely that due to the crimping of the stent, the yield stress for the expansion process will be lower than 375 N/mm^2 owing to the well known Bauschinger effect. This effect, however, is ignored in this chapter.

5.2.2.2 Balloon

The semi-compliant balloon is made of single-layer nylon [135]. De Beule [11] derived the constitutive properties of the semi-compliant Raptor^{TM} balloon, which is used in the expansion

Table 5.2: Different material properties used for FEA of 316L Stainless Steel stents

Group	**Young's modulus (E) (N/mm^2)**	**Yield stress (σ_y) (N/mm^2)**	**Material reference; notes**
Auricchio [121]	196000	205	Metals handbook (American Society of Metals)
Etave [122]	196000	-	Standard stress-strain curve for annealed material
Migliavacaa [123]	196000	205	Auricchio [121]
McGarry [138]	209000	264	Uniaxial tensile test
Gu [124]	196000	205	Auricchio [121]
Migliavacaa [126]	196000	105	Auricchio [121]; included kinematic hardening in the plastic phase
Liang [128]	201000	330	Tensile test on wires
Hall [139]	196000	290	-
Wang [130]	201000	-	Tensile test on annealed wire
De Beule [140]	196000	205	Auricchio [121]
Wu [132]	201000	280	Standard true stress-strain curve for annealed material
Takashima [131]	200000	315	-
Gijsen [133]	196000	105	Petrini [141], Migliavacca [126]
De Beule [134]	196000	375	Murphy [10] (Uniaxial tensile loading of SS 316L strut specimens)
Gervaso [19]	193000	205	-
Zunino [20]	193000	205	Auricchio [121]; included degradation of the hardening modulus

of the CYPHER stent [135], using the compliance chart provided by the manufacturer and using thin shell membrane theory. Their derived values have been adopted for the balloon material properties in this study. The balloon is assumed to be a linear elastic material with a Young's modulus (E) of 920 N/mm^2, and a Poisson's ratio (ν) of 0.4.

5.2.2.3 Artery

The artery material is described as an isotropic hyperelastic material based on a sixth order reduced polynomial strain energy density function, U, given by

$$U = C_{10}(\bar{I_1} - 3) + C_{20}(\bar{I_1} - 3)^2 + C_{30}(\bar{I_1} - 3)^3 + C_{40}(\bar{I_1} - 3)^4 + C_{50}(\bar{I_1} - 3)^5 + C_{60}(\bar{I_1} - 3)^6, \tag{5.3}$$

where, $\bar{I_1}$ is the first invariant of the deviatoric Cauchy-Green tensor –

$$\bar{I_1} = \bar{\lambda_1}^2 + \bar{\lambda_2}^2 + \bar{\lambda_3}^2, \tag{5.4}$$

$$\bar{\lambda_i} = J^{-1/3}\lambda_i, \tag{5.5}$$

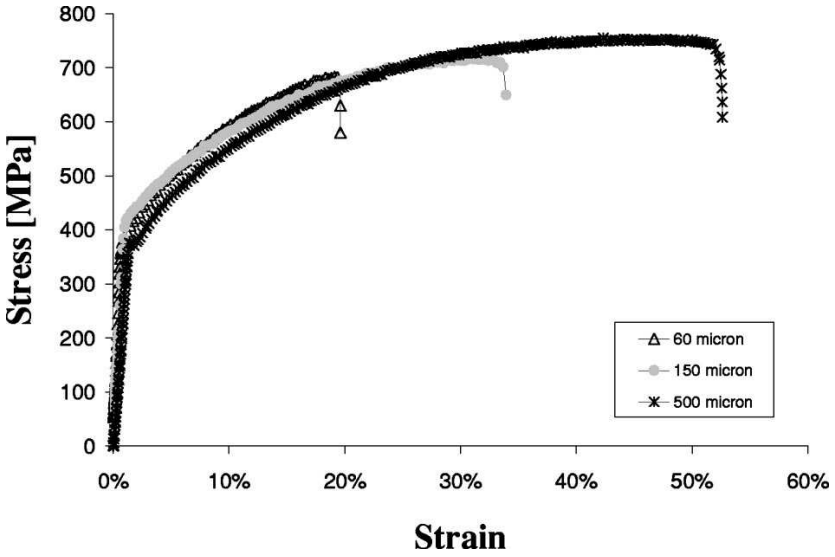


Figure 5.6: Stress-strain curves for coronary stent strut specimens [10]

where λ_i are the principal stretches and J is the total volume ratio (J=det($\mathbb{F}$), where $\mathbb{F}$ is the deformation gradient.

Holzapfel et. al. [142] studied the deformation behavior of 13 non-stenotic human LADs, and proposed a constitutive model for each of the three layers – intima, media, and the adventitia. Their results were adopted for a reduced polynomial strain energy density function (equation 5.3) by Gervaso et. al. [19], and Zunino et. al. [20]. Of these values, which are outlined in Table 5.3, the coefficients for the media layer are used in this study as the artery is assumed to be a single layer.

Table 5.3: Coefficients for strain energy density function for artery layers [19, 20]

Layer	C_{10}	C_{20}	C_{30}	C_{40}	C_{50}	C_{60}
Intima	6.7×10^{-03}	0.54	-1.11	10.65	-7.27	1.63
Media	6.52×10^{-03}	4.89×10^{-02}	9.26×10^{-03}	0.76	-0.43	8.69×10^{-02}
Adventitia	8.27×10^{-03}	1.20×10^{-02}	0.52	-5.63	21.44	0.00

5.2.2.4 Plaque

There is little data available on the material properties of plaque. Holzapfel et. al. [143] sampled 107 specimens from nine human high-grade stenotic iliac arteries, identified eight different tissue types, and quantified the anisotropic mechanical response of these different tissue types using histology and magnetic resonance imaging. The application of such material properties for finite element studies requires realistic geometries and different tissue type volumes identified by MRI scans. A simpler model for plaque was used by Wong et. al. [144]. They used a hyperelastic, Neo-Hookean strain energy function to describe the mechanical behavior of plaque. A Neo-Hookean strain energy density function is defined as follows –

$$U = \frac{\mu}{2}(\bar{I_1} - 3) + \frac{K}{2}(J-1)^2, \tag{5.6}$$

where μ is the initial shear modulus, K is the initial bulk modulus, and $\bar{I_1}$ and J are as defined in Eqns. 5.4 and 5.5. This Neo-Hookean description of the plaque has also been used by Kock et. al. [145], and is used in this study as well. The initial shear modulus, μ, can range from

0.1 kPa to 60 kPa [144]. Smaller values of μ indicate softer plaque while larger values tend towards calcified plaque. In this chapter, μ has been assumed to be 60 kPa. To model a nearly incompressible plaque, which most biological tissues are [144, 146, 145, 147], a value of $K = 20\mu$ is chosen, which gives a poisson's ratio, ν, of 0.475.

Other hyperelastic models such as the Mooney-Rivlin form have been used in studies by Pericevic et. al. [146] and Lally et. al. [148].

5.2.3 Element type

An important consideration in FEA is the choice of element type. Hall et. al. [139] presented an interesting study which compared the use of following element types for stent to model the expansion process –

- C3D8 – 8-node hexahedral solid element.
- C3D8R – 8-node linear hexahedral solid element with reduced integration.
- C3D8I – 8-node linear hexahedral incompatible mode solid element.
- S4 – 4-node linear quadrilateral shell element with 5 integration points through the thickness.
- S4R – 4-node linear quadrilateral shell element with reduced integration and 5 integration points through the thickness.
- B31– 2-node Timoshenko beam element with 5×5 cross section integration points.

Their study showed that all the element formulations provide similar responses even though the dimensionality of all the elements are different. Since evaluation of the post-stenting haemodynamic response is one of the objectives of this thesis, C3D8R elements were chosen for the stent, plaque, and the artery. These elements were also used in earlier studies, e.g. De Beule et. al. [134] and Gervaso et. al. [19]. For the balloon, M3D4R elements (4-node quadrilateral elements with reduced integration) are chosen [134, 19] with a thickness of 0.02 mm [134].

5.3 Simulations

All the simulations are carried out using Abaqus/Explicit 6.9.1 (Dassault Systèmes Simulia Corp., Providence, RI, USA), a commercially available finite element solver. The simulations are carried out in three sets –

1. Deflation of the balloon. The methodology of Gervaso et. al. [19] is adopted to obtain the unexpanded shape of the balloon.
2. Free expansion of stent – without the resistance of plaque and artery.
3. Expansion into the plaque and the artery.

The following sub-sections detail the boundary conditions, loading process, and the mesh and time-step dependence studies, for the three aforementioned steps.

5.3.1 Deflation of balloon

For the deflation of the balloon, the nominal shape of the balloon is taken and the following boundary conditions are applied –

1. Two nodes at the extreme ends of the balloon are constrained to allow displacements only in the axial direction.

2. Three nodes in the central axial cross section, forming an equilateral triangle, are constrained to move only in the radial direction.

Figure 5.7 shows the boundary conditions imposed on the balloon. A negative pressure of 0.01 MPa is then applied to the inner surface of the balloon. This results in deflation of the balloon which can then be used for expanding the stent. The crimping process is neglected in this study and the deflated balloon is used to expand the semi-crimped geometry created directly by CAD modelling. Figure 5.8a shows the deformed balloon, and Figure 5.8b shows the balloon inserted inside the semi-crimped stent.

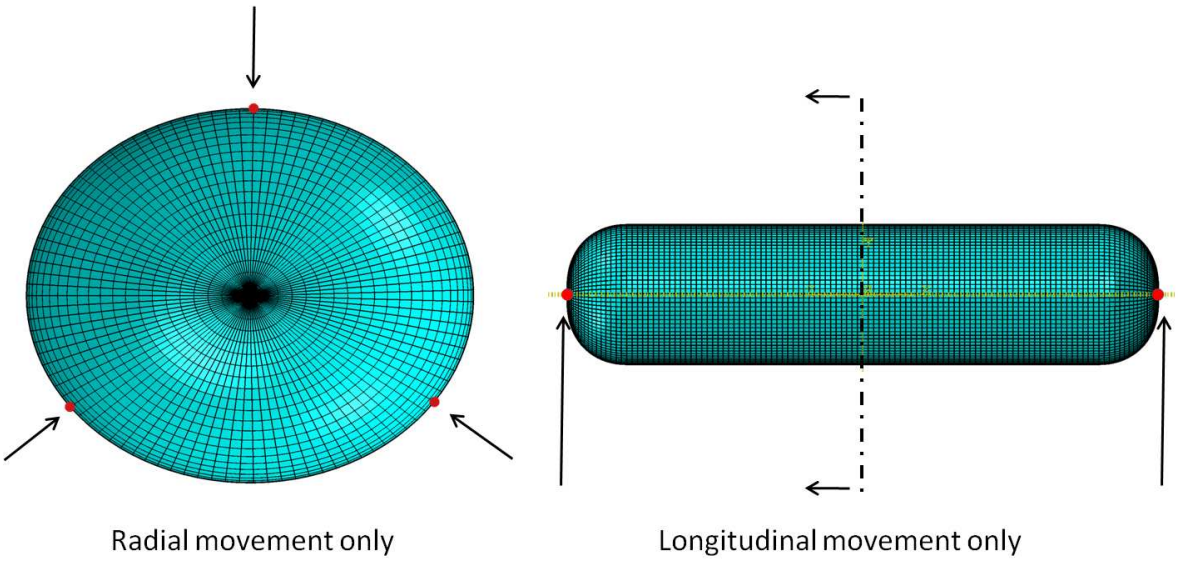
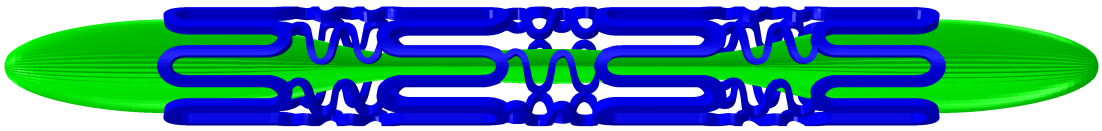


Figure 5.7: Boundary conditions imposed on the balloon for deflation analysis

5.3.2 Free expansion of stent

It is worth recalling that the simulations for free expansions are performed in order to validate the results with those provided by the manufacturer. For these simulations, the balloon configuration from the results of the above step is positioned midway into the stent (cf. figure 5.8b). For the balloon, the same boundary conditions as specified in section 5.3.1 are imposed. For the stent, in-order to constrain any potential rigid displacements [19], three nodes along the central axial cross section of the stent are constrained to move only in the radial direction. A pressure, P, is then applied to the inner surface of the balloon. In clinical practice inertia has negligible effect on the expansion process [134]. Consequently the balloon expansion of stents is usually modelled as a quasi-static process. A quasi-static process is characterized by the ratio of kinetic energy of the deforming material to the total internal energy, such that this ratio remains less than 5%, throughout the analysis [149]. Consequently, the loading rate under the assumption of a quasi-static analysis can be increased considerably, provided the

(a) Deflated balloon configuration



(b) Deflated balloon inside stent

Figure 5.8: Deflated shape of the balloon

energy ratio requirements are met. Figure 5.9 shows the variation of pressure P with clearly defined loading and unloading phases. In reality, there is a 'hold' phase, between the loading and unloading phases, where the pressure is maintained at its peak value for some time. This 'hold phase' is not included in the free expansion primarily because the manufacturer data for validation are available only for the loading phase. A maximum pressure of 1.6 MPa is chosen because the rated burst pressure for the Raptor^{TM} balloon, used for expanding the CYPHER stent, is 1.621 MPa [135]. To model the balloon self contact and the contact of balloon with stent, a surface-to-surface discretisation method is used with a penalty based constraint enforcement and finite-sliding formulation [149]. Finite sliding formulation takes non-linear geometric effects into account and, as opposed to a small-sliding formulation, tracks the relative motions of the surfaces in contact [149]. A Coulomb friction model, with a friction coefficient of 0.2, is used for both the contact formulations, which is valid for both nylon-nylon and nylon-steel contact [134, 150] . Figure 5.10 shows the meshed assembly of balloon and stent for the free expansion simulations.

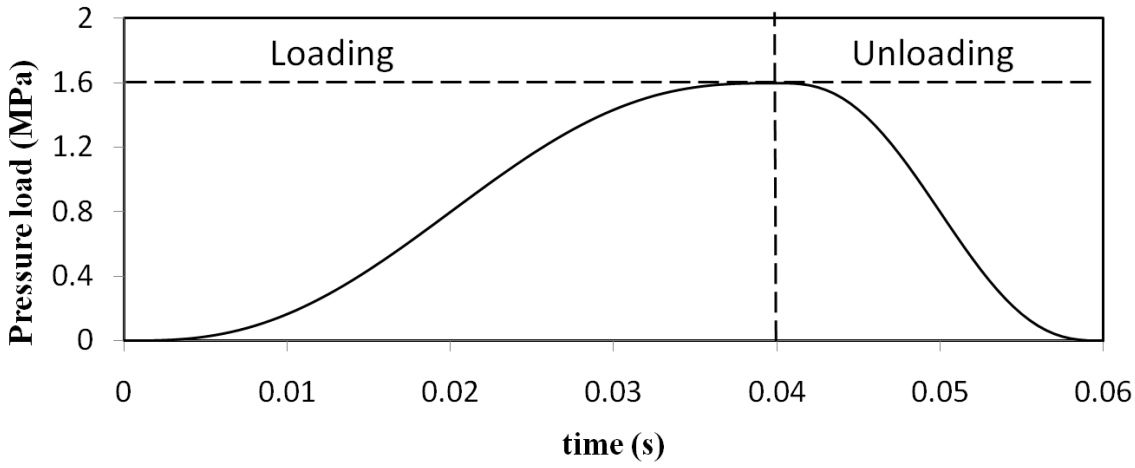


Figure 5.9: Loading profile for the free expansion of balloon

Mesh and time-step dependence studies are performed to be sure that the final simulations are run for a mesh and time-step such that an increase in the mesh density and/or a decrease

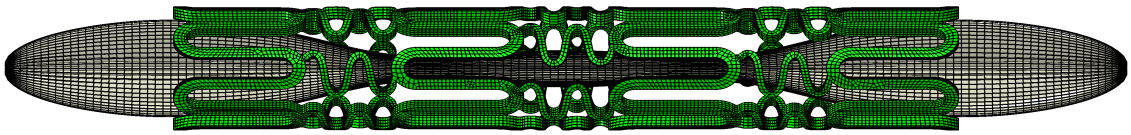


Figure 5.10: Stent and balloon assembly (meshed) for free expansion

in the time-step would have negligible effect on the results. Table 5.4 shows the combinations of different mesh and time-steps used for the quasi-static explicit analysis. Such a study is also helpful in determining the mesh size and time-step for future optimization studies where a number of designs are analyzed while parsimoniously using computational resources. To compare the effect of different mesh densities and time-steps the transient pressure-diameter relationship is used. Since the stent does not expand uniformly along its length, owing to the balloon expanding faster at the ends, an average diameter of stent is used. This average diameter, d, is determined by taking eight nodes, distributed uniformly along the axial direction of the stent, on the inner surface of the stent. Figure 5.11 shows the location of these eight nodes used for calculating d.

Table 5.4: Different mesh and time-steps used for verification studies

Name	**No. of elements Balloon**	**No. of elements Stent**	**time-step (s)**
Base_1e7	13068	19032	10^{-7}
Base_1e8	13068	19032	10^{-8}
Mesh1_1e7	16800	33484	10^{-7}
Mesh1_5e8	16800	33484	5×10^{-8}
Mesh1_1e8	16800	33484	10^{-8}
Mesh1_1e9	16800	33484	5×10^{-9}
Mesh2_1e7	29800	42216	10^{-7}
Mesh2_1e8	29800	42216	10^{-8}
Mesh3_1e7	42960	64740	10^{-7}
Mesh3_1e8	42960	64670	10^{-8}

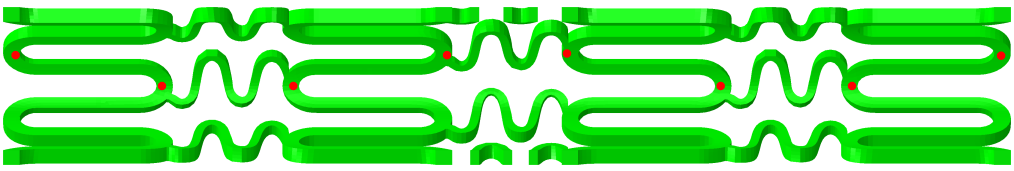


Figure 5.11: Eight nodes to measure the average diameter of stent during transient expansion

Figure 5.12 shows the pressure-diameter, $p-d$, relationship for four different time steps. As can be seen in this figure, decreasing the time-step below 10^{-8} s has negligible effect on the profile. This implies that a time-step of 10^{-8} seconds is computationally most appropriate to accurately capture the transient behavior.

Figure 5.13 shows the $p-d$ relationship for all the four meshes using a time-step of 10^{-8} seconds. This figure suggests that using a mesh density equivalent to Mesh1 for CYPHER like stents would yield negligible numerical error if the mesh density is increased. Hence all the results presented in sections 5.4 and 5.5 are based on Mesh1 using a time-step of 10^{-8} seconds.

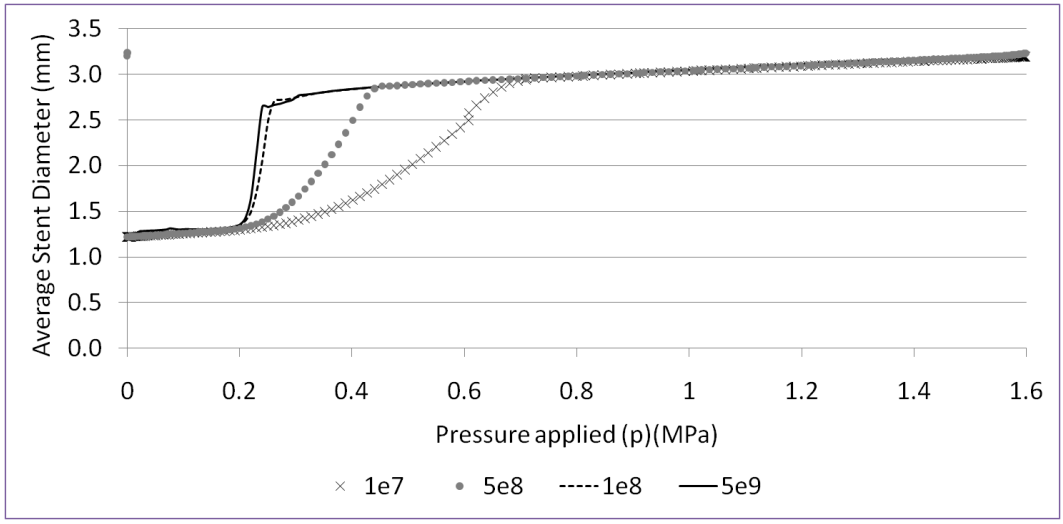


Figure 5.12: $p-d$ relationship for different time-steps on Mesh1

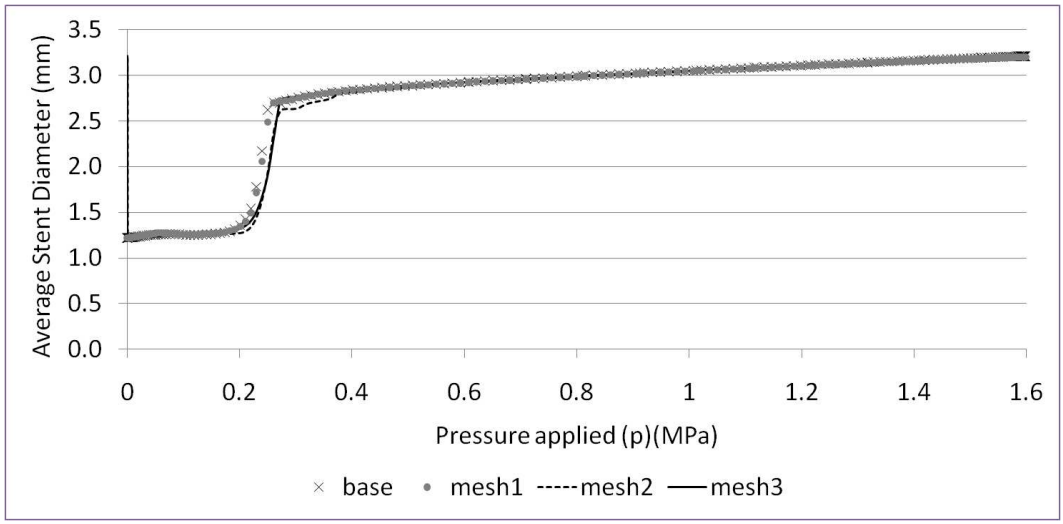
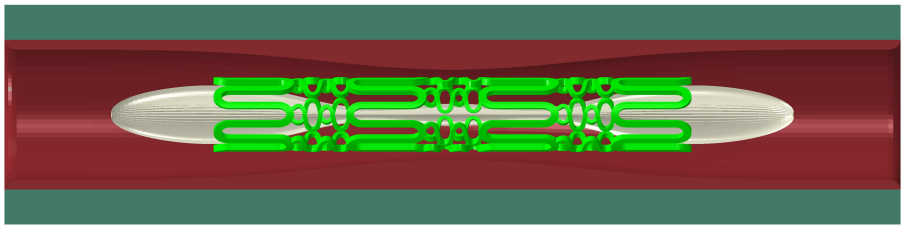


Figure 5.13: $p-d$ relationship for different meshes using a time-step of 10^{-8} s

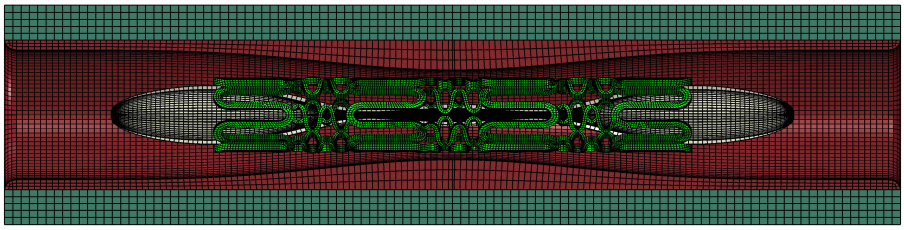
5.3.3 Expansion into the plaque and artery

Having setup the simulations for the free-expansion of the stent the plaque and the artery are now included in the expansion simulations. Figures 5.14a and 5.14b respectively show the geometry and the mesh for the assembly of the balloon, stent, plaque, and the artery. The boundary conditions for the balloon and stent are kept the same as in the free expansion

analysis (section 5.3.2). A tie-constraint is specified between the outer surface of the plaque and the inner surface of the artery. An encastre boundary condition ($U_i = 0;\ i = 1, 2, 3$) is imposed on the longitudinal ends of the artery. Moreover, three nodes on the central cross section of the artery are constrained to move only in the radial direction [19, 20]. Two additional contacts, apart from those in section 5.3.2, are specified – contact between balloon and plaque, and contact between stent and plaque. Both these contacts are specified using a surface-to-surface discretisation and penalty method for contact constraint enforcement. A finite-sliding-formulation for the contact pairs with a Coulomb friction model is used [149]. Since there is no data available (to the best of the author's knowledge) regarding friction coefficient values between either balloon or stent and the arterial tissue, a friction coefficient of 0.2, the same as that between nylon-nylon and nylon-steel is used.



(a) Geometry



(b) Mesh

Figure 5.14: Assembly of balloon, stent, plaque, and artery

To drive the analysis a uniform pressure, p , is applied to the inner surface of the balloon to simulate expansion. Figure 5.15 shows the variation of p with time. Three distinct phases of loading, hold, and unloading can be identified. A maximum pressure of 1.2 MPa is applied as the nominal pressure for the CYPHER stent is 1.115 MPa [135] (see table 5.6). The expansion is modelled as a quasi-static process, keeping the kinetic energy less than 5% of total internal energy. The mesh densities for the stent and the balloon, and the time-step to model the transient expansion, are justified in section 5.3.2 – in particular, Mesh1 and a time-step of 10^{-8} seconds are chosen. To determine the mesh densities for plaque and the artery appropriate mesh dependence studies are conducted. Three different mesh sizes are used for plaque and artery. Table 5.5 shows the different plaque and artery mesh densities and time-steps used for this study. Even though a time-step of 10^{-8} seconds is deemed most appropriate, the tests are conducted for two sets of time steps viz. 10^{-8} and 10^{-7} seconds. The reasons for this will become clear in section 5.3.4.

In order to compare the effect of different mesh densities, the von Mises stresses on the plaque inner surface along a central line at the end of the analysis are used. This final state reflects the equilibrium stage after the balloon pressure has been unloaded. Figure 5.16 shows this line in the undeformed state of the plaque. Figure 5.17 shows a comparison of the von-Mises stresses on this line for different meshes. The average difference between the

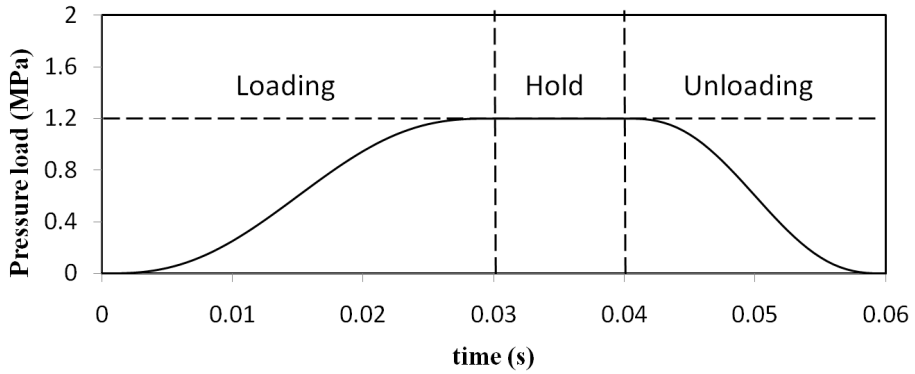


Figure 5.15: Loading profile for the expansion of stent into the plaque and artery

Table 5.5: Different mesh and time-steps used for verification studies of stent expansion into the plaque and artery

Name	**No. of elements Plaque**	**No. of elements Artery**	**time-step (s)**
PA_Base_1e7	11948	19000	10^{-7}
PA_Base_1e8	11948	19000	10^{-8}
PA_Mesh1_1e7	25690	27072	10^{-7}
PA_Mesh1_1e8	25690	27072	10^{-8}
PA_Mesh2_1e7	40216	43200	10^{-7}
PA_Mesh2_1e8	40216	43200	10^{-8}

stress results for PA_Mesh1 and PA_Mesh2 is less than 5%. Even though both PA_Mesh1 and PA_Mesh2 are appropriate for stress calculations, PA_Mesh2 is chosen for future optimization studies, as it gives a better resolution of plaque geometry for subsequent CFD studies (presented in chapter 8).

5.3.4 Choice of time-step for optimisation studies

Since design optimization is a process which needs evaluation of multiple points in a given design space, multiple computational simulations are often not affordable for complex analyses. The expansion analysis presented in Section 5.3.2 suggests the use of a time-step of 10^{-8} seconds, in order to capture the numerical transient expansion which is dependent of the time-step chosen. However, a close look at Figure 5.12 shows that the expansion behavior for a time-step of 10^{-8} s is similar to that for a time-step of 10^{-7} s for pressures higher than 0.6 MPa with minimal errors. Based on this observation it is hypothesized that a time-step of 10^{-7} seconds can be used for optimization studies, as it reduces the physical simulation time from about over a week to 24 hrs. To prove this hypothesis the following four arguments are considered –

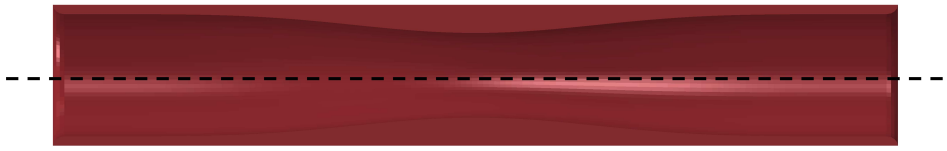


Figure 5.16: Line on the plaque inner surface on which the Von Mises stresses are compared

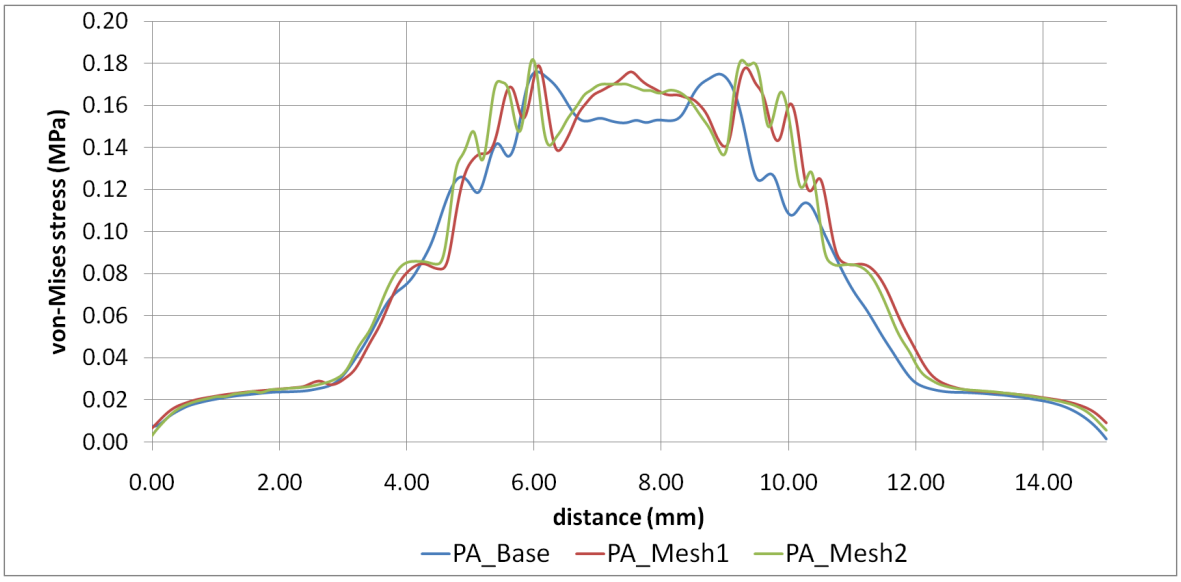


Figure 5.17: Von-mises stress comparison along a line on the inner surface of the plaque for different meshes

1. The material properties for all the components of the expansion analysis, do not have a time-dependent component.

2. None of the objective functions, that are used to judge the efficacy of the stent, are based on the transient response of the stent during expansion, i.e. the metrics are based on the final expanded state of the stent.

3. The ratio of kinetic energy to total internal energy remains below 5% for both the time-steps.

4. The stresses on the components differ by numerically acceptable errors at the end of the analysis for both the time-steps.

While points 1 and 2 are straightforward conclusions based on the modelling methodology, points 3 and 4 are less obvious. For this reason, the ratio of energies is investigated for the two time-steps of 10^{-7} and 10^{-8} seconds. Figure 5.18 shows this ratio for the time steps of 10^{-7} s and 10^{-8} s on PA_Mesh2 . It is found that this ratio remains under 5% for all simulations. Next, along a central line in the plaque (shown in figure 5.16) the von-Mises stresses are compared for the two time-steps of 10^{-8} and 10^{-7} seconds. The results are plotted in 5.19 for PA_Mesh2. A time-step of 10^{-7} s underpredicts the von-Mises stresses, on average by approximately 12% when compared to a time-step of 10^{-8} s. Moreover, the stress pattern along the plaque line is similar for both the time-steps. Since the underprediction of von-Mises stresses applies to all the designs being evaluated in the design space, it is proposed to use a time step of 10^{-7} s for subsequent optimisation studies.

5.4 Validation

To validate the numerical results the data provided by the manufacturer (Cordis Corp., Johnson & Johnson) [135] is used. Table 5.6 lists the data obtained by the manufacturer during *in-vitro* testing at 37^{o}C. The second row in the table (with bold numbers) indicates

the nominal diameter of the stent and the subsequent rows indicate average stent/balloon diameters at specific balloon inflation pressures without taking into account lesion resistance. Since a 3.0 mm CYPHER stent is analysed in this study, the shaded column is applicable for validation. Figure 5.20 shows the comparison of the transient $p-d$ data obtained from numerical results (Mesh1 with a time step of 10^{-8} s) to the data listed in Table 5.6. The maximum difference in the numerical analyses vs. the *in-vitro* data is 8.0%. Apart from a good agreement with the experimental data, the transient expansion relationship obtained in this chapter matches well with the work of De Beule et. al. [134] and Gervaso et. al. [19].

Table 5.6: CYPHER stent expansion data provided by the manufacturer by *in-vitro* experiments

Inflation pressure (MPa)	**Diameter (mm)**					**Note**
	2.25	**2.50**	**2.75**	**3.00**	**3.50**	
0.608	1.95	2.17	2.43	2.64	3.13	
0.709	2.02	2.25	2.51	2.72	3.23	
0.811	2.10	2.32	2.58	2.80	3.32	
0.912	2.16	2.38	2.67	2.87	3.40	
1.1013	2.22	2.44	2.71	2.94	3.47	
1.115	2.26	2.49	2.76	2.99	3.52	Nominal
1.216	2.30	2.53	2.82	3.03	3.57	
1.317	2.34	2.56	2.84	3.07	3.61	
1.419	2.37	2.59	2.87	3.11	3.64	
1.520	2.39	2.62	2.89	3.14	3.67	
1.621	2.42	2.64	2.92	3.16	3.70	Rated Burst Pressure

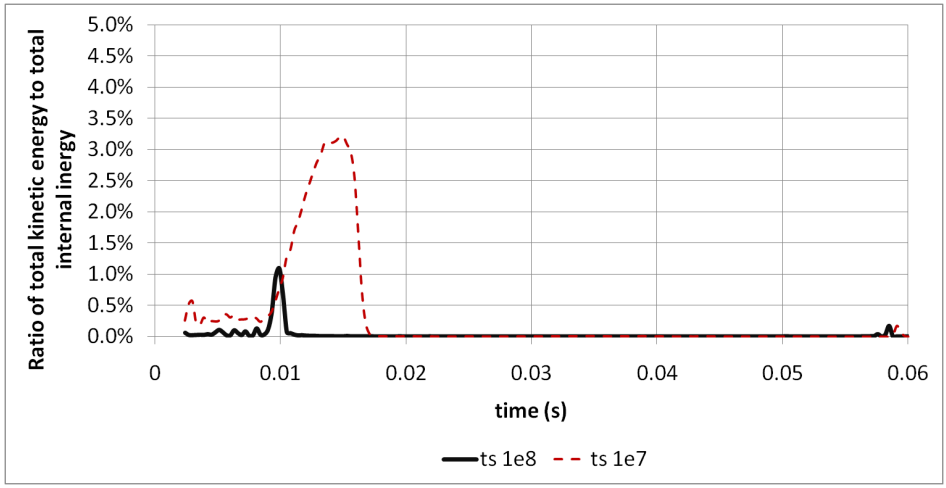


Figure 5.18: Ratio of energies for a time-step of 10^{-7} and 10^{-8} s for PA_Mesh2

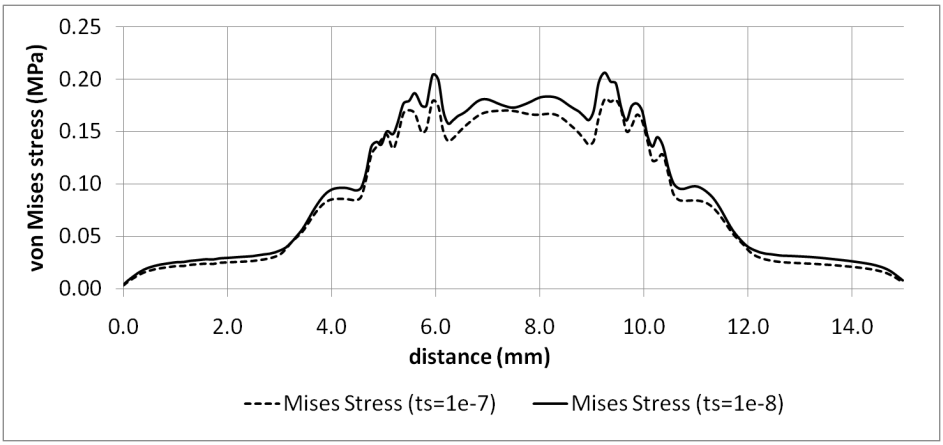


Figure 5.19: Stress comparison for time-steps of 10^{-7} and 10^{-8} seconds on PA_Mesh2

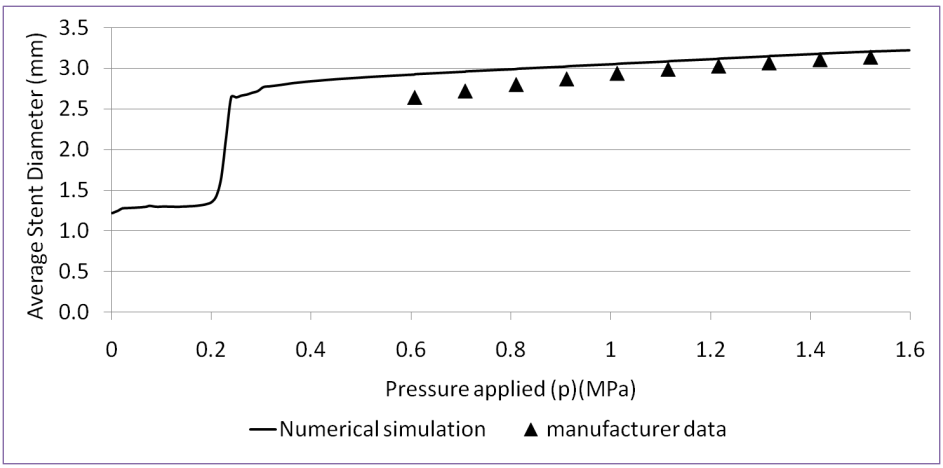


Figure 5.20: Comparison of numerical results with the experimental data

5.5 Results

The results for the free expansion and expansion into the plaque and the artery are presented in the following sub-sections.

5.5.1 Free expansion of the stent

The transient free expansion of the stent is shown in figure 5.21. It can be concluded that the expansion of the stent is not uniform across the longitudinal axis. The stent expands more at the distal ends than at the central part. This phenomenon is often referred to as dogboning, DB, and is defined as –

$$\text{Dogboning } (DB) = \frac{D^p_{distal} - D^p_{central}}{D^p_{distal}}, \tag{5.7}$$

where D^p_{distal} and $D^p_{central}$ are the stent diameters at the distal and central parts respectively at any given load p. The maximum dogboning observed is 2.2%.

It can also be observed that the final length of the stent is smaller than the initial length – as the stent expands in the radial direction, it foreshortens in the longitudinal direction. This phenomenon is referred to as foreshortening, FS, and is defined as follows –

$$\text{Foreshortening } (FS) = \frac{L_{initial} - L_{final}}{L_{initial}}, \tag{5.8}$$

where $L_{initial}$ and L_{final} are the initial and final lengths of the stent respectively. The foreshortening observed in this case is 6.19%.

The most important factor determining the efficacy of the stent is its radial strength, and recoil provides a good measure of this. Even though the stent might recoil many days after the procedure (late recoil), the simulations presented in this thesis can only evaluate the acute recoil (recoil immediately after the implantation procedure). Percentage acute recoil, R_c, is defined as –

$$R_c = \frac{D_{p_{max}} - D_{p_0}}{D_{p_{max}}}, \tag{5.9}$$

where $D_{p_{max}}$ and D_{p_0} are the average stent diameters at maximum loading ($p = p_{max}$) and after unloading ($p = 0$) respectively. The value of acute recoil observed in this case is 4.2%. Figure 5.22 shows the von-Mises stresses and max. principal plastic strains on the stent after the unloading phase.

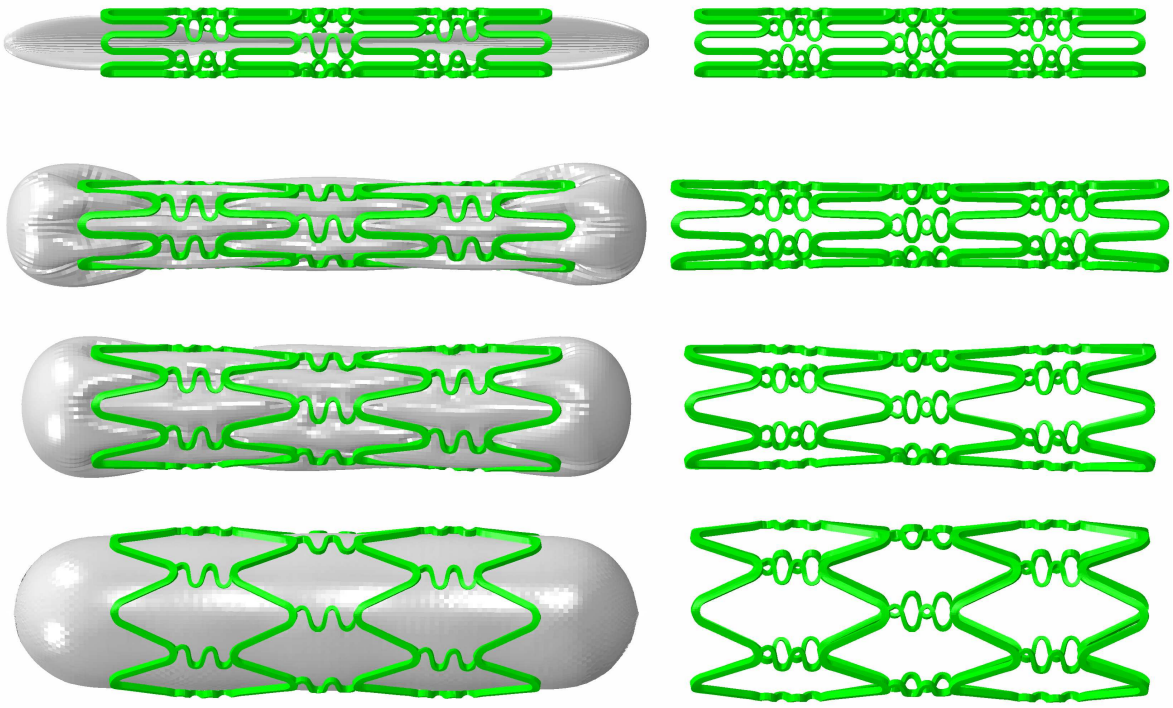


Figure 5.21: Stages of transient free expansion of the stent

5.5.2 Expansion into the plaque and the artery

The expansion of the stent into the plaque and the artery is depicted in figure 5.23. The parameters such as dogboning, foreshortening, and recoil defined in the previous section can be evaluated for this case too. The expansion behavior in this case is different than the free-expansion because of the additional contact/resistance of the lesion. After the acute recoil, the stent is perfectly apposed to the plaque inner surface and since the contact formulations allow embedding of the contact surfaces onto one another, the stent embeds itself into the plaque tissue. Figures 5.24, 5.25, and 5.26, respectively show the von-Mises stresses on the

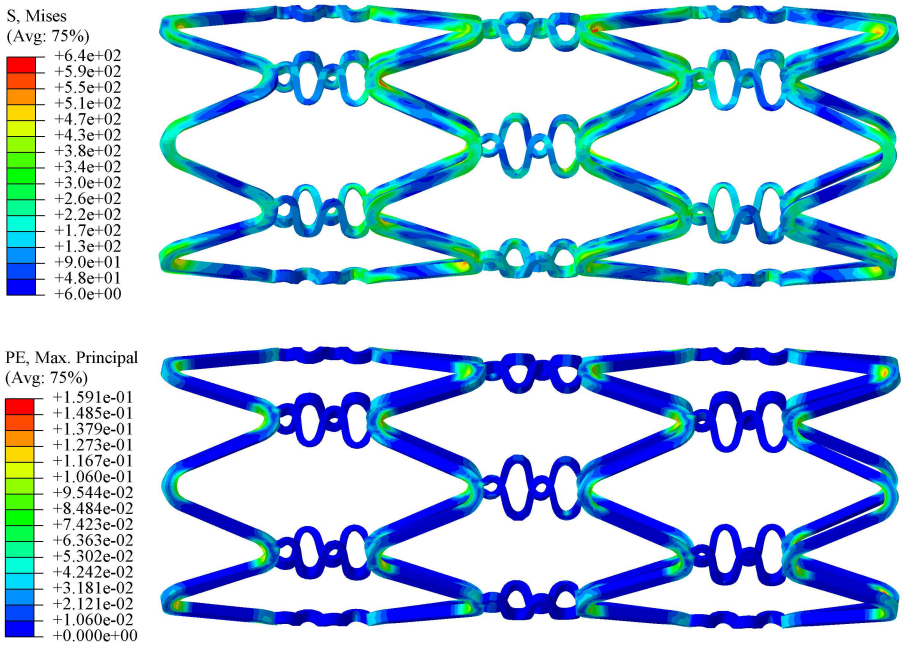


Figure 5.22: Von Mises stresses (MPa) and maximum principal plastic strains on the stent after free-expansion

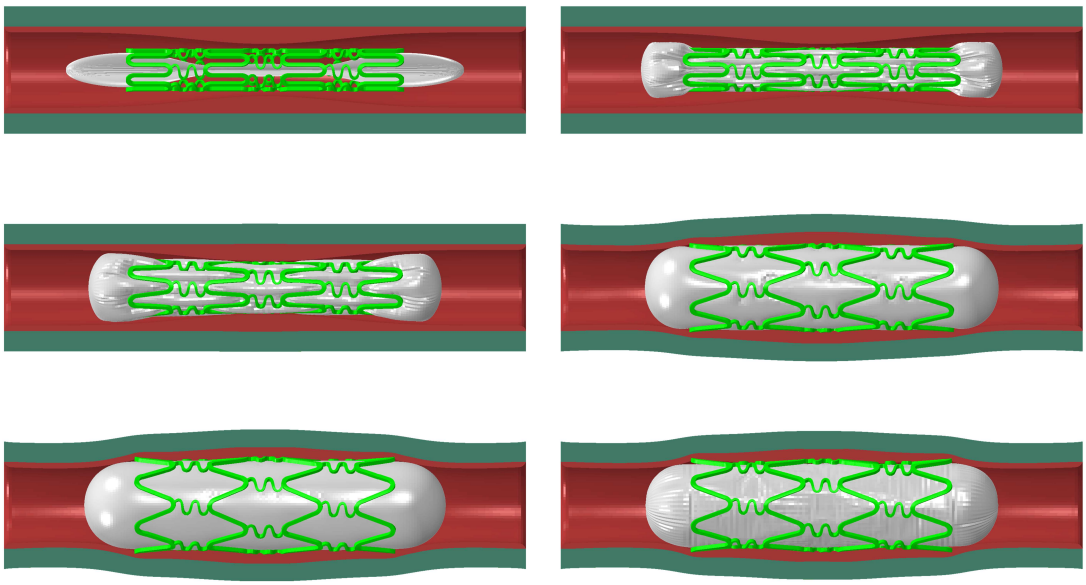


Figure 5.23: Stages of transient expansion of the stent into the plaque and artery

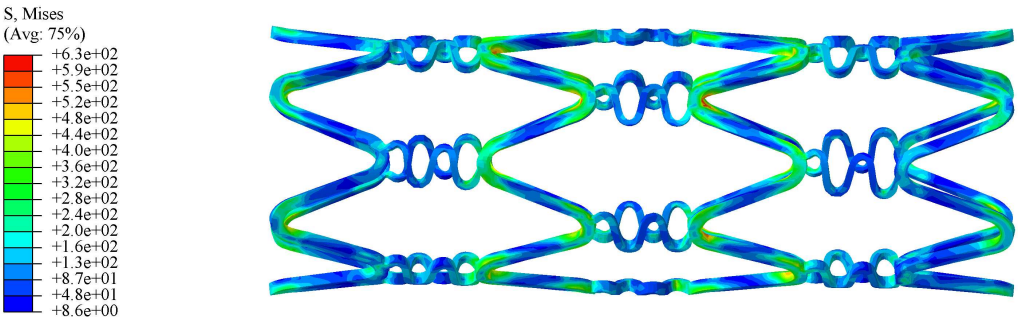


Figure 5.24: Von Mises stresses (MPa) on the stent

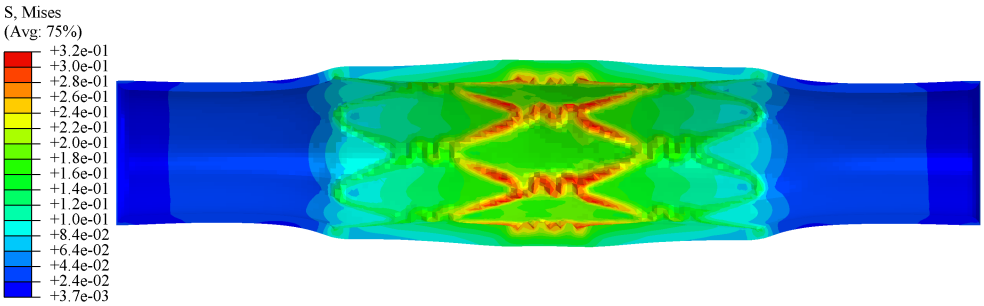


Figure 5.25: Von Mises stresses (MPa) on the plaque

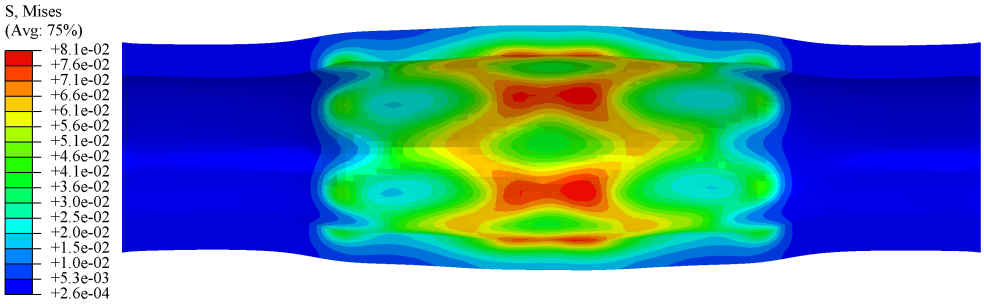


Figure 5.26: Von Mises stresses (MPa) on the artery

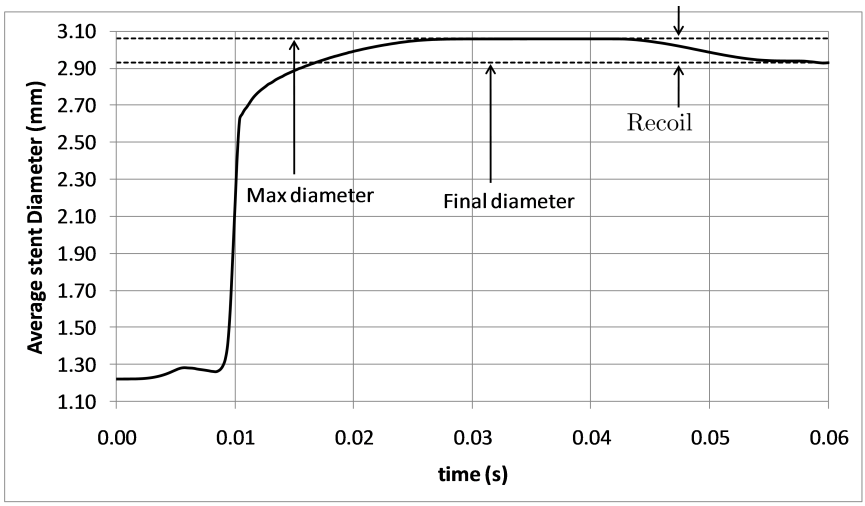


Figure 5.27: Average stent diameter for expansion into the plaque and the artery

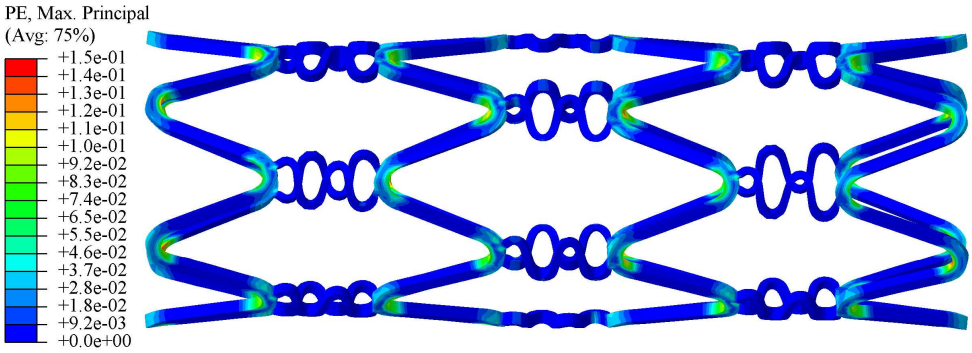


Figure 5.28: Max. Principal plastic strains on the stent after expansion into the lesion

stent, plaque, and the artery after the unloading phase. Figure 5.27 shows a plot of the average diameter of the stent against the pressure applied, and figure 5.28 shows the max. principal plastic strains for the stent at the end of the analysis (i.e. after the unloading phase).

5.6 Discussion

For both the free-expansion and expansion into the lesion, the maximum stresses for the stent are located at the curved parts of the struts (cf. figures 5.22 and 5.24). The maximum plastic strains are also located in the same regions (cf. figures 5.22 and 5.28). Since the simulation does not take into account the failure of the stent material, it is imperative to check that the maximum total strain does not exceed the failure strain [10]. The plastic strains are a key component in designing a stent. A good stent should, on one hand, undergo plastic deformation at the right locations so as to maintain the shape, which would lead to minimal recoil, while on the other hand the plastic deformations should not be large enough so as to cause failure. In the representative CYPHER stent considered in this chapter, the equivalent plastic strains are approximately 15%. The strut thickness (see Table 5.1) is 140 microns, and from Figure 5.6 it can be seen for this range of strut thickness, the maximum permissible strains are greater than 30%. Hence, it can be safely said that the representative CYPHER stent would not fail for the expansion diameters used in this study.

For metallic stents, this constraint of not exceeding the maximum allowed plastic strains and the observation that most stents have circumferential rings of varying shapes leads to the conclusion that circumferential rings represent a vital component of a metallic stent. This is primarily because very large plastic deformations for metals are not feasible, and it seems that the circumferential rings provide a very effective way for the expansion as well as the retention of the expanded shape through plastic deformations. Other ways of expansion usually employ the use of polymeric material properties. The patent by Johnson & Johnson [151] presents very interesting ways for stent expansion by utilizing the material properties of more plastic materials.

Looking at the stresses and strains at the links between the circumferential rings, they seem to not be playing an important part in the expansion of the stent, apart from tying the circumferential rings together. Nonetheless, the links play an important part in determining the flexibility of the stent, and their role is investigated in chapter 6.

Figure 5.25 shows that the stent leaves a heavy imprint in the plaque, causing the stresses to be highest at the contact interface. Moreover, since the stent has to work most in the central region, where the stenosis is highest, the stresses in the central plaque region are the highest. The stent imprint can also be seen in the artery surface in figure 5.26. Observing all these stress patterns in the plaque and the artery it can be concluded that the stress distribution in these tissues is a function of the stent geometry. A good stent would ideally try to distribute the stresses more uniformly, and minimize the maximum stresses. As mentioned earlier that the quantification of stresses is one way of quantifying injury caused to the tissue, a strategy to quantify this numerical information is needed. This forms the content of the next section.

5.7 Formulation of objective function

The two most important parameters relevant to restenosis are stent diameter post-stenting and injury. The following section describe these parameters and their quantification

5.7.1 Stent-diameter post stenting and recoil

Even though any of the parameters defined in section 5.5.1, viz. dogboning, foreshortening, and acute recoil can be used as an objective function to rank the performance of stents, acute recoil is chosen as one objective function in this thesis. This is primarily because acute recoil is directly related to the minimum stent diameter, and which in turn is related to the minimum lumen area, the primary goal of the stenting procedure. Moreover, minimum stent area has direct relevance to restenosis. Clinical trials [152] and other studies [153, 154] have shown that restenosis rates are very low if the post-procedural minimal stent area is high (greater than $9mm^2$). Kasaoka et. al. showed that for every 1-mm^2 increase in the minimum stent area, the predicted risk of restenosis drops by 19%. Thus, a higher stent area post stenting is a key deterrent of restenosis. Based on this the objective function, *Recoil*, without the denominator of R_c defined in equation 5.9, is

$$Recoil = D_{avg}^{max} - D_{avg}^{unload}, \tag{5.10}$$

where D_{avg}^{max} and D_{avg}^{unload} represent the maximum average diameter of the stent during the loading and hold phases and its diameter after the unloading phase, respectively. This *Recoil* is shown graphically in figure 5.27. It should be noted that in the above definition the average diameter is used, instead of the minimum diameter as suggested by clinical studies. This is justified as only one representative plaque geometry is used in this thesis. In reality, depending on the shape of the stenosis, the stent will expand differentially in different locations of the stenosis. Thus, the minimal stent diameter is determined, in part, by the shape and morphology of the stenosis. An average measure used here represents the average radial strength of the stent, and since the stenosis used in this thesis is not severe, is also representative of the minimum stent diameter. Moreover, it is observed that for later stent design variations, material of the stent, and the stenosis shape used in this thesis, the variations in stent diameter along the axial length of the stent are negligible.

Acute recoil is also reflective, in part, of the radial strength of a stent. One way to measure the radial strength of a stent is to apply an increasing inward pressure on the outer surface of the expanded stent and monitor the decrease in diameter. The pressure post which the stent cannot withstand any more increase in pressure (stent collapse) would represent the radial strength of the stent. Acute recoil measures this, in part, as the inward forces that are acting on the outer surface of the stent are due to the elastic nature of the artery which wants to regain its original configuration. Hence, although these forces are not increased to the extent of stent collapse, acute recoil for different stents takes into account the initial relationship between the external inward forces and diameter decrease.

Lastly, it is worth nothing that *Recoil* is also related to thrombosis, albeit indirectly. Incomplete stent apposition is defined as the separation of stent struts from the inner walls of the artery [155], and is shown to be correlated with stent thrombosis [155, 156]. If a coronary stent has high *Recoil*, then it is apparent that it will lead to incomplete apposition, and consequently increase the risk of thrombosis.

5.7.2 Quantifying injury caused by stent implantation

As mentioned in section 5.1, studies have shown a link between vessel injury, i.e. endothelial exposure, intima laceration, and rupture of media [107, 108, 109, 110, 119, 81], and restenosis. These factors can directly be related to the contact pressure between the expanded stent and the inner surface of the artery/plaque [125]. A high contact pressure between the stent struts and the intimal layer of artery can cause endothelial denudation and trauma which can lead to higher neointimal hyperplasia. Therefore, it is reasonable to include this contact pressure, which can be obtained by the FEA balloon expansion simulations developed, as a measure of injury.

Holzapfel et. al. [125] proposed two scalar metrics to evaluate a stent's efficacy by expansion analysis – $D1$ based on intimal pressure concentration caused by stent struts and $D2$ based on stress change in the arterial wall caused by stenting. As discussed above $D1$ is a measure directly relevant to restenosis. On the other hand, $D2$ represents an average change in the stress environment in the arterial tissue. Intuitively, it can be thought that if the average change in the stress environment post stent implantation is low, then the mechanical environment in the post-stented artery is not much different than that of the non-stented artery. This could imply lower injury caused. The results of the study by Holzapfel [125] are analysed and it is found that the two measures, unsurprisingly, are highly correlated ($r > 90\%$). Hence, from an optimisation point of view, it is of little significance whether $D1$ or $D2$ is used for comparing stents. Since the evaluation of the $D2$ metric is relatively easier to compute compared to $D1$, a variant of the $D2$ metric is used in this thesis as an objective function. This metric, volume average stress, VAS, which represents an average value of the von Mises stresses imparted on the artery due to stenting, is defined as follows:

$$VAS = \frac{\int_V \sigma \, dV}{\int_V dV}, \tag{5.11}$$

where σ represents the von Mises stresses and the integrals are calculated over the volumes of plaque and the artery. For numerical simulations, when meshes are involved, VAS can be calculated using the following formula –

$$VAS = \frac{\sum_{i=1}^{i=n} \sigma_i \delta V_i}{\sum_{i=1}^{i=n} \delta V_i}, \tag{5.12}$$

where σ_i represents the von Mises stress in the i^{th} element of plaque/artery, δV_i is the volume of the i^{th} element, n is the total number of elements in the plaque/artery within the axial domain of the stent plus a length of 0.5 mm on either side of the stent (as shown in figure 5.29), and the summation is carried out over the both the domains of plaque and artery.

VAS, averages the stresses over the volumes of plaque and the artery. Since we do not assume any pre-stresses before the stenting procedure, any stresses produced in the artery or plaque are seen as unwanted by the above definition. In a case that there do exist some pre-stresses (σ_{pre}), σ_i can be replaced by $\Delta\sigma_i$ in the above equations, where $\Delta\sigma_i = (\sigma_i - \sigma_{pre})$ [125].

The above two defined objective functions can be used to rank different designs of stents from radial strength and arterial injury perspectives, respectively. A lower value of both the objective functions indicates better designs.

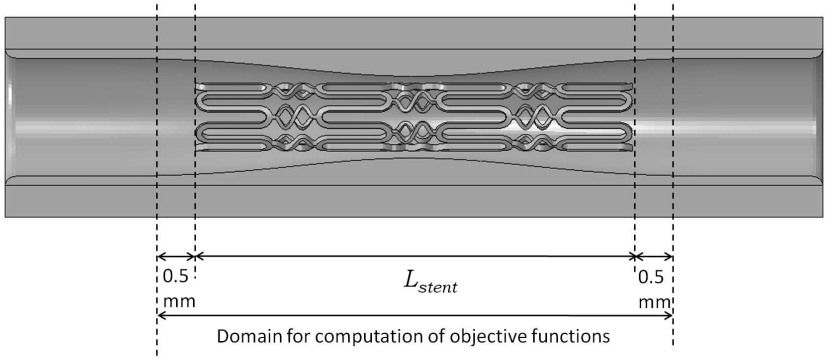


Figure 5.29: Volume domains used for computation of *VAS*

5.8 Conclusions

A methodology to evaluate the transient expansion of a coronary stent is presented in this chapter. Such analysis is then used to extract information such as stresses and strains for all the components, viz. stent, plaque, and the artery. For the CYPHER stent typical evaluation parameters such as dogboning, foreshortening, and acute recoil are calculated. Finally, two objective functions, *Recoil* and *VAS*, which can be used to compare the efficacy of different stents, are formulated for use in optimization studies. It is shown that the maximum stresses and strains for the expanded stent are located in the curved regions of the circumferential rings. It is also observed that for the plaque and artery, maximum stresses occur in regions where: a) stent struts come in contact with the tissue, and b) where the stenosis is high relative to other regions. Apart from analyzing the stent's expansion behaviour, which is critical for design optimization studies, this chapter has also established a procedure for the study of haemodynamics and drug-diffusion in realistically expanded geometries. Although a detailed analysis of haemodynamics is presented in Chapter 4, its primary shortcoming is to use non-realistic geometries to evaluate flow. This chapter links that gap, and provides relatively realistic geometries for haemodynamic and drug-diffusion evaluation (which are presented in chapter 8). Before moving on to such haemodynamic and drug-diffusion analysis, in the next chapter the issue of stent flexibility, which is key for deliverability, is explored.

Chapter 6

Flexibility of stents

The deployment of a stent involves the delivery system (comprising the balloon catheter and the stent) to be manoeuvred through highly curved anatomical pathways to the stenosis site [129]. Consequently, the flexibility of the stent becomes a concern, as potentially rigid stents give rise to various problems during deployment. The aims of this chapter are as follows

1. to develop a FEA model to model bending of an unexpanded coronary stent, and
2. to formulate an objective function which can be used to compare stents based on their flexibility.

This chapter starts with a review of methods that have been proposed to evaluate the flexibility of a stent. Thereafter an FEA model to simulate bending of stents is presented. Finally, a figure of merit, which can rank stents based on their flexibility, and hence in part deliverability, is proposed.

6.1 Introduction

Although there are no standard tests to evaluate flexibility of a stent, several methods have been used in the past. These methods can be broadly classified as either experimental or numerical. In the experimental regime either a one-point [157], three-point [158], or a four-point [157, 159] bending test is generally used. In the numerical methodology, bending is either measured by specifying the rotational displacements of the longitudinal ends and measuring the moments required [141], or by applying the moments at the longitudinal ends and measuring the rotational displacements [160]. Ormiston et. al. [158] compared 13 stent designs (two coil, two hybrid, and nine slotted tube designs) both before and after stent expansion by using a three-point bending test. They reported a strong correlation between expanded stent stiffness and number of longitudinal strut interconnections, and a poor correlation between expanded stent stiffness and strut-thickness. They also reported that expanded stents were more stiff than their crimped state counterparts. Szabaďts et. al. [157] compared six designs by using one-point and four-point bending tests. Mori et. al. [159] also used four-point bending tests on four different designs and proposed a simplified 2-D FEA method to evaluate the effect of stent parameters on flexibility. Petrini et. al. [141] compared the flexibility of the Bx VELOCITY stent and the Carbostent (Sorin Biomedica, Saluggia, VC, Italy) using

FEA. They proposed specifying the rotational displacements of the longitudinal ends to simulate bending, and then measuring the corresponding moments. They also showed that the response of one unit of stent closely followed the response of the full stent, and suggested that one unit of stent, being computationally cheaper, is ideal for flexibility measurement. Wu et. al. [160] proposed the use of multipoint constraints (MPC) in FEA to simulate the four-point bending tests. Their proposed method, where the analysis is driven by application of moments, can be applied to achieve more complex bending shapes (such as an 'S' shape, which is hard to simulate by specifying rotations) apart from the widely used pure-bending shape. They reported that flexibility of stents after they undergo plastic deformation is roughly 16 times larger than when they deform elastically only, and hence concluded that stents which can be plastically deformed easily are more flexible. Ju et. al. [161] proposed a repeated unit cell (RUC) approach to simulate pure bending of coronary stents, and reported that the geometry of the link was a key determinant of stent flexibility.

In this chapter the methodology suggested by Wu et. al. [160] is used to simulate bending. Thereafter, flexibility, both in the elastic and plastic regimes, is quantified to be used in optimization studies. The following sections discuss the approach used for this purpose.

6.2 Methodology

6.2.1 Geometry and material

Following the work by Petrini et. al. [141], which showed that the results of one unit of stent are very similar to that of a full length stent, only a unit model of stent is used for flexibility analysis in order to save computational time. The parameters of the stent geometry are the same as described in section 5.2.1.1, except that from the full length model, one unit is extracted. Figure 6.1 shows this model. The material of stent used in this chapter is the same as described in section 5.2.2.1.

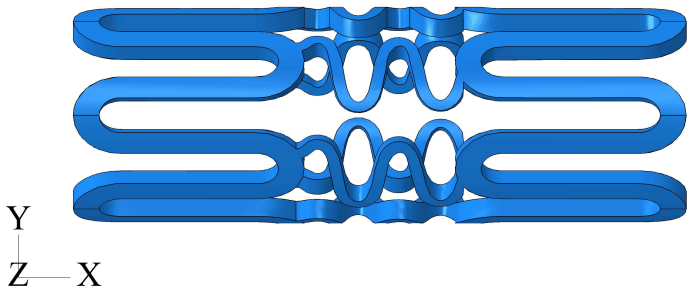


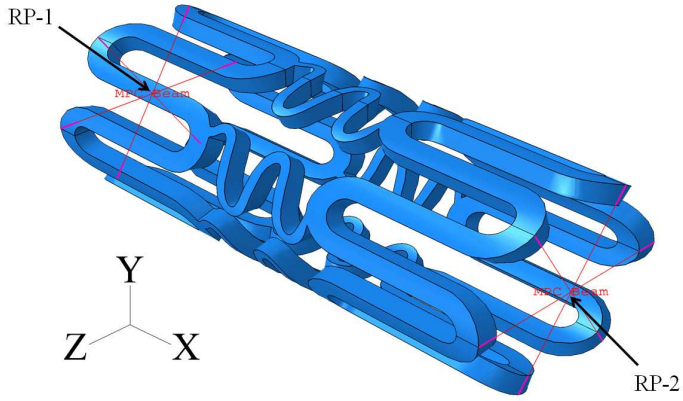
Figure 6.1: A unit model of the stent

It must be noted that due to the plastic deformations in the stent during delivery, the stress-strain history of the stent material for the expansion analysis will change. Consequently, kinematic hardening effects such as the Bauschinger effect, will come to play during the expansion of stent depending on the kind of plastic deformation occurred during the delivery of the stent. However, as will be seen in this chapter, and in chapters 7 and 9, the plastic deformation in the stent links is primarily important for delivery, and the plastic deformation in the circumferential rings is primarily important for the expansion of the stent. In this sense, for the stent designs considered in this thesis, these plastic deformations, and hence

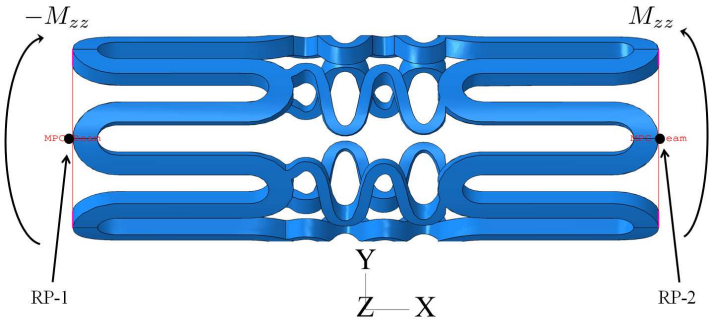
their strain history effects on each other are de-coupled. For these reasons the effect of strain history during delivery on the expansion analysis is ignored.

6.2.2 Boundary conditions and analysis

Abaqus/Standard 6.9.1 (Dassault Systèmes Simulia Corp., Providence, RI, USA) is used for the flexibility analysis. Six edges on each longitudinal end are tied to their respective reference points (*RP*-1 and *RP*-2) on the axial ends (as shown in Fig. 6.2a) using multi-point beam constraints. *RP*-1 is constrained not to move in either of the three X,Y, or Z directions and *RP*-2 is constrained to move only in the X direction. Moreover, the rotation of *RP*-2 is constrained about the X-axis. As the stent bends, the link on the compression side may come into contact with itself. Hence, a self contact for the stent is specified using surface-to-surface discretisation, Coulomb friction model, with a friction coefficient of 0.2, and penalty method for constraint enforcement [149]. Linearly increasing equal and opposite moments (M_{zz} and $-M_{zz}$) from 0 to 1.5 N-mm are then applied to these reference points as depicted in Fig. 6.2b. 10-node tetrahedral (C3D10) elements [160] are used for the static flexibility analysis.



(a) MPC constraints for application of moments



(b) Moment loads for flexibility analysis

Figure 6.2: Boundary conditions for flexibility model

6.2.3 Mesh dependence studies

Mesh dependence studies are performed to determine an appropriate mesh size for the analysis. The moment vs. curvature index graph is used for this purpose. M is the moment applied and χ is the corresponding curvature index ($\chi = 2\phi/Lu_{stent}$), where ϕ is the deformation angle

as shown in figure 6.3 and Lu_{stent} is the length of one unit of the stent. Table 6.1 details the three meshes used and the corresponding percentage differences relative to mesh-1. The results show that mesh-1 sufficiently discretizes the domain and that increasing the mesh density has minimal effect on the results.

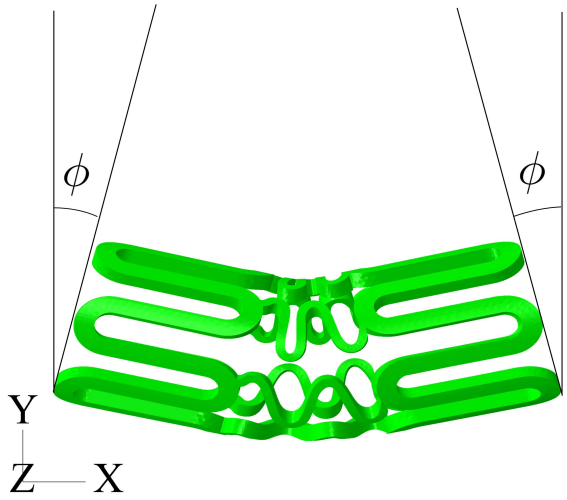


Figure 6.3: Phi (ϕ) measurement in the bending analysis

Table 6.1: Mesh dependence study for flexibility analysis

	No. of elements	**max. % error from mesh-1 in $M-\chi$ curve**
mesh-1	55,047	-
mesh-2	72,167	0.25%
mesh-3	105,392	1.69%

6.3 Results and discussion

Figure 6.4 shows the bending of the stent. Figures 6.4 (c) and (d) show how links on one side compress and on the other side expand to allow for a flexible shape. Such a winding feature of the links, which allows for both expansion and contraction, determines the flexibility of a stent. A stent lacking such a design feature, for example, one with no curves (straight links), would require considerably higher moments to bend by the same angle ϕ and hence would be relatively less flexible. Relative to the expansion, contraction of the links requires more consideration. The links, owing to their design, come into self contact after a certain ϕ. For a flexible stent, this should be avoided as far as possible because self-contact leads to the requirement of higher moments for further increases in ϕ. Consequently, flexibility decreases and this makes the stent less suitable for manoeuvring through high curvatures.

Contact of the links is depicted in Fig. 6.5. The images show how increasing moment, and consequently increasing χ leads to more points of self contact in the links. While in Figs. 6.5b and c there is only contact location, in Fig. 6.5d there are two contact locations. It is observed that as the moment is increased further, the two more pairs (marked as 'contact imminent' in Fig. 6.5 (d)) come into contact. This leads to a significantly higher increase in the moments required to further increase χ. This phenomenon is also reflected in the $M-\chi$

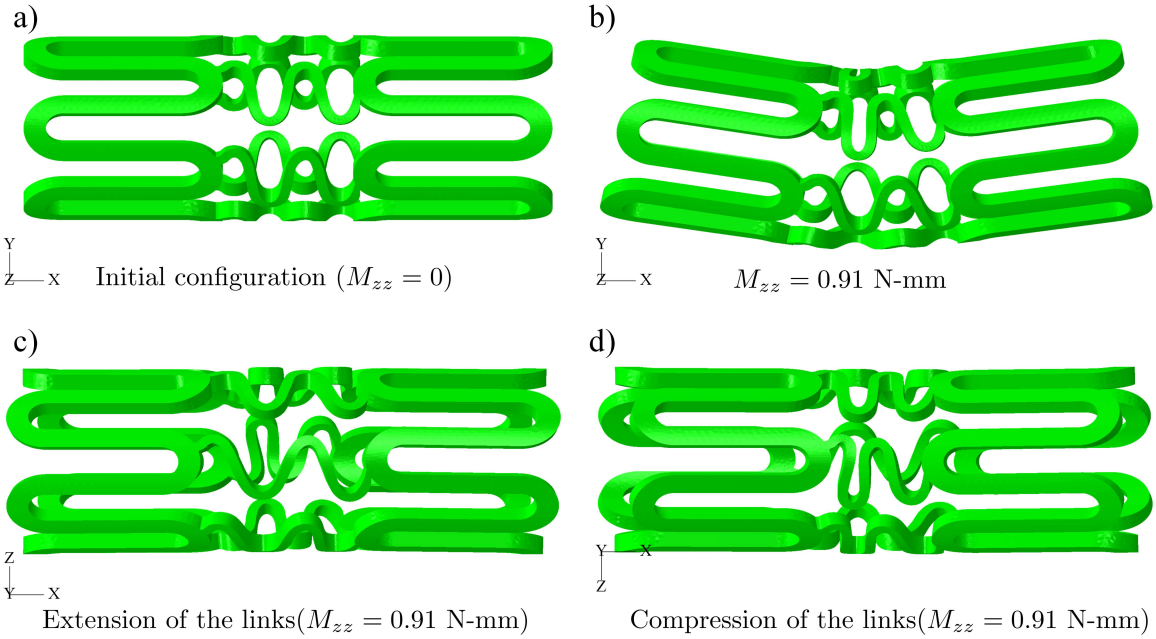


Figure 6.4: Flexibility results: snapshot at $M_{zz} = 0.91$ N-mm; a) initial shape; b) deformed shape; c) y^- view; and d) y^+ view

curve shown in Fig. 6.6. Initially, when the deformations are only elastic, an almost linear behaviour is observed. As the links deform plastically, the incremental moments required to further increase χ are very small until contact occurs. When the first contact occurs, and until there are only one or two contact locations, a slight increase in the moments is observed. After the contact locations increase to more than two, i.e. the link is in self contact and also in contact with the stent struts, a very steep change in the slope of the $M - \chi$ curve is observed. Thus, the design of the links, which involves consideration of both allowing the links to expand and contract and simultaneously avoid self contact, is a key determinant of flexibility in stents.

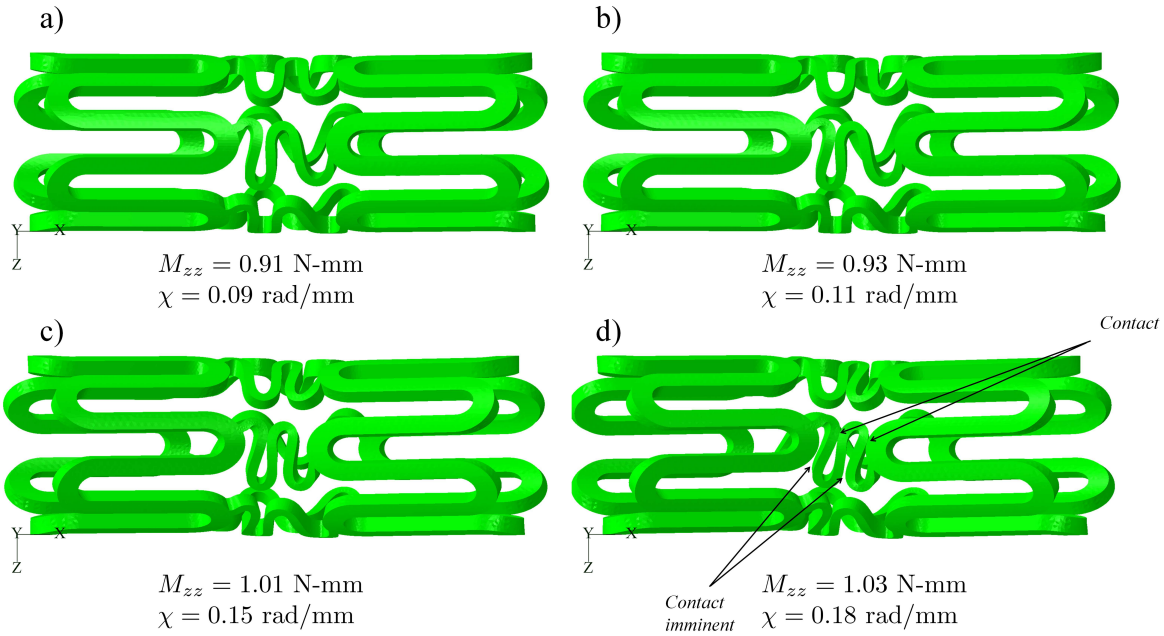


Figure 6.5: Contact in bending of stents

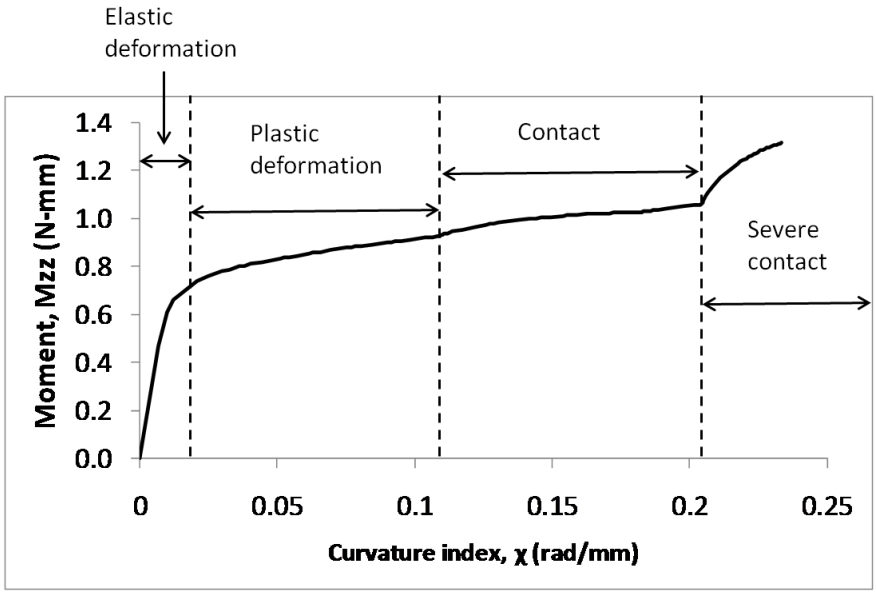


Figure 6.6: Moment vs. curvature index curve

6.4 Validation

The model used in this chapter is validated against the numerical results obtained by De Beule [11]. Figure 6.7 shows the numerical result, $M - \chi$ curve, obtained in this chapter against the numerical results obtained by De Beule. The results match well both qualitatively and quantitatively: the maximum difference being less than 10%. The differences can be attributed to the differences in the geometry of the CYPHER stent used by De Beule and the geometry used in this chapter.

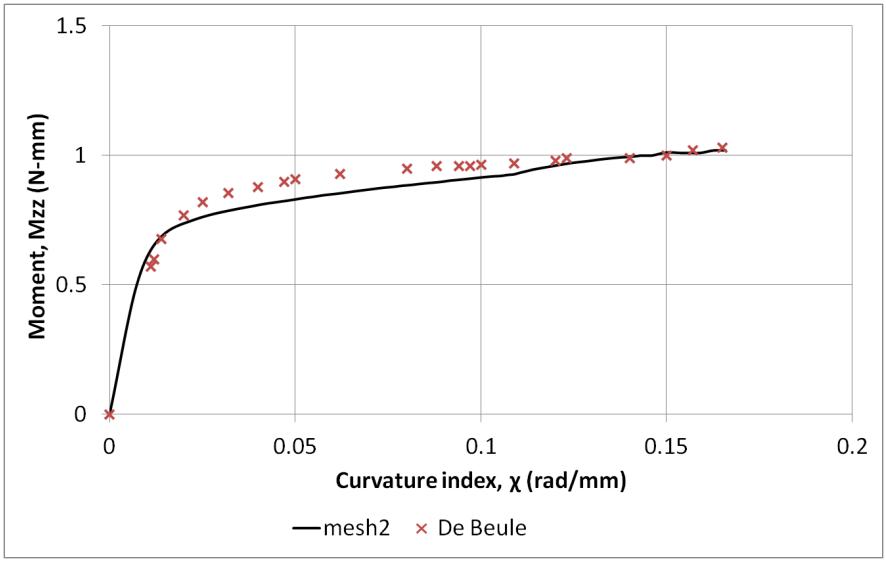


Figure 6.7: Validation against numerical results obtained by De Beule et. al. [11]

6.5 Formulation of objective function

The moment vs. curvature index graph gives a measure of flexibility of a stent. Petrini et. al. [141] have calculated the slopes both in the elastic and the plastic regions and used them

to compare stents. Such differentiation of slopes in various regions is thought inappropriate for purposes of design studies for two reasons – first that the number of such slopes, which depends on the severity of contact at different values of ϕ is not known *a priori*, and second that it is difficult to compare stents as two or more metrics are involved (one stent could have a lower slope in the elastic region and a higher slope in the plastic region and the other stent vice versa). Besides, the concept of slope in the non-elastic regions is a confusing one and can lead to errors. To overcome this problem the area of the $M-\chi$ curve is proposed. A smaller area indicates that, on average, a smaller moment is required to deflect the stent. Such a formulation overcomes the two aforementioned problems while dealing with slopes. Consequently, the flexibility metric, FM, is defined as –

$$FM = \int_0^{\chi_{max}} M(\chi)\, d\chi, \tag{6.1}$$

where χ_{max} is either determined by the parameterisation used for a particular design study, or the maximum curvature that is expected in the path of the stent to the stenosis site. Since data about the latter are not available, while using the former method, the value of χ_{max} should be cross-checked to ensure that sufficient plastic deformation is taken into account. It should be noted that since the CYPHER stent is symmetric in the circumferential direction, i.e. it has uniformly repeating strut segments and links in the circumferential direction, the proposed bending test is required only once by application of moments in a direction perpendicular to one of the links. For stents which do not have this property, bending tests should be performed in all directions since the $M-\chi$ curve will vary depending on which direction the moment is applied. In such cases, the maximum of FM in all the bending directions can be taken as the objective function.

6.6 Conclusions

A method to evaluate flexibility is implemented in this chapter. The importance of the design of links is depicted both qualitatively and quantitatively. It is concluded that a design which avoids self contact and number of contact locations, for higher values of curvature index is, in general, more flexible. Based on the results obtained, an objective function, called the flexibility metric (FM), to quantify flexibility into a single numerical quantity is proposed.

As part of the analysis of stents, in order to formulate the comparison figures of merit, this thesis has so far covered haemodynamics (flow), balloon expansion (structural), and flexibility analyses. The drug diffusion merits still need to be formulated. These will be formulated in chapter 8, but before that in the next chapter, a preliminary constrained optimisation study is presented.

Chapter 7

Constrained optimisation of coronary stents

This chapter[1] presents a constrained optimisation study on coronary stents. The thesis has already formulated metrics for haemodynamic alteration (in chapter 4), structural strength (in chapter 5), arterial injury (in chapter 5), and flexibility (in chapter 6). In this chapter, the metrics from balloon expansion analysis from chapter 5, viz. VAS and $Recoil$, and flexibility analysis from chapter 6, FM, are used. The haemodynamic metric from chapter 4, $HLRFI$, is excluded in this chapter because of two reasons: a) CFD models to use the expanded geometry from balloon expansion analysis for haemodynamic evaluation have not yet been explored; and b) the metrics for VAS, $Recoil$, and FM are all related to finite element analysis, and hence from a practical point of view can be obtained from one FEA package, thereby eliminating the hard task of transferring geometries between FEA and CFD packages. Nonetheless, in this chapter a fast model to evaluate drug-distribution is proposed. In line with the thought of keeping all analysis in an FEA package, this model does not include flow in the lumen (as that would be in the regime of CFD) and hence is formulated using a heat-transfer analogy with a finite element formulation. With this background, the aims of this chapter are

1. to develop an FEA based model to evaluate drug-distribution in a stented artery,

2. to develop a parameterisation technique which can be used to create varying stent geometries,

3. to perform a constrained optimisation study, in order to demonstrate improvement in stent performance in terms of radial strength, arterial injury, flexibility, and drug delivered, and

4. to evaluate trends between various merit functions and the parameters representing the stent geometry.

[1]The contents of this chapter are published in the below mentioned article with the author's supervisory team.
Pant, S., Bressloff, N. W., and Limbert, G., Geometry Parameterization and Multidisciplinary Constrained Optimisation of Coronary Stents. *Biomechanics and Modeling in Mechanobiology*, 11:61–82, 2012.

This chapter is organised as follows: first, a review of the few optimisation studies in the area of coronary stent design is presented; then an FEA model for drug-release is developed and an objective function from this model is formulated; thereafter, a parameterisation technique to represent stent geometry is proposed; and finally, the details and results of the constrained optimisation study are presented.

7.1 Introduction

Relatively few studies in the past have addressed the problem of stent design optimisation. Atherton and Bates [162] have presented different approaches to design optimization and their application in the field of stent design using computational fluid dynamics simulations. Li et. al. [163, 164] performed an optimization study using a simplified expansion model for the MAC STENT^{TM}(amg international GmbH, Germany). Their analysis was based on the free expansion of the stent (without considering the lesion) and they sought designs maximising radial gain while minimising radial reduction, radial loss, and dogboning (uneven expansion of the stent along its axial length). After imposing proper constraints on maximum stresses , foreshortening etc., they combined the multiple objectives onto a single objective. Blouza et. al. [165] performed an optimization study using a 2-D model for the stent by evaluating the fluid-stent interaction. For the multi-objective problem of minimising the mean-square wall shear stress and minimising the mean swirl value near the struts, they used an ϵ-multiobjective evolutionary algorithm. Wang et.al [130] evaluated six different combinations of balloon lengths and stent types to investigate design features to reduce dogboning and foreshortening. Bedoya et. al. [166] proposed a three parameter based technique to generate stent designs and after evaluating eight designs concluded that large strut spacing, radius of curvature, and large amplitude of the circumferential rings result in lesser stresses in the artery. Timmins et. al. [167] took the data of Bedoya et. al. and using Lagrange interpolation approximated the values for relevant stresses, lumen gain, and maximum cyclic radial deflection. They then identified designs minimising a weighted sum of these three metrics. Recently Wu et. al. [168] used a 2-D morphing technique on biodegradable magnesium alloy stents to sequentially minimise maximum principal strain and maximise mass.

7.2 A model for drug release

In this section a model for drug-release is developed as a heat transfer problem in Abaqus/ Standard 6.9.1 (Dassault Systèmes Simulia Corp., Providence, RI, USA). As mentioned in the beginning of this chapter the motive behind such a formulation is that all analyses for the optimisation study can be performed in the FEA package. Hence, a finite element formulation of drug-diffusion is needed. The similarity between the diffusion equations and the heat equation is utilised for this purpose [169, 170, 171]. Since the objective of this model is to evaluate the differences in the performance of a stent for different geometric designs only, the drug release is modelled as a steady state diffusion problem. Hence, in the regions of plaque and artery the following equations are solved respectively

$$D_p \nabla^2 C_p = 0 \qquad \text{in plaque region,} \tag{7.1}$$

$$D_a \nabla^2 C_a = 0 \qquad \text{in artery region,} \tag{7.2}$$

where C_p and C_a are drug concentrations in the plaque and artery, respectively, and D_p and D_a represent the drug diffusivities in the plaque and artery regions, respectively. The stent region is set to unity drug concentration. Such a Dirichlet boundary condition for drug release has also been used by Kolachalama et. al. [16]. The perivascular side, as shown in Fig. 7.1, is set to a zero flux boundary condition. For the luminal side, shown in Fig. 7.1, there are two possibilities for the boundary condition: first, to assume a zero flux boundary condition; and second, to assume a flux on the assumption that the drug concentration in the lumen is zero, which in turn is based on the assumption that all drug in the lumen is washed out owing to the luminal flow. In this chapter, the former approach is adopted. The latter approach is explored in chapter 8, where the drug concentration in the lumen is also solved for, and hence the flux on the boundary is calculated more accurately. The longitudinal ends are set to a zero concentration boundary condition. For the stent-plaque interface and the plaque-artery interface, the following formulation is adopted –

$$q(C_s, C_p) = k_{sp}(C_s - C_p) \qquad \text{for the stent-plaque interface,} \tag{7.3}$$

$$q(C_p, C_a) = k_{pa}(C_p - C_a) \qquad \text{for the plaque-artery interface,} \tag{7.4}$$

where q represents the flux along an interface, C_s is the drug concentration in stent, and k_{sp} and k_{pa} represent the permeabilities of the stent-plaque and plaque-artery interface, respectively. For the values of the diffusivities of the drug in the plaque and the artery, the work by Feenstra et. al. [171] is used. Their diffusivity values of 2.2×10^{-5} mm^2/s and 3.2×10^{-5} mm^2/s for intima and media are as D_p and D_a, respectively, in this model. Disregarding the porosity field, that was proposed by Feenstra et. al. [171], in order to simplify the model, k_{sp} and k_{pa} are assumed to be 10^{-4} mm/s and 1 mm/s, respectively.

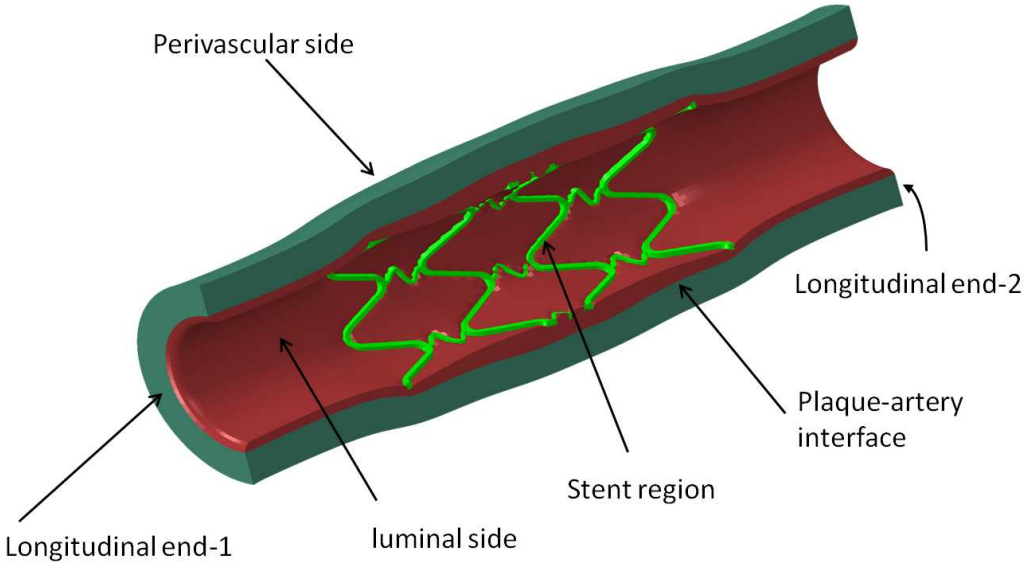


Figure 7.1: Boundary condition surfaces for drug release analysis (baseline geometry)

7.2.1 Formulation of objective function

The drug related measure of performance can be considered in two aspects – a) a measurement of the drug delivered into the tissue globally, and b) a measurement of the tissue volumes which

receive low drug (less than therapeutic levels). In this thesis, the first aspect is considered for two reasons – first that the determination of drug levels below which the effects are non-therapeutic (specially for the steady state drug distribution model adopted in this chapter) is not obvious, and second that depending on the minimum level set (most likely different for different types of drug) the optimisation results will be different. The volume average metric for drug, VAD, to measure the amount of drug transported into the tissue globally, is defined as follows –

$$VAD = \frac{\int_V cdV}{\int_V dV}, \tag{7.5}$$

where c is the drug concentration in the tissue region (i.e. plaque and artery combined) lying within the axial domain of the stent plus a length of 0.5 mm on either side of the stent and V is the volume. For discretized domains VAD can be calculated by

$$VAD = \frac{\sum_{i=1}^{i=n} c_i \delta V_i}{\sum_{i=1}^{i=n} \delta V_i}, \tag{7.6}$$

where c_i is the concentration of drug in an element of volume δV_i, and the summation is carried out over all the elements in the tissue lying in the axial domain of stent plus a length of 0.5 mm on either side of the stent. This domain is similar to that used in the formulation of VAS and is shown in figure 5.29.

$$VAD = \frac{\sum_{i=1}^{i=n} c_i \delta V_i}{\sum_{i=1}^{i=n} \delta V_i}, \tag{7.7}$$

A measure of uniformity of the drug-distribution is excluded from this chapter, and is presented in chapter 8.

7.2.2 Stent geometry and parameterisation

Most stent designs used in clinical practice today have two distinct features – circumferential rings and links/ connectors. The circumferential rings are the primary feature for expansion as they undergo plastic deformation at the curved parts and the links provide flexibility to the stent during deployment. The circumferential rings are crucial for metallic stents as an alternate way of expansion, which would typically involve extremely large plastic deformations, seems improbable. A change in the material properties (e.g. by using polymers), however, can lead to very interesting stent designs which can sustain high plastic deformation (cf. patent by Contiliano and Zhang [151]). Nonetheless, confined to a circumferential ring and connector design topology, the present work uses a parametrization to parsimoniously alter the shape of the stent. Since a stent is one single structure, composed of two parts (rings and links), two parameters are used to describe the rings and two parameters to describe the links.

Geometry construction/ parameterisation is performed in Rhinoceros 4.0. The base structure for the circumferential rings is first created. Figure 7.2a shows this structure (note that the lines depicted in Fig. 7.2a are construction lines). The radius of the initial shape (semi crimped state) of the stent, R_{stent}, and its length, L_{stent}, are fixed to be equal to 0.75 mm and 8.0 mm respectively. Four rings in the longitudinal direction are used and there are 12 (N_y) curved parts in each circumferential ring. Consequently, dH is equal to $2\pi R_{stent}/N_y$, where dH

is the distance between successive linear regions in the circumferential rings. The parameter h_c, as shown in Fig. 7.2a, is allowed to vary. Since L_{stent} is constant, h_c controls the length of each circumferential ring, and hence the length of the links too. The small horizontal lines in Fig. 7.2a show the location where the links are joined to the circumferential rings. From each curved peak in the circumferential ring, these lines are at a distance of half the value of strut width (W_{strut}), which is the second variable parameter. In order to parameterise the link, a two parameter (p_1 and p_2) based polynomial of 5^{th} order is used. While seeking the form of the polynomial two considerations are kept in mind – first that the polynomial should take a value of zero at the ends, and second that the polynomial should have zero slope at the ends. The second condition ensures tangency at the ends so that the connectors can smoothly join the circumferential rings (c.f. Fig. 7.2b). Such a polynomial could take the following form

$$f(t) = p_1 t^2 (1-t)^2 \ ; \text{for } 0 < t < 1. \tag{7.8}$$

Observing that many stents (CYPHER stent, Multi-link ZETA stent etc.) have an oscillating link shape, a second link parameter, p_2, is introduced and the parametrisation is represented as a fifth order polynomial

$$f(t) = p_1 t^2 (1-t)^2 (p_2 - t) \ ; \text{for } 0 < t < 1 \ \ \& \ \ 0 < p_2 < 1. \tag{7.9}$$

Equation 7.9 now represents a smooth curve which takes a value of zero at $t = p_2$ and has opposite sign values for intervals $(0 < t < p_2)$ and $(p_2 < t < 1)$. Adding more parameters such as p_2 would introduce more such oscillations in the function value. However, to keep the number of parameters low, this is avoided here. Since the peak values of $f(t)$ in Eq. 7.9 depends on the value of p_2, the final form chosen for this chapter is the following –

$$f(t) = \frac{p_1 t^2 (1-t)^2 (p_2 - t)}{|t_0^2 (1-t_0)^2 (p_2 - t_0)|} \ ; \text{for } 0 < t < 1 \quad \& \quad 0 < p_2 < 1, \tag{7.10}$$

where, t_0 is

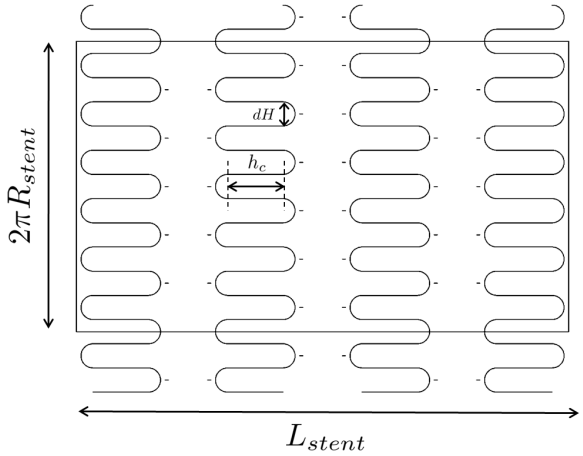
$$t_0 = \begin{cases} t_1 \text{ if } F(t_1) > F(t_2) \\ t_2 \text{ if } F(t_1) < F(t_1) \end{cases}, \tag{7.11}$$

and t_1 and t_2 are the extremum values (besides 0 and 1) for the function–

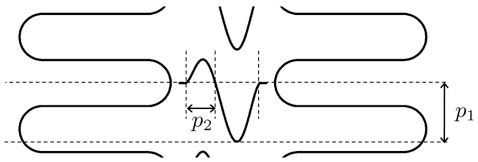
$$F(t) = |t^2 (1-t)^2 (p_2 - t)|. \tag{7.12}$$

Such a formulation ensures that p_1 is the maximum value that the curve takes. Once the curves for both the circumferential rings and the links are obtained, they are offset by a distance of $W_{strut}/2$ and $n_{width}/2$ respectively on either side, where n_{width} is the width of the links (constant in this chapter with a value of 0.07 mm, which is roughly half of the strut-width value for the CYPHER stent, Cordis Corporation, Johnson & Johnson). Figure 7.2d shows this network of curves after the offset operation. All the regions bounded by the curves are then converted to surfaces, the surfaces extruded to a value equal to strut thickness (radial dimension) T_{strut} (constant and equal to 0.14 mm, based on the strut height of the CYPHER stent [34], in this chapter), and the resulting structure wrapped to make the final stent shown in Fig. 7.2e.

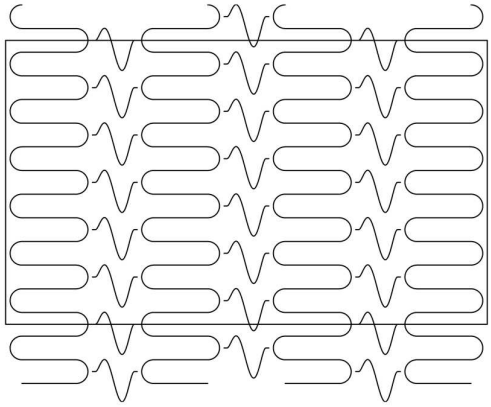
In essence, there are four parameters which are allowed to vary for each stent design – W_{strut}, h_c, p_1, and p_2. For the baseline geometry, shown in Figure 7.3, W_{strut}, h_c, p_1, and p_2



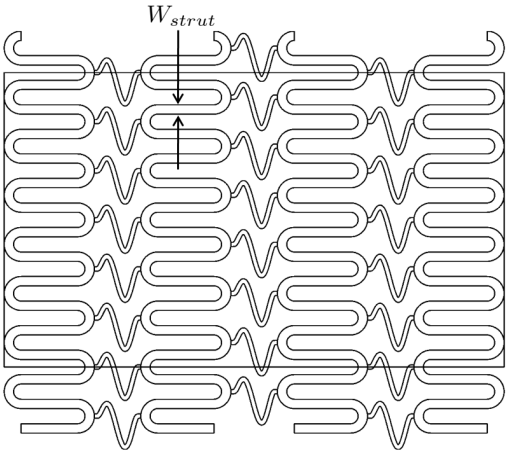
(a) The base shape for parameterisation (h_c as the parameter)



(b) Parameterisation for the link ($p1$ and $p2$ as parameters)



(c) Figure with links and the base structure



(d) Shape after offsetting the curves(W_{strut} as the parameter)

(e) Final stent geometry

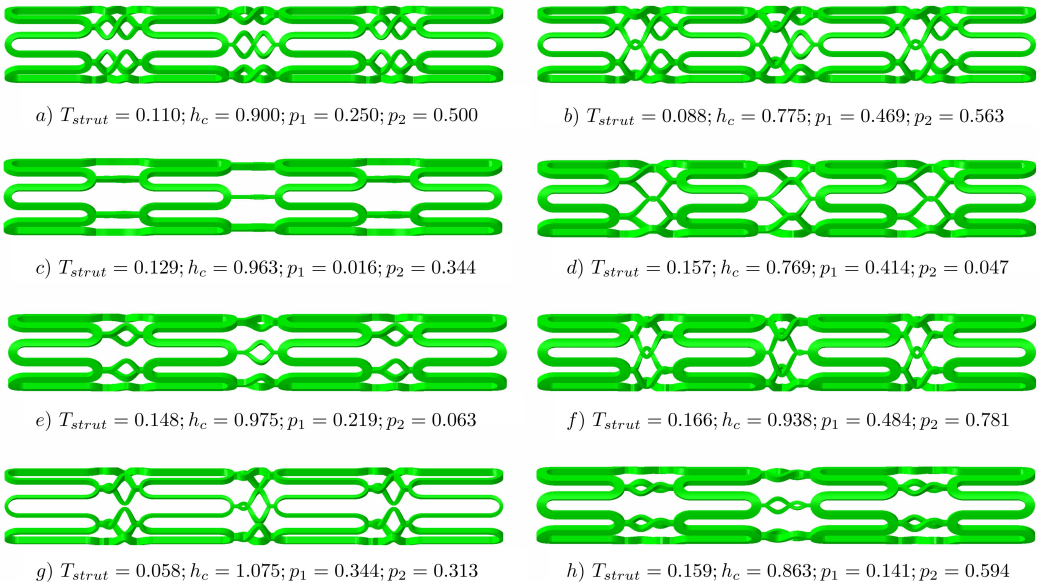
Figure 7.2: Stent parameterisation: stages in creating the stent geometry

Figure 7.3: Baseline geometry for constrained optimisation studies

Table 7.1: Limits imposed on the design parameters

	W_{strut} mm	h_c mm	p_1 mm	p_2
Min.	0.05	0.7	0	0
Max.	0.17	1.1	0.5	1

are specified to be equal to 0.11 mm, 0.9 mm, 0.25 mm, and 0.5 units respectively. Figure 7.4 shows some alternate designs created by using this parameterisation.



a) $T_{strut} = 0.110; h_c = 0.900; p_1 = 0.250; p_2 = 0.500$

b) $T_{strut} = 0.088; h_c = 0.775; p_1 = 0.469; p_2 = 0.563$

c) $T_{strut} = 0.129; h_c = 0.963; p_1 = 0.016; p_2 = 0.344$

d) $T_{strut} = 0.157; h_c = 0.769; p_1 = 0.414; p_2 = 0.047$

e) $T_{strut} = 0.148; h_c = 0.975; p_1 = 0.219; p_2 = 0.063$

f) $T_{strut} = 0.166; h_c = 0.938; p_1 = 0.484; p_2 = 0.781$

g) $T_{strut} = 0.058; h_c = 1.075; p_1 = 0.344; p_2 = 0.313$

h) $T_{strut} = 0.159; h_c = 0.863; p_1 = 0.141; p_2 = 0.594$

Figure 7.4: Sample designs generated from the parameterisation described in section 7.2.2 and Figure 7.2

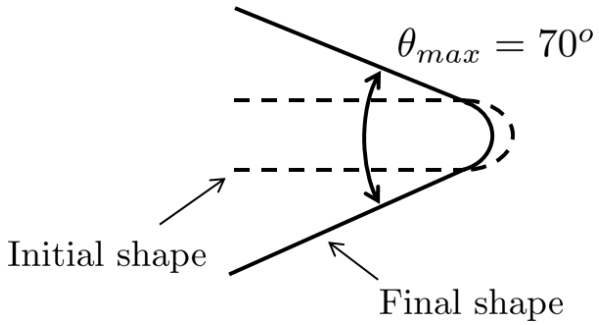


Figure 7.5: Explanation of the lower bound on the parameter h_c

Table 7.1 shows the bounds imposed on each of the four free parameters. The choice of the bounds is decided by the following:

1. W_{strut}: The upper bound on W_{strut} is decided by the fact that a very high value of W_{strut} results in self contact of the circumferential rings (since the number of curved regions in each circumferential ring is constant). The lower bound on W_{strut} is decided by meshing constraints.

2. h_c: From Fig. 7.2a and 7.2d it is clear that the axial length occupied by one circumferential ring is $h_c + dH + W_{strut}$. The value of dH is fixed (0.393 mm) owing to the fact there are 12 curved regions in each circumferential ring. For maximum values of W_{strut} and h_c the axial length of one ring is 1.663 mm. Consequently the length of four rings is 6.650 mm. This leaves about 17 % of the total axial length (8 mm) for three sets of links, implying just over 5% of the total axial length for each link. Any increase in the upper bound of h_c makes this percentage even smaller which is not considered desirable.

   Based on a simplistic idea of seeing the expansion as a stretching of the circumferential rings in a flat plane, the lower bound reflects a maximum limit of an angle of 70^o on the angle between the curved regions of the circumferential rings. Figure 7.5 shows this scenario.

3. p_1: The upper bound for p_1 is chosen considering that a very high value of p_1 results in contact of one link with the one vertically above it. The lower bound for p_1, $p_1 = 0$, results in straight links.

4. p_2: The parameter p_2 by definition lies between 0 and 1.

7.2.3 Baseline results

The baseline geometry, defined by the middle values between the upper and lower bounds for all the four parameters, is shown in Figure 7.3. The values of the various metrics of $Recoil, VAS, VAD$, and FM for this design are 0.168 mm, 0.046 MPa, 1.278 units, and 0.505 N-rad, respectively. The balloon expansion and flexibility analysis for this design yields results similar to those reported in chapters 5 and 6. These results can be found in appendix C. Here, the results of the drug-distribution model on the baseline geometry are presented.

The steady state drug contours, from the drug release simulations, are shown in Fig. 7.6. Due to higher penetration of the stent into the plaque tissue in the central region, and the fact that the central region receives drug from both ends of the stent, the drug concentration is higher in these areas. The impression of the stent can be seen in the plaque drug concentration contours too. This is expected as the stent is the source of the drug. The volume average drug in the plaque and the tissue region calculated from Eq. 7.7 is 78.2×10^{-2} units.

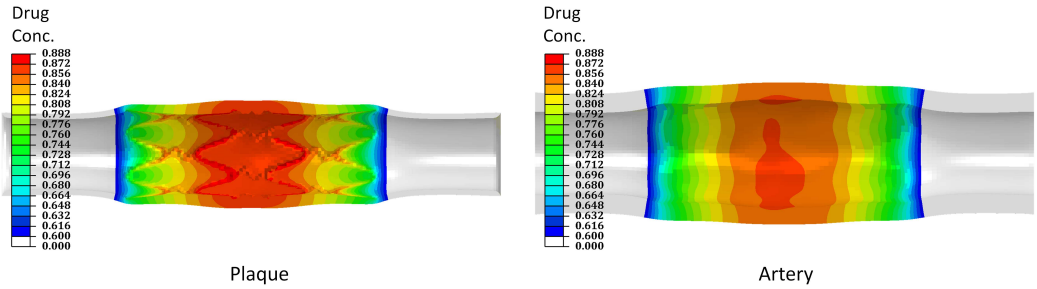


Figure 7.6: Steady state drug contours for the baseline geometry

7.2.4 Optimization

7.2.4.1 Optimization problem

The four metrics used in this chapter for evaluation of stents are: *Recoil* (equation 5.10), *VAS* (equation 5.12), VAD^{-1} (cf. equation 7.7), and FM (equation 6.1). Note than an inverse of *VAD* is chosen so that a lower value across all the four merits, implies a better design. With these four metrics, three optimisation problems are formulated as single objective constrained problems. The setting up of the inequality constraints of the form $g_i(\mathbf{x}) < g_i^{max}$ requires an understanding of the minimum allowed values of *Recoil*, *VAS*, *VAD*, and FM. Such minimum allowed values for each objective are hard to determine. To overcome this problem, the approach used in this study is to start from the baseline geometry and then search for improvement in one of the objectives without compromising on any of the other three. So, the optimisation problems are–

$$\begin{aligned} &\textit{Minimise} \quad f(\mathbf{x}) \ , \ \textit{s.t.} \\ &g_1(\mathbf{x}) < g_1(\mathbf{x})_{baseline}, \\ &g_2(\mathbf{x}) < g_2(\mathbf{x})_{baseline}, \\ &g_3(\mathbf{x}) < g_3(\mathbf{x})_{baseline}, \end{aligned} \tag{7.13}$$

where $\mathbf{x}$ is the vector containing the four parameter values and $f(\mathbf{x})$ is either of the three *VAS*, VAD^{-1}, or FM. Depending on which of the *VAS*, VAD^{-1}, or FM is being minimised, the other three metrics are formulated as constraints, $g_1(\mathbf{x})$, $g_2(\mathbf{x})$, and $g_3(\mathbf{x})$, with one of them always being *Recoil*. The idea of having *Recoil* always as a constraint stems from the discussion in section 5.7.1, where it was mentioned that clinical trials have shown a significant reduction in restenosis rates for minimum stent area greater than certain thresholds. Hence, in the process of designing a stent, *Recoil*, is more likely to be formulated as a constraint, as opposed to other metrics, for example *VAS*, for which clinical thresholds are hard to determine.

7.2.4.2 Optimisation methodology

Surrogate modelling (see section 3.2) is used to model the response of each of the four metrics. Surrogate modelling is pertinent as the evaluation of one design for all the four metrics in this chapter takes over 24 hours (wall time), using 8 parallel compute processes (Microsoft Windows 64-bit high performance computing platform, Intel quad core 2.8 GHz processor, 16 GB RAM). In particular, the Gaussian Process surrogate modelling approach (see section 3.2.3) is adopted in this chapter.

An LP_τ [172] based sampling plan is used to construct a 40-point sample. The choice of 40 points is based on the general rule of sampling 10 times the number of parameters. High fidelity simulations, i.e. the analyses presented in chapters 5, 6 and section 7.2, are carried out for each of these 40 designs, and corresponding metrics of *Recoil*, *VAS*, VAD^{-1}, and FM are extracted. The first five columns of Table 7.2 show the normalised values of the four parameters for the 34 successful evaluations out of 40. The analysis for six points failed due to issues faced in the automated processes of geometry export from the CAD package, geometry import into the FEA package, and automated meshing.

Based on the metric values obtained for the 34 designs, individual Gaussian Process models are constructed for each of the metrics of *Recoil*, *VAS*, VAD^{-1}, and FM. These models are validated using the SCVR methodology discussed in section 3.2.3.2.

Table 7.2: Design table showing the parameter values (normalised) for the points analysed and respective objective function values

Design	W_{strut}	h_c	p_1	p_2	*Recoil* mm	*VAS* MPa	VAD^{-1} units	*FM* N-rad
DOE-1	0.500	0.500	0.500	0.500	0.168	0.046	1.278	0.505
DOE-2	0.250	0.750	0.250	0.750	0.262	0.039	1.515	1.096
DOE-3	0.750	0.250	0.750	0.250	0.112	0.046	1.233	0.587
DOE-4	0.625	0.125	0.375	0.375	0.123	0.045	1.283	0.743
DOE-5	0.375	0.375	0.625	0.125	0.183	0.043	1.422	0.590
DOE-6	0.875	0.875	0.125	0.625	0.139	0.047	1.192	1.894
DOE-7	0.063	0.938	0.688	0.313	0.316	0.036	1.740	0.510
DOE-8	0.563	0.438	0.188	0.813	0.145	0.046	1.285	1.401
DOE-9	0.313	0.188	0.938	0.563	0.186	0.044	1.440	0.416
DOE-10	0.813	0.688	0.438	0.063	0.132	0.047	1.207	0.941
DOE-11	0.188	0.313	0.313	0.688	0.222	0.040	1.876	0.852
DOE-12	0.688	0.813	0.813	0.188	0.162	0.047	1.214	0.620
DOE-13	0.438	0.563	0.063	0.438	0.179	0.044	1.361	2.072
DOE-14	0.938	0.063	0.563	0.938	0.089	0.043	1.214	0.700
DOE-15	0.031	0.531	0.406	0.219	0.331	0.033	2.091	0.689
DOE-16	0.531	0.031	0.906	0.719	0.125	0.045	1.323	0.424
DOE-17	0.781	0.781	0.656	0.469	0.144	0.048	1.198	0.609
DOE-18	0.156	0.156	0.531	0.844	0.232	0.039	1.875	0.610
DOE-19	0.656	0.656	0.031	0.344	0.147	0.046	1.243	2.199
DOE-20	0.906	0.406	0.281	0.594	0.108	0.046	1.206	0.947
DOE-21	0.594	0.969	0.344	0.906	0.184	0.046	1.234	1.132
DOE-22	0.344	0.719	0.594	0.656	0.217	0.043	1.384	0.597
DOE-23	0.844	0.219	0.094	0.156	0.101	0.045	1.227	2.077
DOE-24	0.719	0.344	0.719	0.031	0.118	0.045	1.237	0.615
DOE-25	0.469	0.094	0.469	0.281	0.139	0.044	1.379	0.676
DOE-26	0.969	0.594	0.969	0.781	0.115	0.047	1.185	0.487
DOE-27	0.016	0.797	0.953	0.672	0.323	0.035	1.801	0.367
DOE-28	0.516	0.297	0.453	0.172	0.149	0.045	1.313	0.742
DOE-29	0.266	0.047	0.703	0.422	0.197	0.042	1.604	0.406
DOE-30	0.766	0.547	0.203	0.922	0.124	0.046	1.222	1.406
DOE-31	0.141	0.422	0.078	0.297	0.265	0.037	1.903	1.468
DOE-32	0.641	0.922	0.578	0.797	0.174	0.047	1.223	0.896
DOE-33	0.891	0.172	0.828	0.047	0.095	0.045	1.214	0.552
DOE-34	0.078	0.234	0.266	0.984	0.272	0.036	2.164	0.947

Once the Gaussian process models are constructed and validated for each metric, the following two tasks are performed

1. sensitivity indices and main effects (see section 3.2.3.3) of each variable on each of the merit functions are calculated in order to understand the relationship between the merit functions and parameters representing stent geometry.

2. a Fortran feasible sequential quadratic programming (see section 3.4.6 for an introduction to SQP) search (FFSQP) [173], is made over the surrogates for the three constrained optimisation problems represented by equation 7.13.

7.3 Results

7.3.1 GP model validation

Based on the objective function values of the 34-point sample data (listed in table 7.2), Gaussian Process models for each of the objectives are constructed. To assess the validity of the GP model, the leave-one-out process is adopted (see section 3.2.3.2). In this method, one point is left out for calculating the correlation matrix in the Gaussian Process model, thereby excluding its effect, and a prediction for the function value at that point is made. This process is repeated for all the points. A plot between the actual values and the predicted values, if showing a linear behavior with a slope of 45 degrees, suggests a good model. The leave-one-out plots are shown in Fig. 7.7. These plots suggest that the Gaussian process models predict function values close to the actual values. These plots also suggest that the predictor for FM is the relatively most inaccurate of all the four GP predictors. Another method to assess the validity of a Gaussian Process model is to evaluate the *standardized cross validated residual (SCVR)* values in the leave-one-out process (see section 3.2.3.2). The SCVR value at a point i denotes the number of standard errors by which the predicted and actual values for the left out point differ. The Gaussian Process model predicts, with approximately 99.7% confidence, that the values lie within the mean prediction plus or minus three standard errors. Thus, if the SCVR lies in the interval $[-3, +3]$, the Gaussian Process model is appropriate. Figure 7.8 shows the SCVRs for the four Gaussian Process models for each point, and all points lie within the interval $[-3, +3]$, thereby asserting the validity of these models.

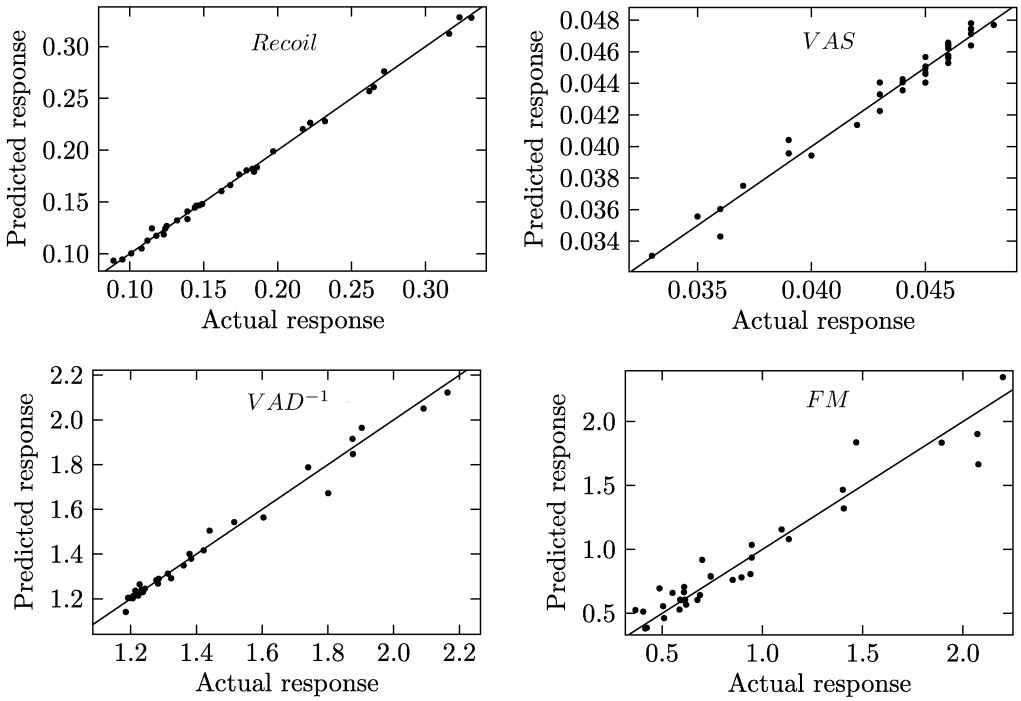


Figure 7.7: Surrogate model validation: Leave-one-out plots for recoil (*Reoil*), volume average stress (*VAS*), volume average drug (*VAD*), and the flexibility metric (FM)

7.3.2 Response surfaces and errors

Having constructed the GP predictors for the four metrics, the prediction values can be plotted over the search domain. Figures 7.9 – 7.12 show the tile plots for the four merit functions

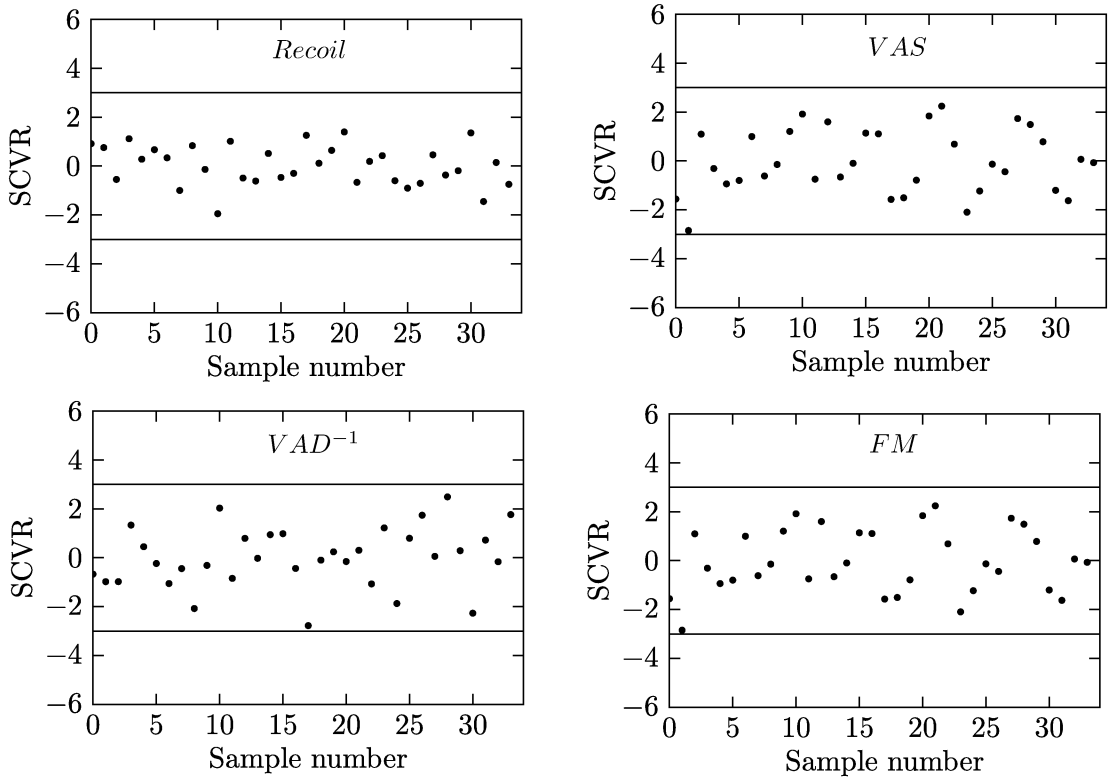


Figure 7.8: Standardised cross validated residual (SCVR) values for all points analysed – recoil (*Recoil*), volume average stress (*VAS*), volume average drug (*VAD*), and the flexibility metric (*FM*)

Recoil, *VAS*, VAD^{-1}, and *FM*, respectively. These tile plots show the four dimensional function in terms of various slices through the $W_{strut} - h_c$ planes; each subplot represents W_{strut} on the x-axis and h_c on the y-axis, parameter p_1 increases horizontally from left to right (while p_2 remains constant in each row), and parameter p_2 increases vertically downwards (while parameter p_2 remains constant in each column). Figures 7.13 – 7.16 show the error plots, as predicted by the GP model, for the merit functions *Recoil*, *VAS*, VAD^{-1}, and *FM*, respectively in a similar fashion. These plots are highly informative: for example, Figure 7.13 shows that except in the small red regions (top regions of each plot), where the predicted errors are quite high, in most of the search domain, the errors are of the order of 10^{-3} mm. Observing the column for *Recoil* (column 5) in table 7.2, one can see that the differences in *Recoil* values for the designs are of the order of 10^{-2} mm. This implies that the GP predictor can be trusted in most of the search domain. Similar arguments hold for the GP predictors of *VAS* and VAD^{-1}. Finally, as observed in previous section from the leave-one-out plots, the relatively high inaccuracy of the *FM* predictor is also apparent from Figure 7.16. In particular high uncertainty is observed for extreme values of the parameters p_1 and p_2.

The accuracy of these models can be improved by using more points to create the GP predictors. A naive approach is to use a larger space filling sample plan. Other more efficient ways include searching the error function to find points where the error is maximised and appending these points to the training set. Another approach is to use the expected improvement criterion proposed by Jones et. al. [50] that balances efforts to both improve the minimum function value and error simultaneously [50]. However, owing to the high computational expense of running more high fidelity simulations, this is avoided in this chapter. The

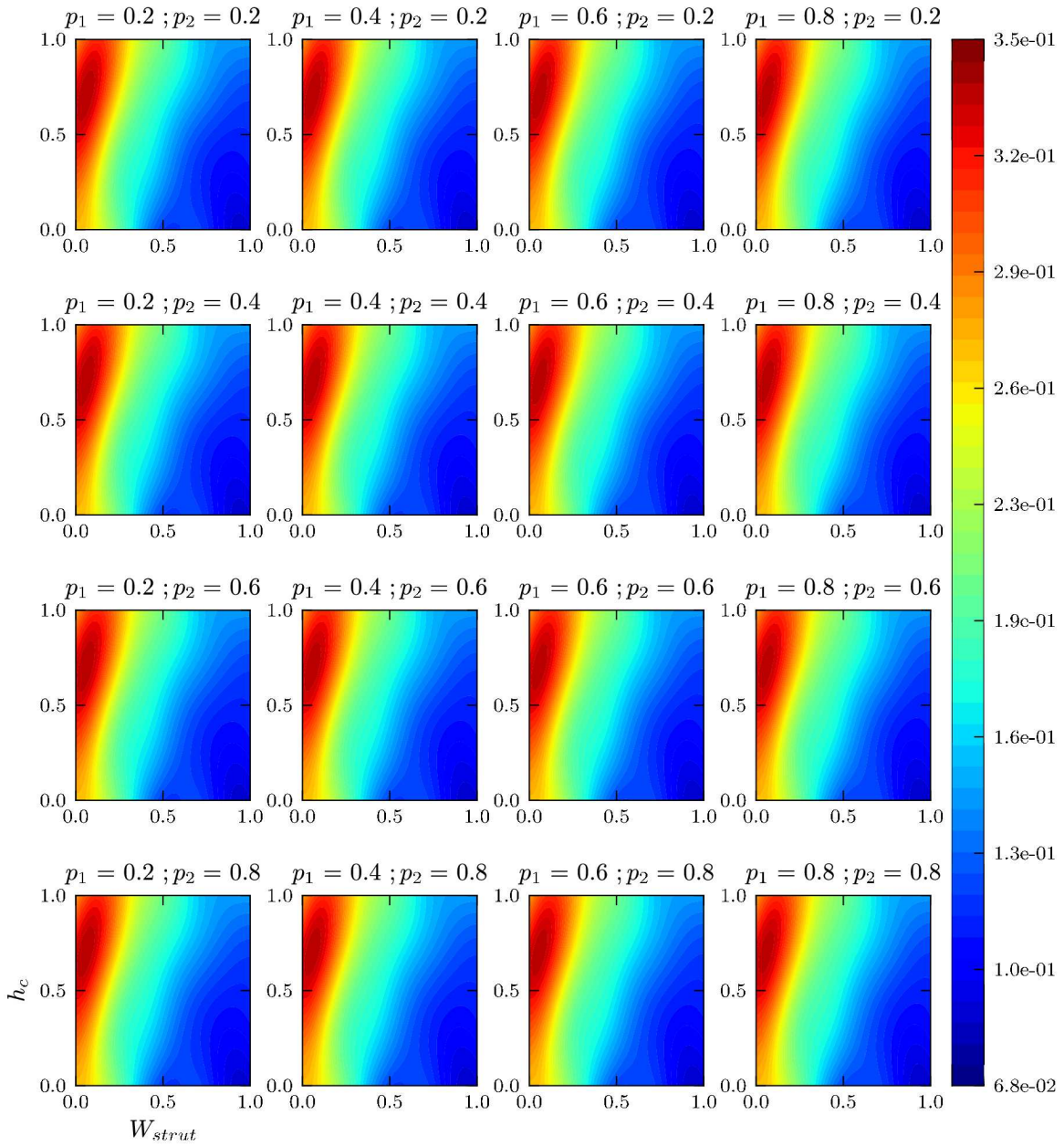


Figure 7.9: Response contours for *Recoil* as predicted by GP models: each subplot, with W_{strut} on x-axis and h_c on y-axis, shows function value contours at fixed p_1 and p_2; p_1 varies horizontally from left to right, and p_2 varies vertically from top to bottom

current predictors are searched using an SQP method and the results obtained are analysed using the high fidelity simulations. These results are presented in section 7.3.4. Returning to figures 7.9 – 7.12, some key observations are made. Figure 7.9 suggests that varying the parameters p_1 and p_2 has negligible effect on *Recoil* since all the slices in this plot are very similar. This figure also suggests that higher values of W_{strut} and lower values of h_c yield best result in terms of *Recoil*. Similarly, Figure 7.10 suggests that *VAS* too is not influenced by changes in p_1 and p_2, and that lower values of W_{strut} and higher values of h_c are preferred in terms of *VAS*. This is an apparent conflict between *Recoil* and *VAS*, which is quite evident by observing the blue and red regions of Figures 7.9 and 7.10. Figure 7.11 suggests that the parameter p_2 does not influence VAD^{-1} but unlike *Recoil* and *VAS*, p_1 mildly effects VAD^{-1}. This figure also suggests that higher values of W_{strut}, higher values of h_c, and higher values of p_1 produce best stent designs in terms of VAD^{-1}. Finally, figure 7.12 shows that both while

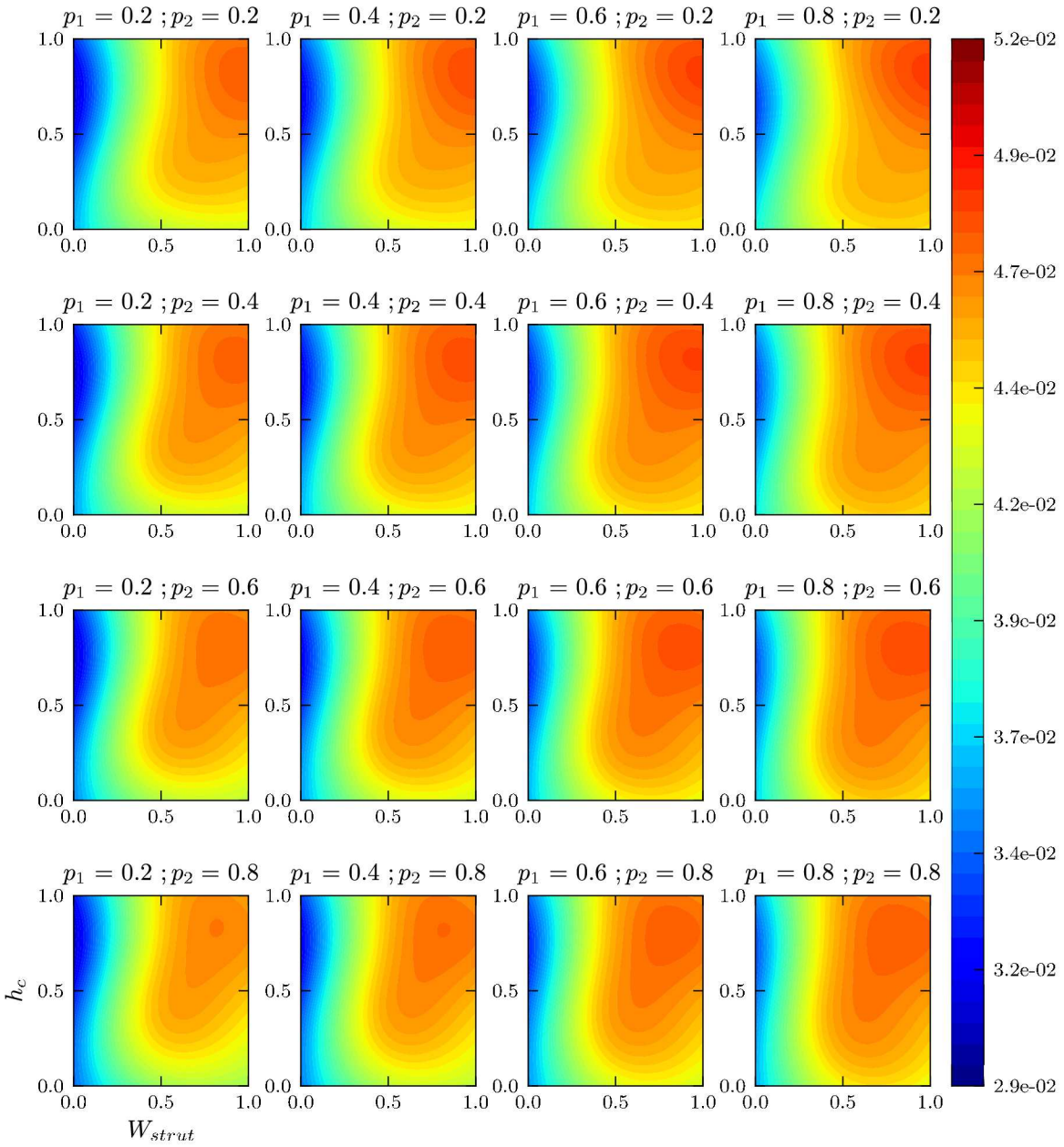


Figure 7.10: Response contours for *VAS* as predicted by GP models: each subplot, with W_{strut} on x-axis and h_c on y-axis, shows function value contours at fixed p_1 and p_2; p_1 varies horizontally from left to right, and p_2 varies vertically from top to bottom

W_{strut} and h_c do not influence FM significantly, FM is most sensitive to the parameter p_1 (higher values preferred in terms of FM followed by W_{strut} and p_2.

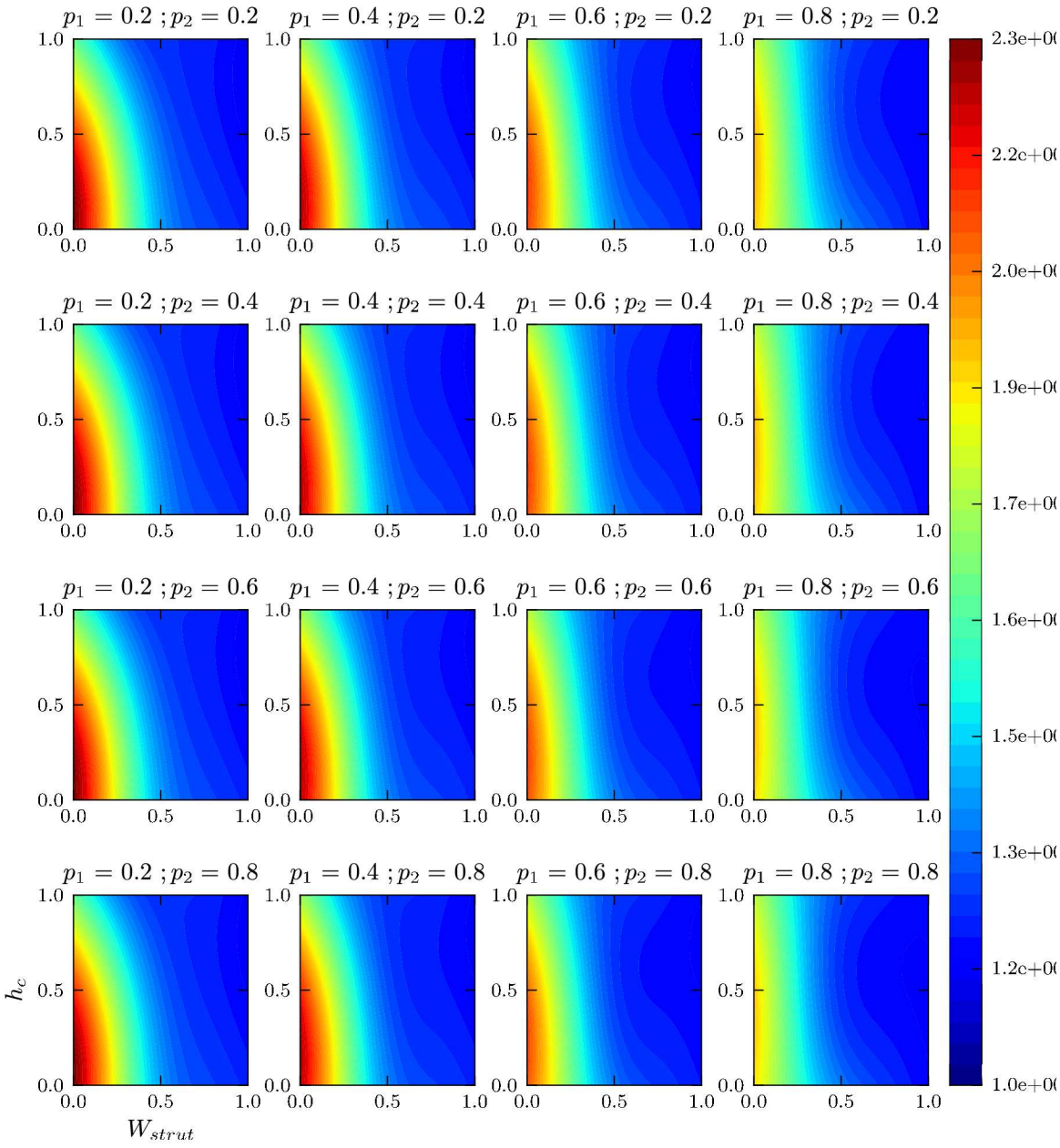


Figure 7.11: Response contours for VAD^{-1} as predicted by GP models: each subplot, with W_{strut} on x-axis and h_c on y-axis, shows function value contours at fixed p_1 and p_2; p_1 varies horizontally from left to right, and p_2 varies vertically from top to bottom

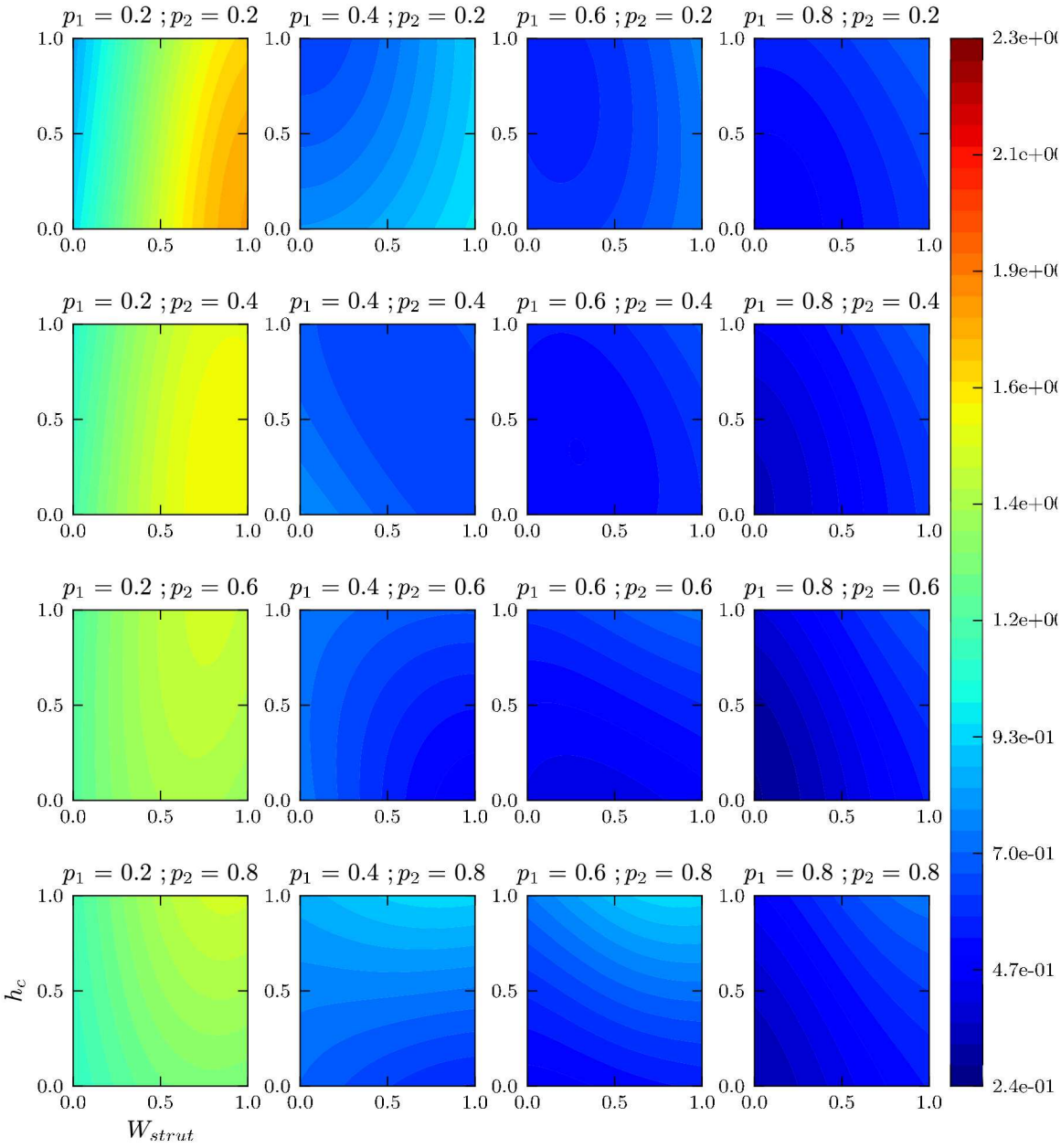


Figure 7.12: Response contours for FM as predicted by GP models: each subplot, with W_{strut} on x-axis and h_c on y-axis, shows function value contours at fixed p_1 and p_2; p_1 varies horizontally from left to right, and p_2 varies vertically from top to bottom

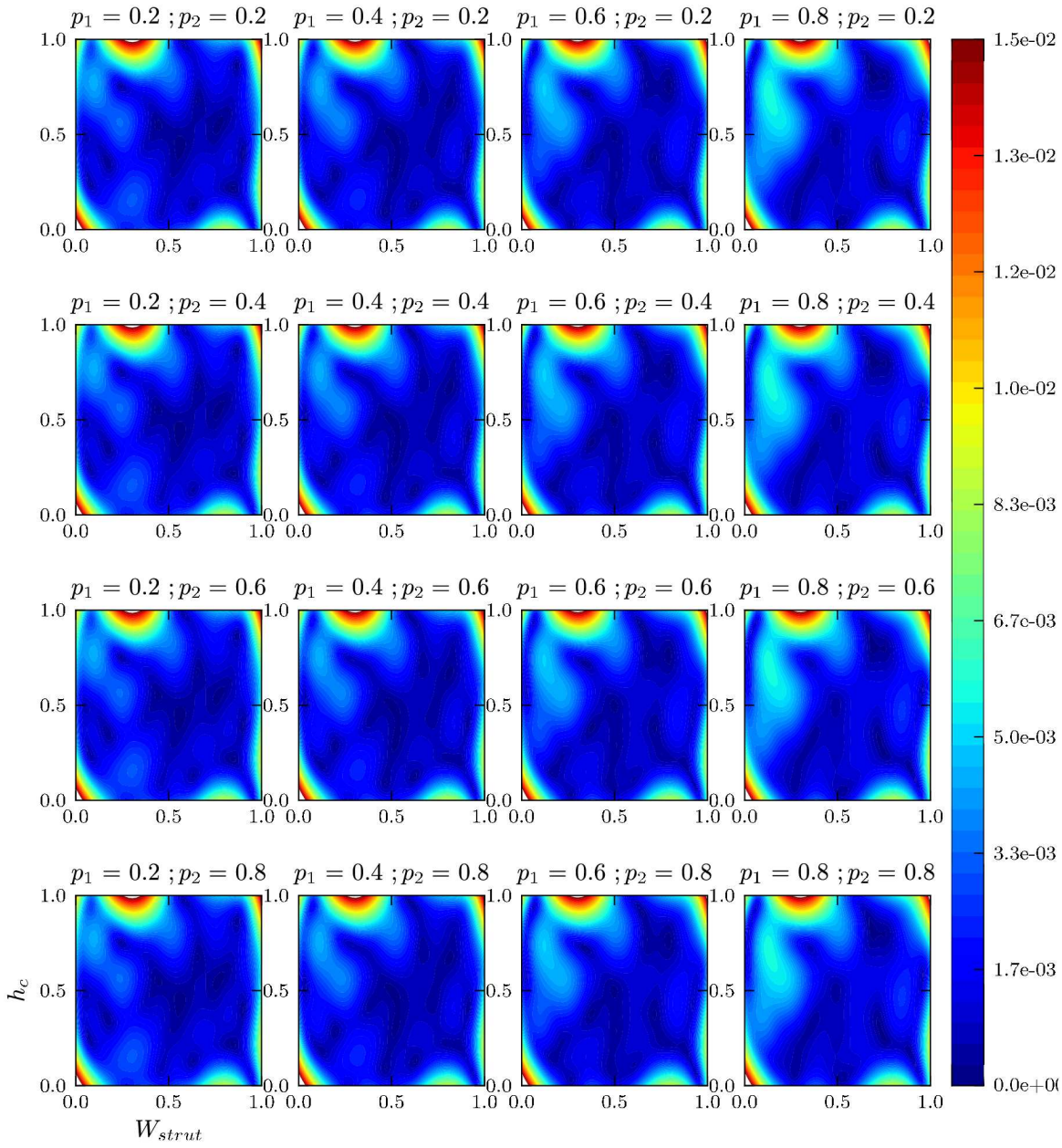


Figure 7.13: Error contours for *Recoil* as predicted by GP models: each subplot, with W_{strut} on x-axis and h_c on y-axis, shows the error in prediction contours at fixed p_1 and p_2; p_1 varies horizontally from left to right, and p_2 varies vertically from top to bottom

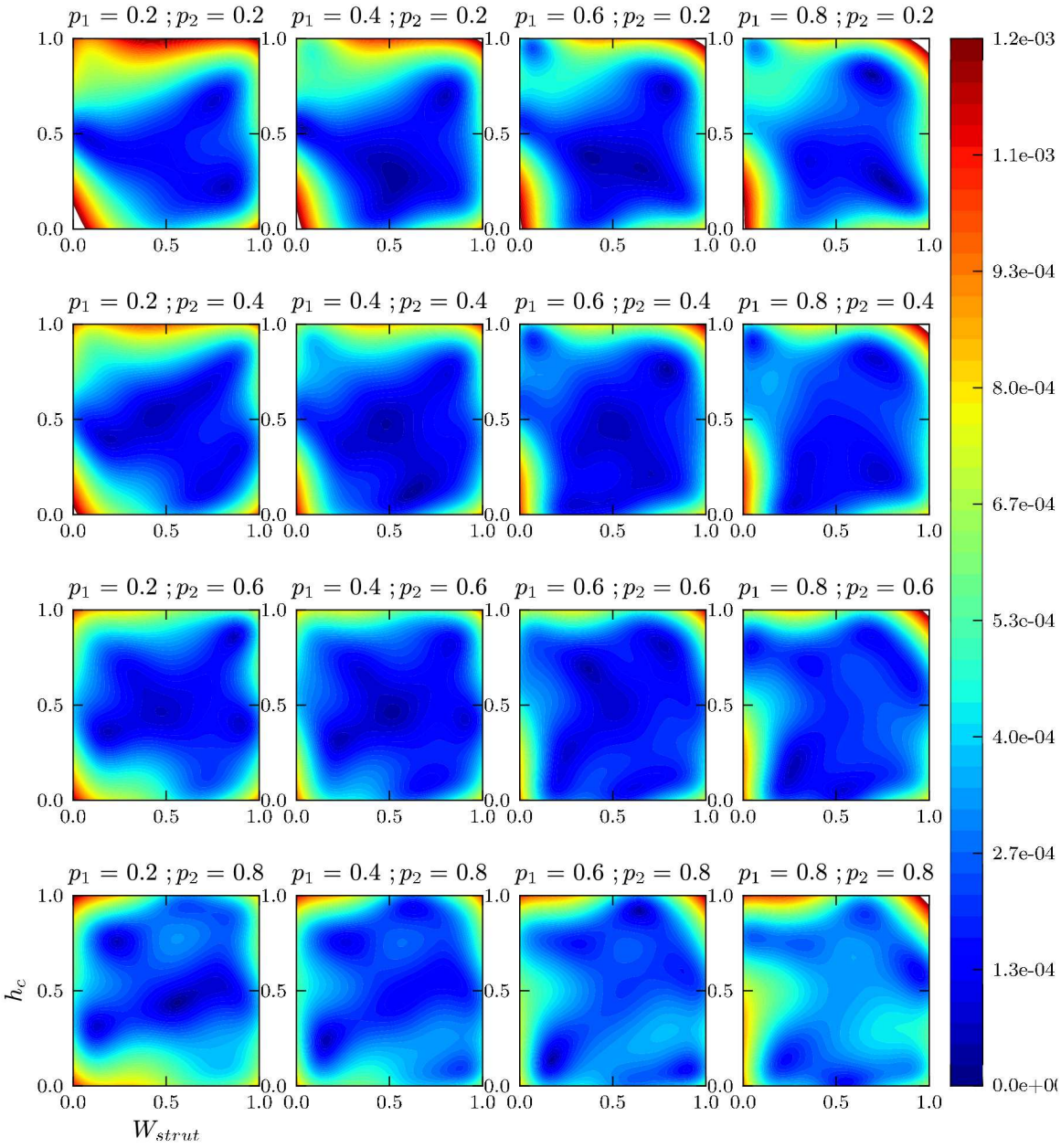


Figure 7.14: Error contours for *VAS* as predicted by GP models: each subplot, with W_{strut} on x-axis and h_c on y-axis, shows error in prediction contours at fixed p_1 and p_2; p_1 varies horizontally from left to right, and p_2 varies vertically from top to bottom

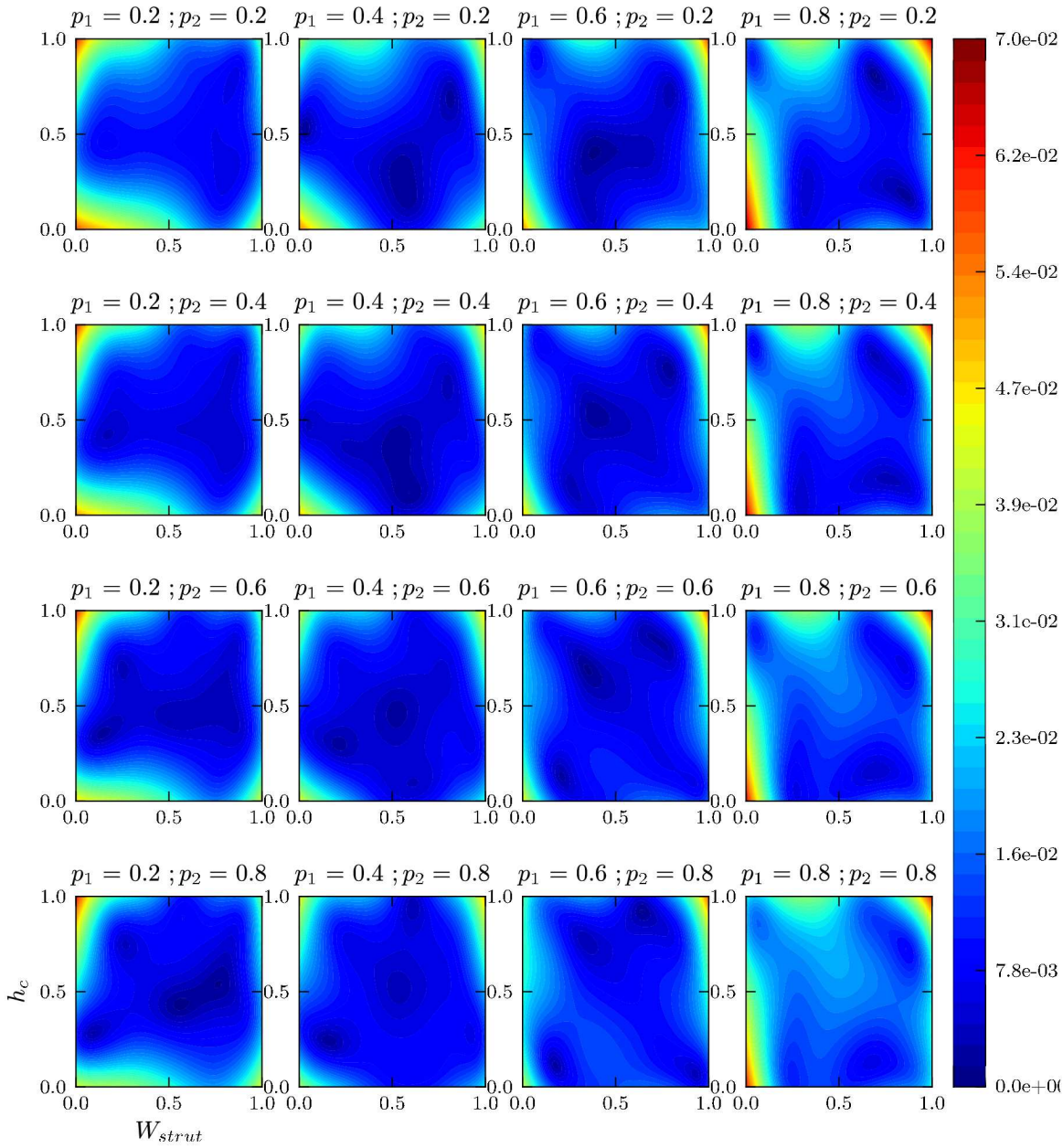


Figure 7.15: Error contours for VAD^{-1} as predicted by GP models: each subplot, with W_{strut} on x-axis and h_c on y-axis, shows the error in prediction contours at fixed p_1 and p_2; p_1 varies horizontally from left to right, and p_2 varies vertically from top to bottom

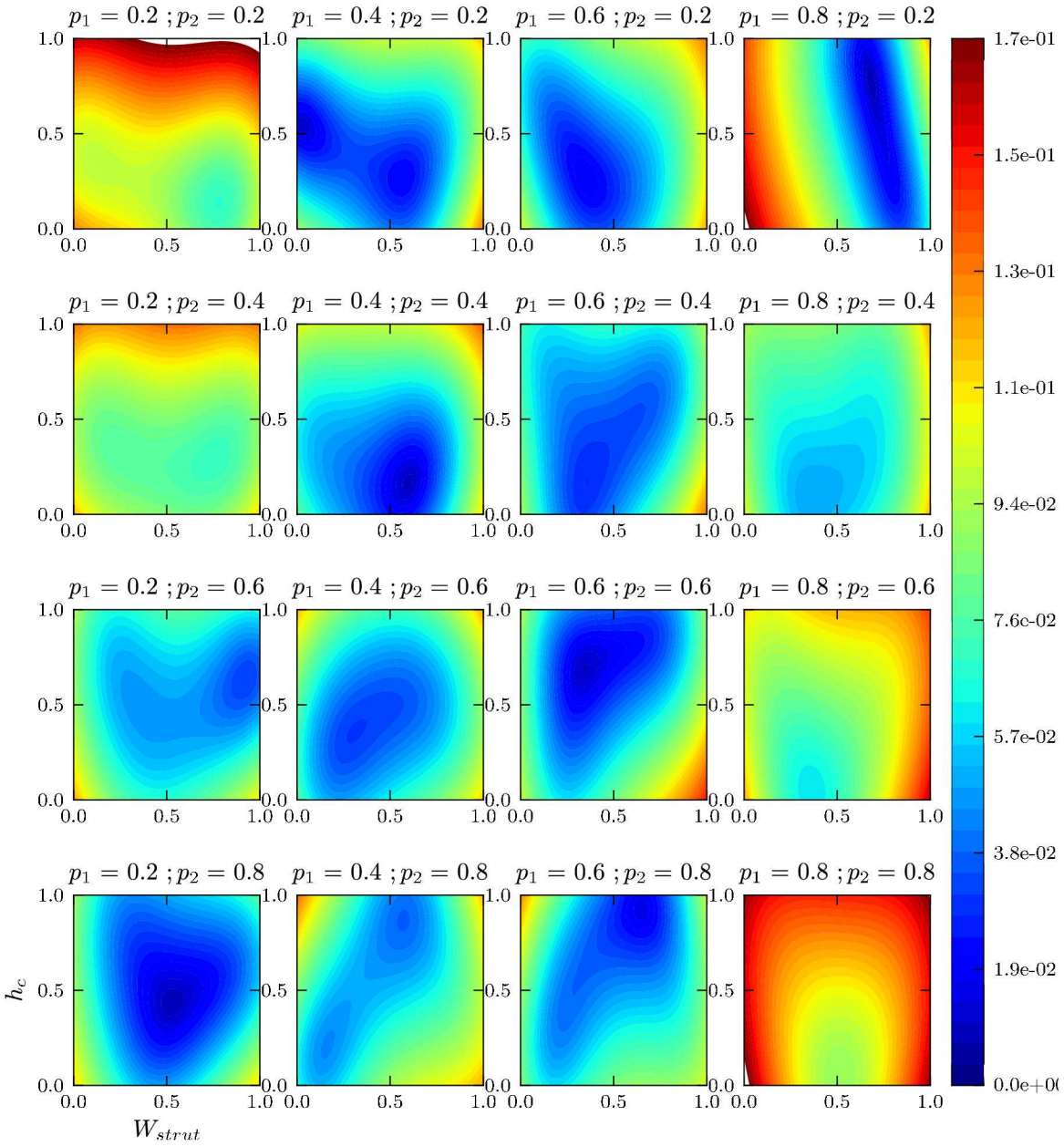


Figure 7.16: Error contours for FM as predicted by GP models: each subplot, with W_{strut} on x-axis and h_c on y-axis, shows the error in prediction contours at fixed p_1 and p_2; p_1 varies horizontally from left to right, and p_2 varies vertically from top to bottom

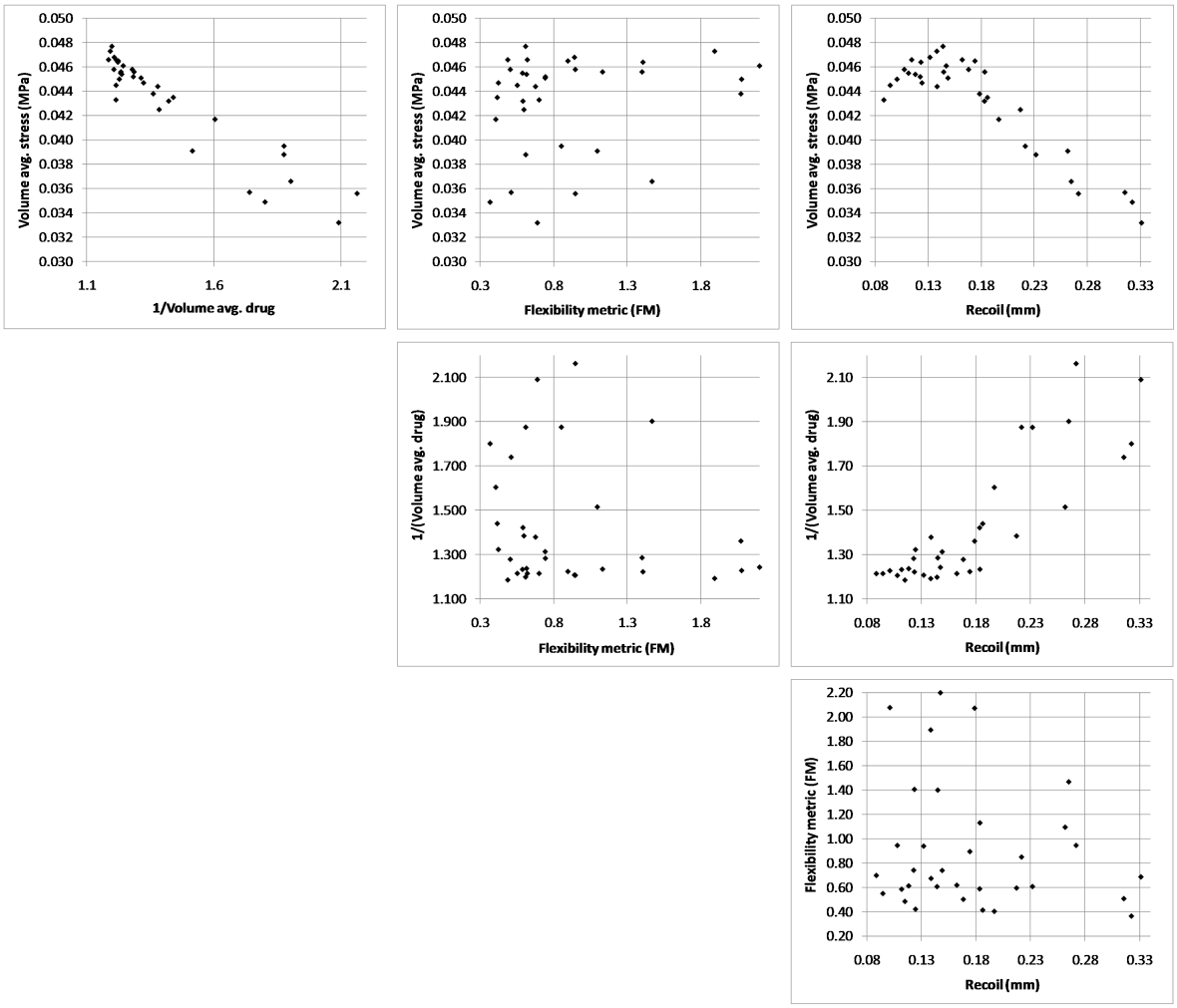


Figure 7.17: Trade-off plots for all combinations of the four metrics

7.3.3 Objective function trade-offs

Figure 7.17 plots the four metrics against each other for the 34-point sample, thereby showing the relationship between the objective functions. Across all these plots better designs tend to lie towards the bottom left corner. The centre-right plot shows that volume average drug and recoil are not in conflict with each other – a lower recoil generally implies a higher volume average drug. The top-left and the top-right plots suggest that volume average stress is in conflict with both volume average drug and recoil – a lower volume average stress would generally imply a higher volume average drug and higher recoil. The rest of the plots where flexibility is one of the variables imply that flexibility is not highly correlated with any of the other metrics. To locate the features in the parameterisation to explain the observations of these plots one can use the information provided by the surrogate models (Figures 7.9 – 7.12). Moreover, main-effects calculated from the GP predictors can be used to aid this understanding. Hence, the sensitivity indices of the main effects (see section 3.2.3.3) for each objective, which evaluate the relative effects of the individual parameters on the GP predictor, are calculated. Representative pie charts for the main effects are plotted in Fig. 7.18. These provide a graphic indication of the parameters that are most influential in determining the physical response in terms of the four evaluation metrics. This figure reinforces the observations made in the previous section, and readily explains the behaviour of flexibility. While the parameter p_1, the peak height of the links, affects flexibility most, it

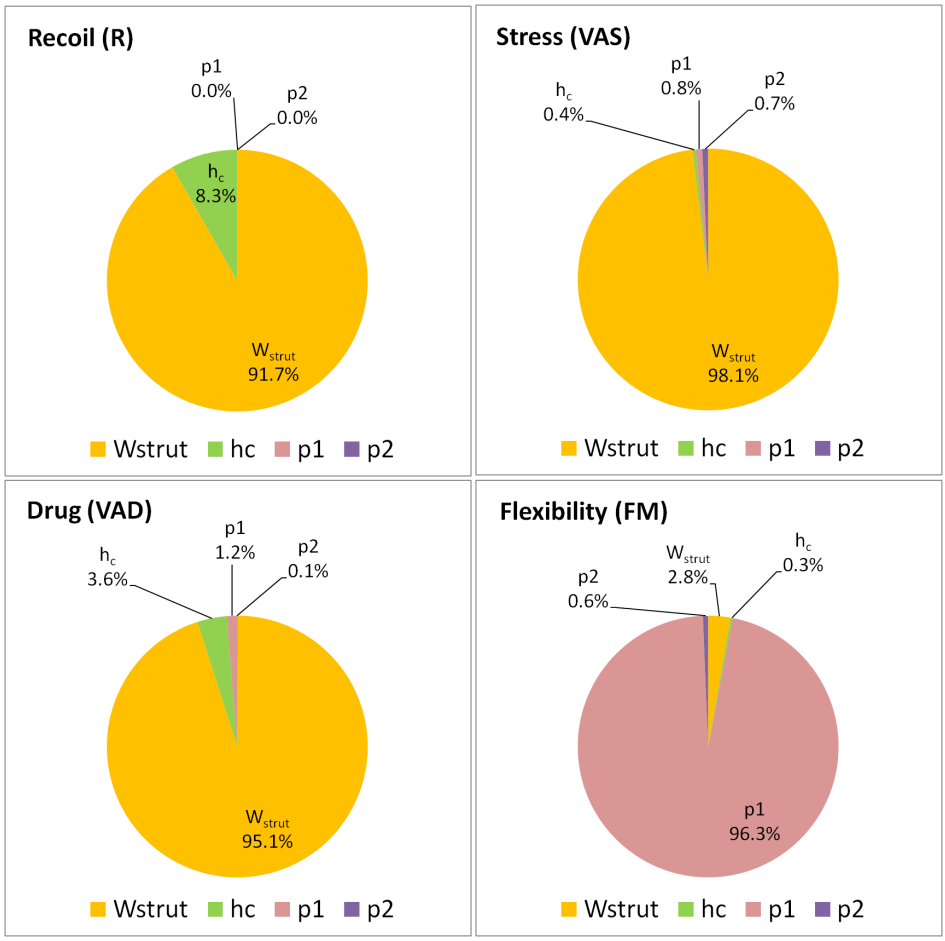


Figure 7.18: Sensitivity indices for each of the four metrics

has minimal effect on the other objectives. This explains the independence of flexibility with other objectives as varying p_1 could improve the flexibility without having much effect on the other metrics. The effect of p_1 seems most appropriate if one observes the bottom pictures of Fig. C.4 in the appendix. A higher peak value of p_1 gives more length to the curved link which helps it expand as well as contract. A straight link, $p_1 = 0$, would not have such a designed feature for expansion/contraction and consequently will not be as flexible as with a finite p_1.

The main effects and GP predictor plots also explain the conflict between volume average stress with drug and recoil observed in Fig. 7.17. The strut width, W_{strut}, is the parameter influencing all three of these objectives most. Intuitively, a higher W_{strut} implies more metal in the stent, leading to higher contact area between the stent and the tissue, and consequently higher stresses. This simple generalisation that more metal (in terms of higher W_{strut}) in the stent leads to higher stresses needs more explanation, even though it is quite clear from figure 7.10 that this is the case. One would think that for the same inward force by artery, trying to regain its original configuration, a thinner strut will lead to a higher contact pressure owing to a low contact area. This is apparently against the observations of figure 7.10 where clearly higher W_{strut} leads to higher stresses. The explanation of this counter-intuitive phenomenon lies in figure 7.9, where it can be seen the *Recoil* is also influenced by W_{strut}. High W_{strut} values lead to lower recoil, which implies that for these stents the artery is more stretched (higher expanded diameter) when compared to the stent designs with lower W_{strut}. Consequently, for stent designs with higher W_{strut}, there is more force imparted by the artery on to the stent

relative to the designs with lower W_{strut}, and hence explains why higher W_{strut} values lead to higher stresses. Such high W_{strut} values, though unwanted from the point of view of stresses, seems favourable for the drug as a higher W_{strut} leads to higher drug availability, and a higher contact area implies more drug transport from the stent into the tissue. It could be argued, at this point, that the uniformity of the drug distribution is as important, as the amount of drug in the tissue. An easy way to look at the uniformity of the drug, and easily implemented in discretised domains, is to calculate the mean value of the drug concentration in all the elements and then look at the standard deviation of the data. This value effectively implies the root mean square error of the difference between each element's drug concentration and the overall mean. It is found that for the parameterisation and modelling techniques used in this chapter, the volume average drug is negatively correlated to the standard deviation, implying that a higher amount of drug is also reflective of a relatively lesser uniform distribution around its mean. This standard deviation of drug concentration can be used as one of the objective functions in design studies. It is important that such a metric for uniformity must be used in conjunction with the average drug concentration, as a device allowing for uniform distribution could have a low mean. This would imply uniform distribution of non-therapeutic levels for the drug. Such a measure of uniformity is avoided in this chapter to keep the number of evaluation metrics low.

An explanation of the apparent correlation between the inverse of volume average drug and recoil is now sought. Both these quantities are most affected by W_{strut} followed by h_c. For the same expansion radius, thicker struts undergo more plastic deformation at the curved ends of the circumferential rings (all other parameters being equal) and consequently reduce recoil. This partly explains why volume average drug inverse and recoil seem to be correlated. The role of h_c in recoil, with a relatively large main-effect value of 8.3%, can be explained by consideration of Fig. 7.19. The expansion of a stent can be seen as stretching of the circumferential rings in a flat plane. For the same stretch, d, circumferential rings with smaller h_c undergo higher plastic deformation at the curved regions. A higher plastic deformation consequently results in a lower recoil. The effect of h_c can also be understood for volume average drug and stresses. A smaller value of h_c implies a smaller length of the circumferential rings. Such designs have a relatively lower surface area and hence lesser contact area between the stent and the plaque. This affects both volume average drug (negatively) and stresses (positively).

Table 7.3: Results for constrained optimisation (Normalized design parameters)

	W_{strut}	h_c	p_1	p_2	*Recoil*	*VAS*	VAD^{-1}	*FM*
	mm	mm	mm	-	mm	MPa	unit^{-1}	N-rad
Baseline	**0.5000**	**0.5000**	**0.5000**	**0.5000**	**0.1685**	**0.0458**	**1.2783**	**0.5045**
min (*VAS*)	1.0000	0.0001	0.9599	0.8508	0.1685	0.0432	1.2783	0.5045
min (VAD^{-1})	0.9402	0.3993	0.9316	0.8712	0.1685	0.0458	1.1760	0.5045
min (*FM*)	0.5063	0.2433	0.8658	0.6850	0.1685	0.0458	1.2783	0.4138

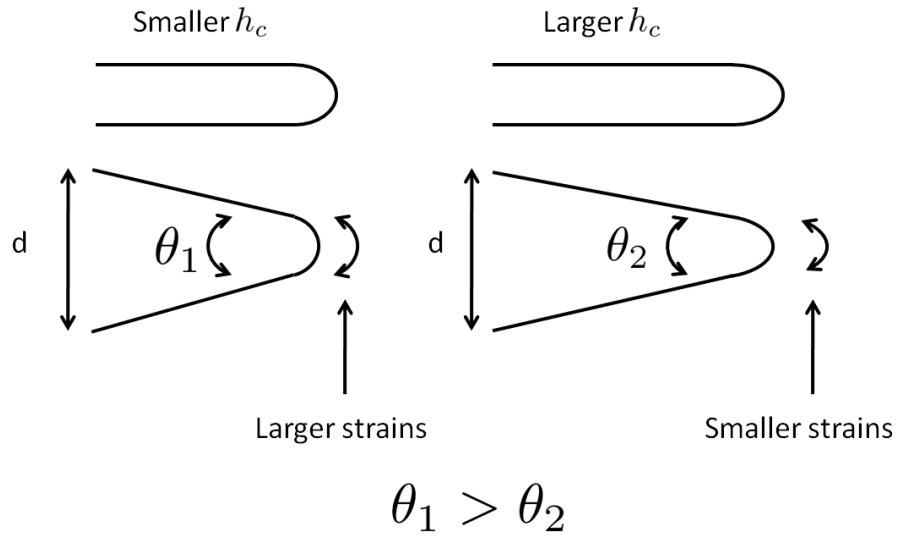


Figure 7.19: Effect of h_c on plastic strains

7.3.4 Optimisation results

After the construction of GP predictors for each metric, a Fortran feasible sequential quadratic programming search (FFSQP) [173], is made over the surrogates for the constrained optimisation problems. The results for the three different sets of optimisation problems are tabulated in Table 7.3. As discussed in the above paragraph, both W_{strut} and h_c affect the metrics for stress, drug, and recoil. It is primarily the interplay between these two parameters which results in the trade-off between various metrics and hence influence the results for the optimisation studies.

For the case of stress minimisation, the optimal designs have a high value of W_{strut} but low values of h_c, both being pushed towards their respective bounds. A high value of W_{strut} to minimise the stresses appears to be paradoxical but can be partly explained when combined with a low value of h_c. It is worth recalling that starting from the baseline geometry, design improvement is being sought which minimises stresses without compromising on the other three metrics. Given the broad relationship between the parameters and the metrics, as discussed in the previous paragraph, the first thought should be to decrease W_{strut} to decrease stresses. However, decreasing W_{strut} comes at a price – any decrease in the value of W_{strut} can increase both the recoil and the volume average drug relative to the baseline geometry. It is here that the interplay between all the parameters has to be considered. A high value of W_{strut}, as suggested by the optimal solution, satisfies the constraints imposed on the recoil and drug metric, and a design improvement in the stress metric is obtained by decreasing h_c. Minimising h_c leads to shorter circumferential rings, which, even with higher W_{strut} could have a lower surface area, which minimises stresses. Furthermore, a high value of p_1 balances the decrease in flexibility (relative to the baseline geometry) due to increased W_{strut}. A high value of p_1 also ensures better artery wall coverage to improve drug delivery. The result of the optimisation algorithm is confirmed by performing verification simulations on the optimal design predicted. For the solution of minimising the stress, the FEA analysis gives a value of 0.0421 MPa, the predicted value being 0.0432 MPa (error $\approx$ 2.5 %) for the optimized solution. The resulting improvement relative to the baseline geometry being 8%. Note that the optimal solutions suggested by the optimisation algorithm push the constraints to zero for

all the three optimisation problems, implying that the optimal solution lies on the boundary of the constraint functions. This constraint value of zero, however, is based on the GP predictors. FEA verification analyses, for the problem of minimising stress gives errors of 0.7%, 1.3%, and 3.8%, respectively, for *Recoil*, VAD^{-1}, and FM between the predicted and actual values.

Similar arguments as in the above paragraph hold for the results of maximising the drug metric. An increase in W_{strut} is desirable as it allows for more contact area which results in higher transfer of drug into the tissue. However, the increase in the stress values due to increase in W_{strut} must be balanced by a decrease in h_c in order to satisfy the specified constraint on the stress metric. Moreover, the decreased flexibility due to higher W_{strut} is balanced by high values of p_1. FEA verification for the optimal design for maximizing drug metric gives a value of 1.19 mm^3/unit for VAD^{-1} (the predicted value being 1.176 mm^3/unit; error $\approx$ 1%). The resulting improvement relative to the baseline geometry is approximately 7%. FEA verification analyses, for this problem gives errors of 1.2%, 2.5%, and 4.1%, respectively, for *Recoil*, VAS, and FM between the predicted and actual values.

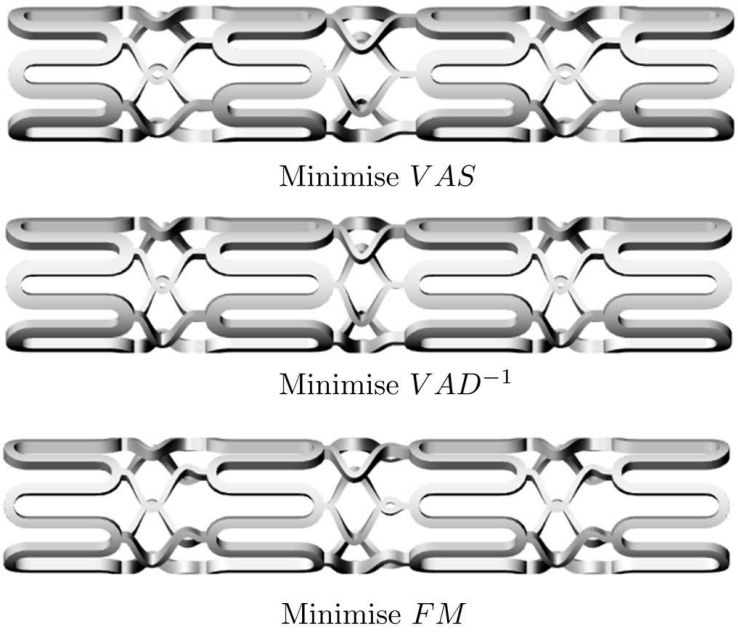


Figure 7.20: Optimal designs from single-objective constrained optimisation

The problem of minimising the flexibility metric does not face such serious issues as W_{strut} is not the parameter having most effect. The solution methodology in this case seems to be simpler – since the value of the flexibility metric is most dependent on p_1, and the fact that p_1 does not significantly affect the other metrics, the optimal solution has a high value of p_1 without much change in the value of W_{strut} from the baseline geometry. Although W_{strut} has an effect on flexibility, with lower strut width stents being more flexible, the optimised solution does not show a decrease in its value as that would reduce *VAD* when compared to the baseline geometry. Figure 7.20 shows the optimal stent geometries as suggested by FFSQP when run over the constructed Gaussian process models. It is worth noting that the optimal stents have the links curved only once, unlike the baseline geometry whose links are curved twice. This can be explained by the fact that a flexible stent design is more likely to avoid the self contact in the links in the compression part (c.f. Fig. C.4: bottom-left image). If the

links cross the medial line again, then there is a higher likelihood of self contact in the link. A self contact would result in increased moment requirement for high curvatures implying reduced flexibility. The validation FEA run on the optimized solution for minimising the flexibility metric gives an FM value of 0.4332 N-rad (the predicted value being 0.4138 N-rad; error $\approx$ 4.5%) which is an improvement of approximately 14% over the baseline geometry. FEA verification analyses, for this problem gives errors of 1.0%, 2.3%, and 1.6%, respectively, for $Recoil$, VAS, and VAD^{-1} between the predicted and actual values.

It should be noted that the accuracy of GP models can be further enhanced by running high fidelity simulations at more points obtained by searching either the error function for the predictors or by searching the expected improvement function, see Jones et. al. [50], for each objective. If such a procedure is adopted then there is a possibility that further improvement in designs can be obtained. However, this comes at the added expense of performing more high-fidelity analysis, practical issues of geometry construction, geometry import and export from CAD packages to FEA packages, automated meshing, and various peculiarities of each analysis package used in the process. Nonetheless, improving the accuracy of the surrogate models can help better understand the relationships between the design parameters and the objective functions, and forms an obvious extension for future work.

7.4 Conclusions

This chapter proposes a parameterisation technique for stent designs with circumferential rings and links. A 34-point sample using the LP_τ sampling plan is constructed, and the resulting designs using the proposed parameterisation are evaluated for four performance metrics. Thereafter, Gaussian process models are constructed to approximate the response for each evaluation metric. The relative effects of each of the parameters on individual performance metrics are quantified. The results of single objective constrained optimization show significant improvement in the metrics relative to the baseline geometry demonstrating the applicability and suitability of the Gaussian process modelling to approximate the physical response by efficient sampling at only a small number of points. The conflict between various metrics and the parameters to achieve improvement in a particular metric while not altering the others is shown and discussed. For the parameterisation technique used in this chapter, the optimal solutions have a higher strut width compared to the baseline geometry, but a lower value of circumferential ring length. Although the effect of parameters such as strut thickness in the performance of stents is very well known and researched, both experimentally and numerically, the effect of other parameters, such as W_{strut}, h_c, and p_1, are shown here to have significantly influenced stent performance. The strut width, W_{strut} is shown to be a key determinant of recoil, stresses, and amount of drug delivered. The plastic deformation produced by smaller h_c, in the obtained optimal designs, implies that a deviation from the circumferential ring type structure for new stents, which allows for designed plastic deformation at specific locations, could potentially improve the stent design. This conclusion can be seen independently, i.e. without any influence from this study, implemented in the geometric design for Boston Scientific's latest ELEMENT stent series, as shown in Figure 7.21. The design of the ELEMENT series platform has been modified to specifically introduce higher plastic strains in the curved segments of the stent circumferential rings [7]. Similarly, a lower value for stent segments, i.e. smaller value of the parameter h_c used in this study, is preferred

in the design of the ELEMENT stent series [7]. This study also concludes that the parameters defining the peak cross length of the links, p_1 in the current parameterisation, can be manipulated to produce flexible stents without affecting other metrics such as volume average stress and drug. Finally, the results suggest that minimising the number of curved regions in the link design, in order to delay self-contact, can be beneficial in terms of flexibility, stresses, and drug delivery. In the ELEMENT stent series, see Figure 7.21, the self contact is avoided by having a nested peak structure.

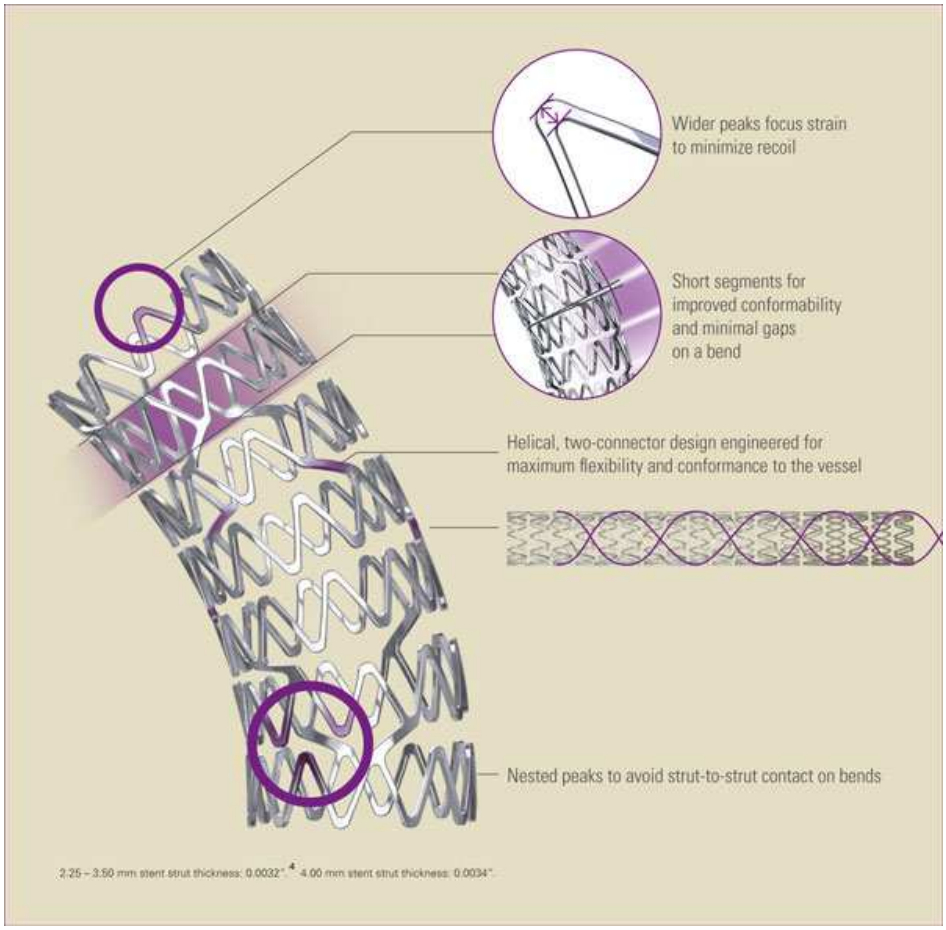


Figure 7.21: Boston scientific's geometric platform for the latest ELEMENT stent series [7]

Having performed a preliminary optimisation study in this chapter, the next chapter focuses on the development of CFD drug-distribution models that account for haemodynamic flow in the lumen. Subsequently, chapter 9 uses all the metrics used in this chapter, along with the drug uniformity metric defined in the next chapter, for a multiobjective design study. A constrained optimisation formulation, as the one used in this chapter, is useful in two scenarios: first, if the constraint limits are known; and second, if a good design is available which performs well on all aspects but one, and it is required to improve its performance in this one aspect without deteriorating others. The former scenario stems from the thought that for many objectives, although hard to estimate, minimum thresholds could be available in future. For example it was already discussed how some clinical trials have suggested a minimum stent area threshold to arrest restenosis. Similarly, if the maximum curvature of the arteries is known, then the stent need not be more flexible than it is required for easy deliverability through the most difficult artery sections. Based on this data, a minimum threshold for FM could be determined. Finally, for the drug, once the biochemistry of the

drugs is fully understood, the minimum and maximum levels of drug that are allowed to be administered, depending on the toxic-to-therapeutic ratio of a particular drug, could be ascertained. These values could then be represented in the constrained optimisation problem as two constraints: one for the minimum therapeutic level and one for the maximum toxic level. On the other hand, when such thresholds are not known, but it is known, for example, that lower recoil, higher flexibility, lower injury, and higher drug, are preferred, then a multi-objective formulation of the design problem is more suitable. This is primarily because, such a formulation, focuses on minimising all objectives simultaneously, rather than focusing on only one objective at a time. Hence, a family of optimal designs, when all objectives are considered simultaneously, can be obtained using such a formulation. Such a formulation, which also results in a better understanding of the trade-offs involved in improving one objective with respect to others, forms the content of chapter 9.

Chapter 8

Drug distribution in stented vessels

This chapter concerns the final aspect of engineering analysis of stents considered here, namely drug distribution. The distribution of anti-proliferative drug plays an important role acting against restenosis. However, for an ideal stent design, an adequate and uniform drug distribution is required. If adequate drug is not delivered in the tissue, the therapeutic effects of the drug might abate; and if excessive drug is delivered then the toxic effects of the drug might aggravate. Hence, both adequacy and uniformity from a drug-distribution, which are governed by the stent design, are required. In chapter 7, a basic model for drug-release, which considered pure diffusion of the drug in tissue region and neglected luminal blood flow was developed. The development of this model was primarily governed by practical reasons of keeping all analyses in a finite element analysis package. In this chapter, a CFD model for drug-release which accounts for the haemodynamic flow in the lumen is developed. Similarly, the idealised CFD model for haemodynamic evaluation presented in chapter 4, is extended in this chapter to include post-expansion geometry of the stent-artery assembly obtained from chapter 5. The aims of this chapter are

1. to apply the model for haemodynamic evaluation developed in chapter 4 to the expanded assembly obtained from chapter 5. This provides a one way coupling between the FEA and CFD analysis, and consequently gives a more realistic evaluation of the *HLRFI* metric (defined in chapter 4, equation 4.9),

2. to develop a CFD model to evaluate drug-distribution in a stented artery, and

3. to formulate an objective function which measures the uniformity of the drug distribution in the tissue to complement the *VAD* metric developed in chapter 7.

This chapter is divided into two parts – part I deals with the results of the haemodynamic evaluation in the geometry obtained by FEA expansion, and part II details the model for the drug distribution. Part II begins with a brief introduction to the workings of the different drugs used in DES. Then, after a review of drug-distribution models proposed in the literature, the details of the drug distribution model proposed in this thesis are presented. Finally, a measure of uniformity of drug delivered in the tissue is formulated.

8.1 Part I : Haemodynamics results

The details of the haemodynamic model are provided in chapter 4. Here the results of the same model when applied to the geometry obtained through the FEA analysis of chapter 5 are presented. The deformed geometry after the unloading phase (c.f. Figure 5.15) in section 5.5.2 is first obtained and exported to the CFD package, Star-CCM+ 3.06.006, Melville, NY USA. Since only the flow volume is needed, only the geometries of plaque and the stent are exported. Figure 8.1 shows this geometry.

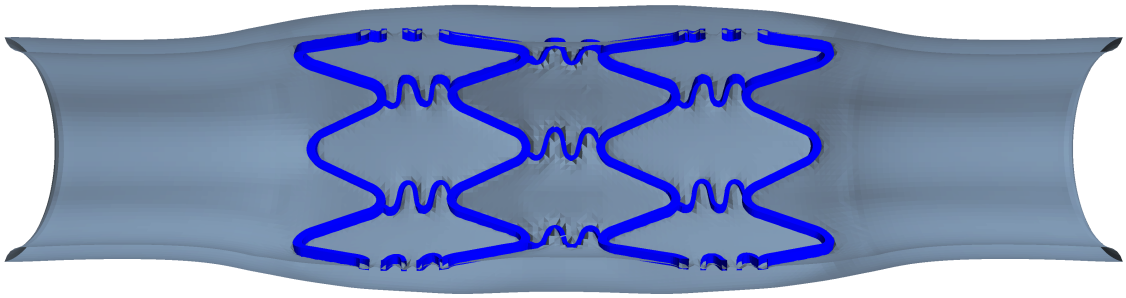


Figure 8.1: Imported geometry of plaque and stent in the CFD package

The imported geometry is meshed using polyhedral volume cells. Figure 8.2 shows this mesh. The irregularities in the central part of the mesh can be attributed to the embedding of the stent into the plaque during the expansion simulations. Since the stenosis is highest in the central region, which leads to higher contact stresses in the central region, a higher embedding of the stent into the tissue occurs. Such embedment leads to distortion of the plaque elements and can only be avoided by using a very high mesh density for the plaque during the expansion analysis. This, although possible, is avoided as it increases the computational time for the expansion analysis to impractical limits. Nonetheless, to overcome convergence issues, a fine surface mesh for the plaque is used in this chapter for CFD simulations. Figure 8.3 shows a closeup of the volume mesh (number of cells $\approx$ 1.6 million) which sufficiently resolves the irregularities and does not cause convergence issues. Pulsatile flow calculations are performed on this mesh to evaluate the wall shear stress and velocity profiles.

Figure 8.4 shows the axial wall shear stress contours on the surface of the plaque at point 5 of the cardiac pulse (c.f. figure 4.4). Figure 8.5 shows the velocity profiles on a cross section of the plaque-stent assembly. Even though the dimensions of the stents are not the same as in Chapter 4 and this chapter, one observation can be made – the recirculation zones in the geometry of this chapter are smaller. This can be visually observed in Fig. 8.5, and can be attributed to tissue prolapse. Tissue prolapse refers to the tissue volume protruding into the lumen area between the stent struts. This can be seen in Figure 8.7. The protruding tissue reduces the severity of the effective backward facing step that the flow faces. Immediately after any strut, the blood flow hits the plaque surface earlier due to this protrusion, and consequently reduces the extent of the recirculation zones. Similar to the results presented in chapter 4, the percentages of plaque area exposed to low wall shear stress (below 0.5 Pa) and reverse flow are calculated. These values are then combined using equation 4.9 to calculate *HLRFI*. The value for *HLRFI* is 27.4 %. Figure 8.6 shows the secondary recirculation of flow in a layer adjacent to the plaque wall in the lumen.

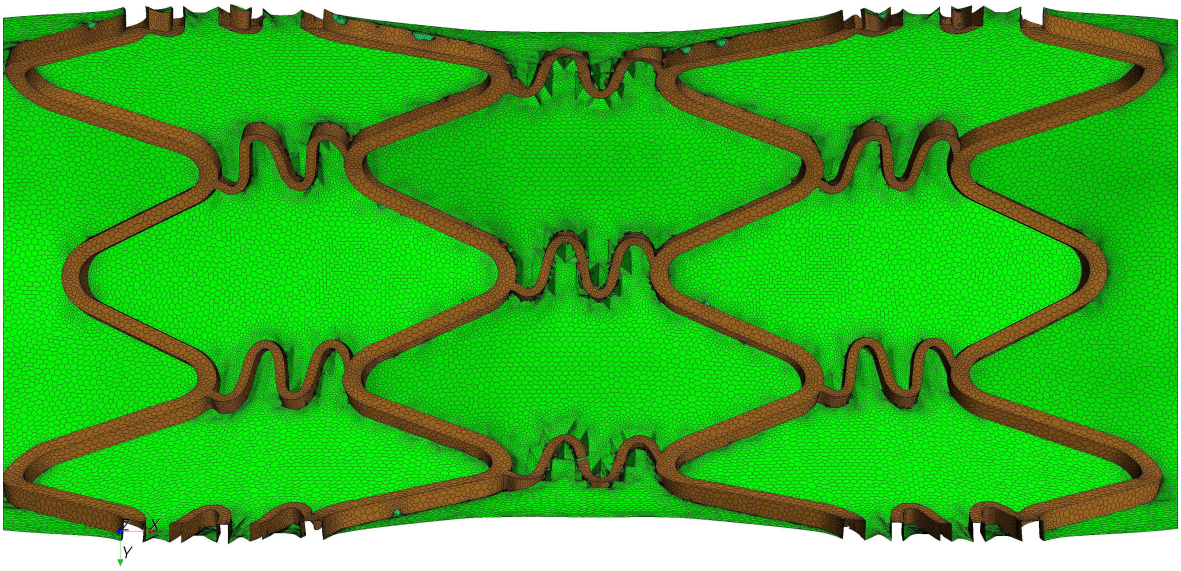


Figure 8.2: Volume mesh of the imported geometry

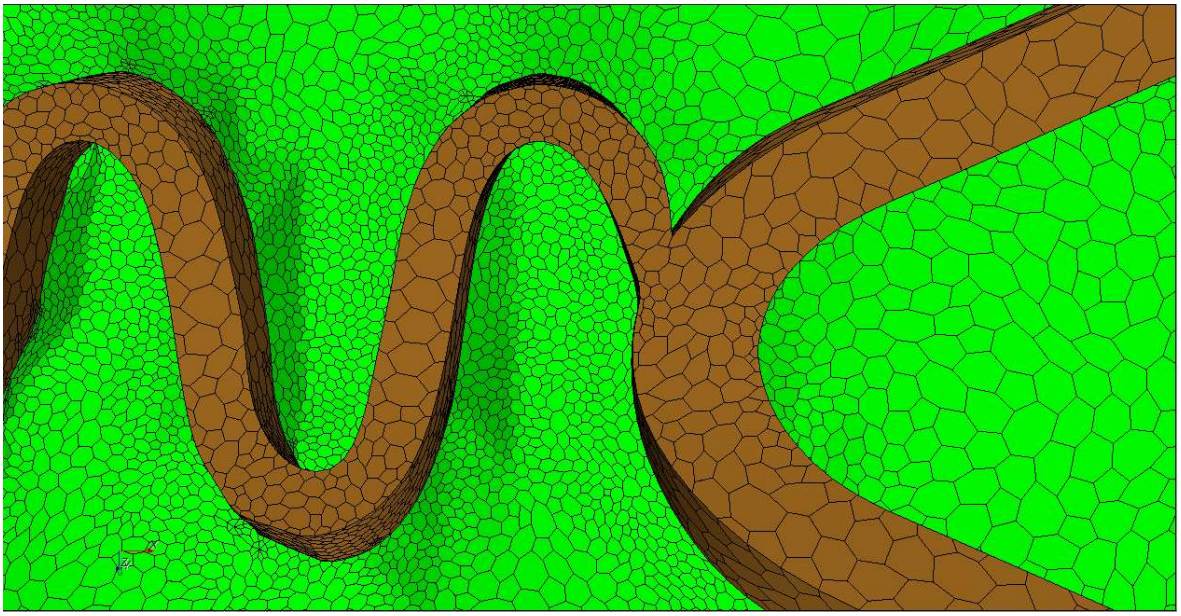


Figure 8.3: Volume mesh of the imported geometry: closeup

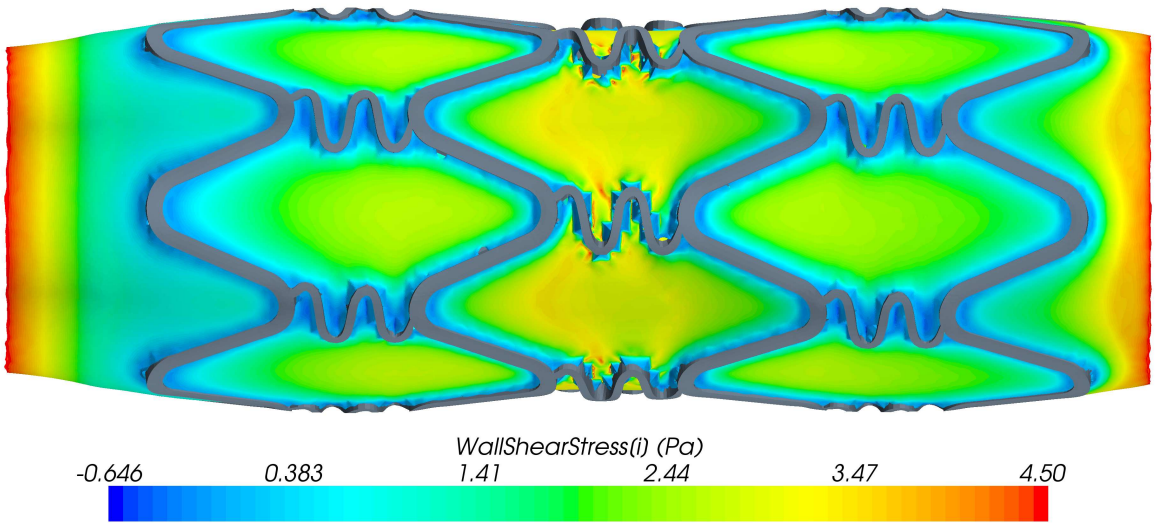
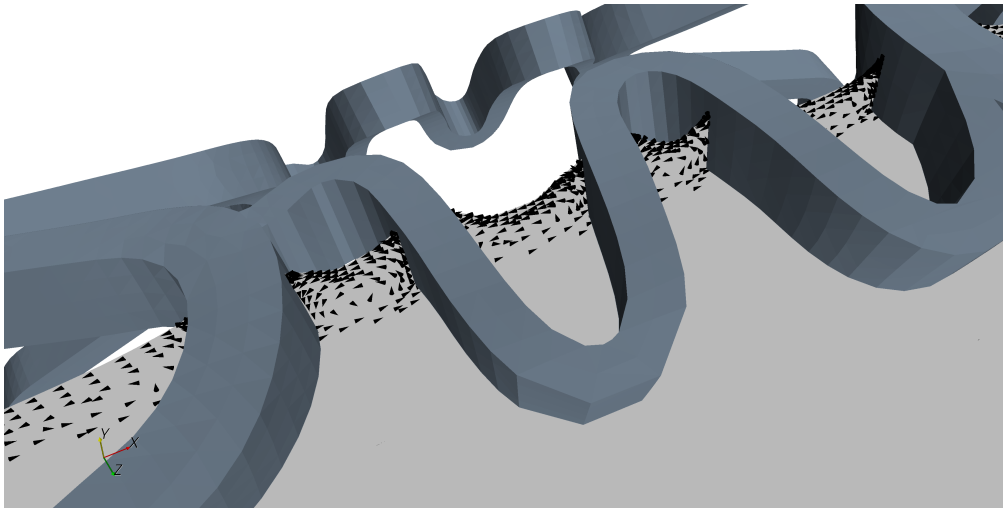
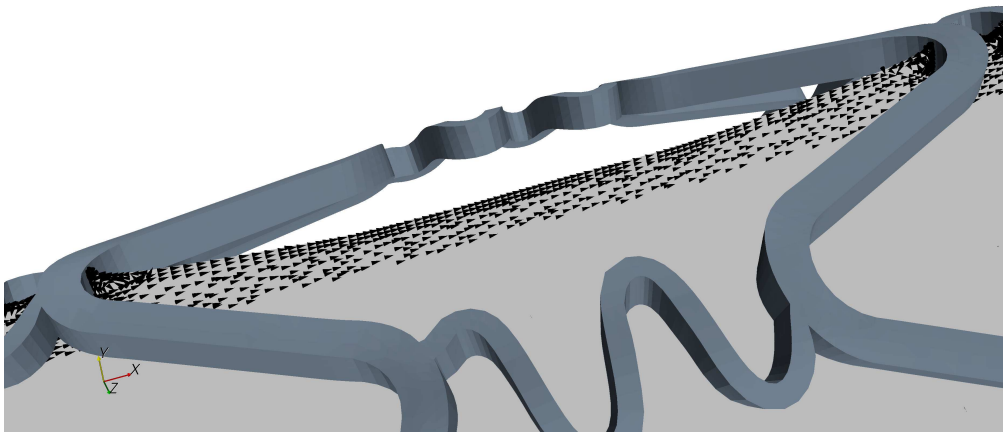


Figure 8.4: Axial WSS at point 5 of the cardiac pulse (c.f. figure 4.4)



(a) Velocity profile: links



(b) Velocity profile: central region

Figure 8.5: Velocity profiles for flow on geometry obtained post FEA analysis: point 3 (c.f. figure 4.4) of cardiac pulse

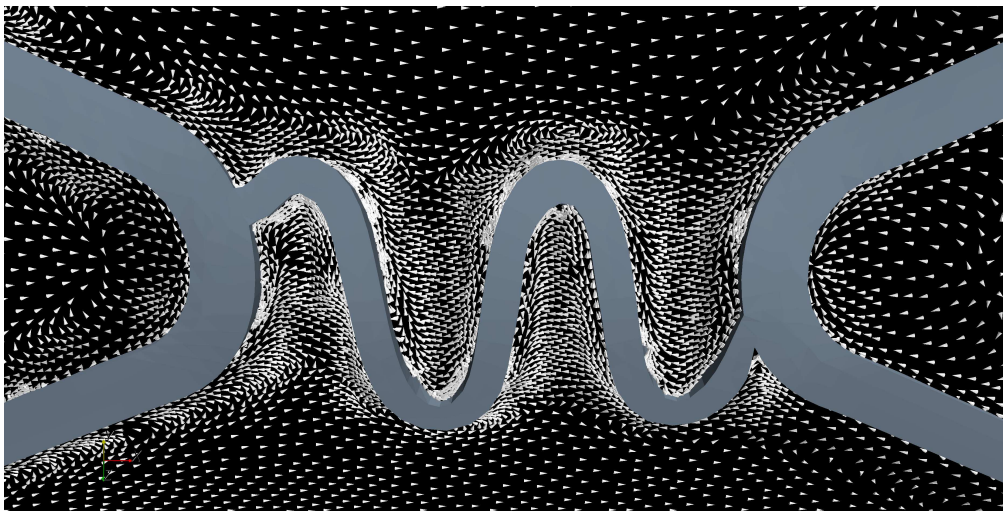


Figure 8.6: Secondary recirculation

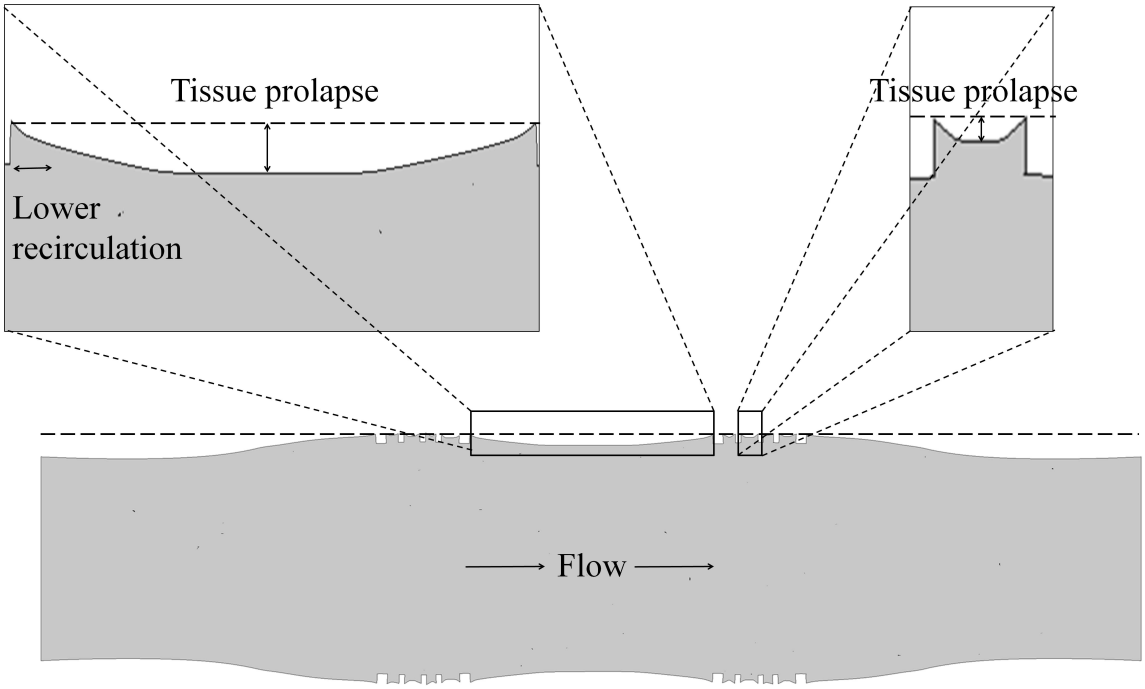


Figure 8.7: Tissue prolapse and it effect on recirculation

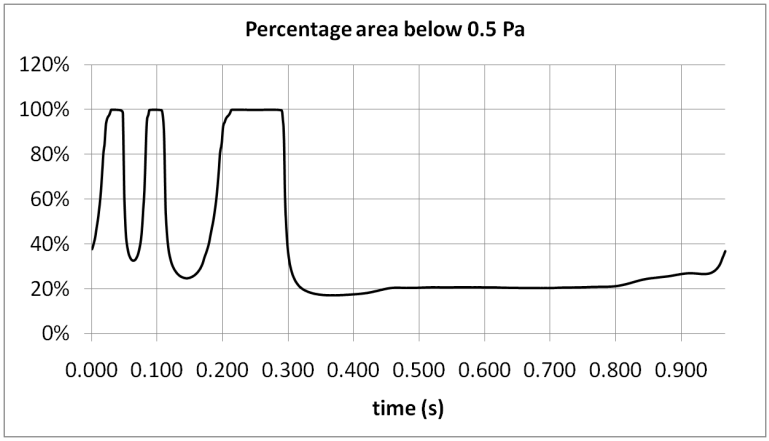


Figure 8.8: Percentage area exposed to WSS magnitude below 0.5 Pa over the entire cardiac pulse

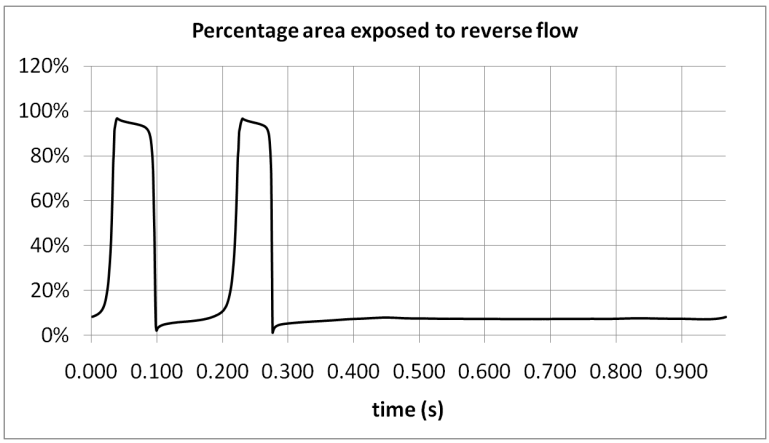


Figure 8.9: Percentage area exposed to reverse flow over the entire cardiac pulse

8.2 Part II : Drug-distribution

In this section a CFD model to evaluate the distribution of drug in a stented artery segment is discussed.

8.3 Introduction

The modelling of drug diffusion in the stented vessel applies to drug eluting stents which release a drug to inhibit inflammatory response and smooth muscle cell migration and proliferation. Since most drugs do not bind to the metallic stent surface easily, a polymer coating is often needed both to fix the drug on the stent and control the release kinetics [174]. Although a uniform surface coating is the norm in DES, another technology known as the reservoir (RES) technology is being researched by Conor Medsystems LLC, Johnson & Johnson. RES technology stents have small multiple reservoirs on the stent struts, each reservoir comprising various layers of drug and polymers to control drug release [175, 176]. The drug, a biologically active agent, arrests any one of the phases of the cell cycle. Figure 8.10 shows this cell cycle [12], the series of events after stenting which lead to in-stent restenosis, and the main targets of common drugs to inhibit restenosis. The cell cycle has four phases –

1. G1 phase – This is also known as the gap-1 phase (gap referring to the time between the previous M phase and the beginning of DNA synthesis). During this phase various enzymes are synthesized to be used for the S phase.

2. S phase – This phase involves DNA synthesis and replication of the chromosomes.

3. G2 phase – Also known as the gap-2 phase, this phase involves preparation (biosynthesis) for the M phase.

4. M phase – This phase involves Mitosis which leads to separation of chromosomes and cytoplasm.

Different drugs attack the cell cycle at different points, and consequently have a different mechanism of suppressing restenosis. The two most common drugs used in DES today are Sirolimus and Paclitaxel. While Paclitaxel attacks the M-phase, Sirolimus inhibits the G1 phase. The implication of using different drug types to arrest restenosis is out of the scope of this thesis. Focusing on the engineering evaluation of drug-distribution, a general methodology for drug transport, the diffusivity values of the drug depending on the type of drug used, is presented here.

In the past many studies have been performed on the modelling of drug diffusion in stented vessels. Hwang et. al. [177] performed experimental drug distribution studies by spray-coating Palmaz-Schatz Crown stents with fluorescein sodium ethlylene vinyl acetate copolymer solution in dicholoromethane, and implanting these stents in bovine carotid arteries. The resulting fluorescence microscope images revealed a high spatial heterogeneity in drug concentrations. They also evaluated the effect of strut placement in the circumferential direction by setting up a 2-D model and randomly placing the struts. They reported that inhomogeneous stent strut placement along the circumference can result in a highly non-uniform spatial distribution of drug in the tissue (implying higher concentration near struts and a lower concentration in the inter-strut spaces). Lovich et. al. [178] performed perfusion

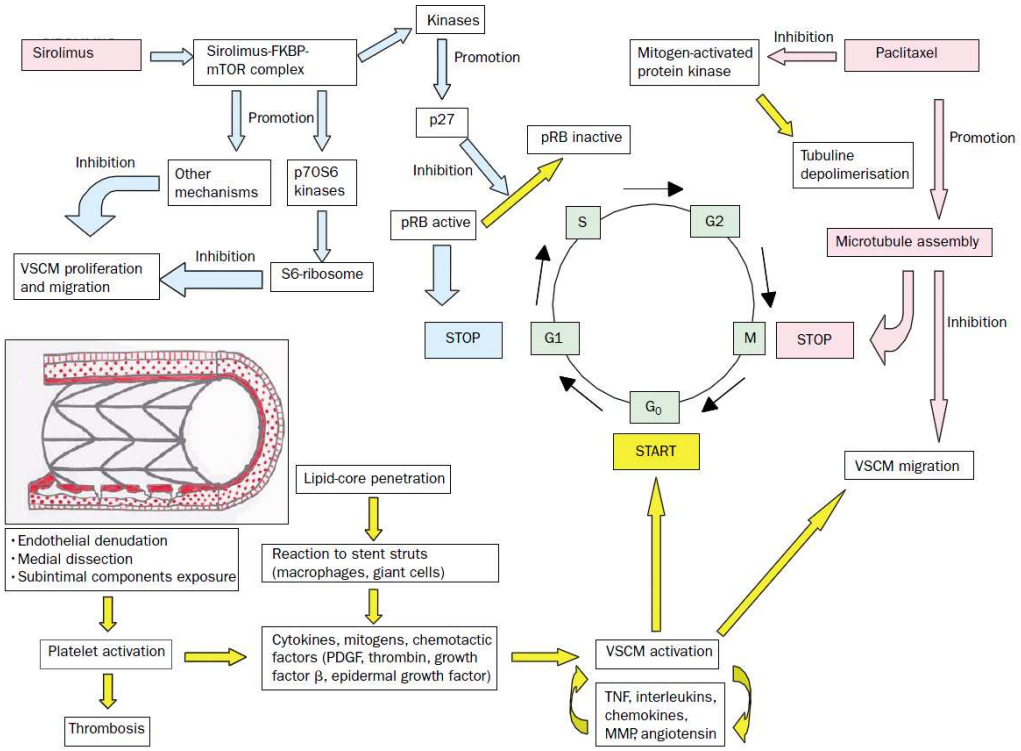


Figure 8.10: The cell cycle and targets for drugs used in DES [12]

experiments on calf carotid arteries and concluded that convective drug transport in the tissue played a significant role in de-endothelialised arteries when compared to those where the endothelium was intact. Hose et. al. [170] studied Paclitaxel distribution from a BiodivYsio stent using a computational model. They reported significant heterogeneity in spatial distribution of the drug in the tissue and attributed this to the design of the stent. Sakharov et. al. [179] implemented reverse binding of the drug in the tissue and numerically showed that high polymer diffusion resistance increased average concentration of the drug in the tissue. Pontrelli et. al. [180] presented analytical solutions for the problem of 1-D drug transport from a polymer layer into the tissue. Zunino [181] modelled the drug-release dynamics from the stent coating in a 2-D setup and evaluated the effect of drug properties, stent coating, and arterial wall on drug dynamics. Borghi et. al. [182], in an axi-symmetric geometry, evaluated the difference between pure diffusive transport and reverse binding. Balakrishnan et. al. [183] implemented a coupled CFD and mass transfer model to study drug-release dynamics in a 2-D setup. They concluded that flow alteration and location of struts were very important to achieve higher drug deposition. Migliavacca et. al. [184] developed a model to study drug-diffusion on a geometry obtained by FE stent expansion analysis. They proposed a model for plasma infiltration into the tissue and then used the advection-diffusion equations to model drug release. Their model accounted for reverse binding of the drug in the tissue and the drug dissolved in plasma. Zunino et. al. [20] proposed a unified methodology to evaluate the expansion, haemodynamics, and drug-release for DES. Their drug release simulations involved a standard advection-diffusion process with no reverse binding of the drug with the tissue. Feenstra et. al. proposed a sequential expansion and drug-release analysis. Along with the geometry, they extracted the interstitial fluid velocity, from the expansion analysis and included reverse binding of the drug with the tissue. Similar to the work by Balakrishnan

et. al. mentioned above, Kolachalama et. al. [16] studied the effect of flow in the lumen on arterial drug deposition. Their steady state and pulsatile flow models, implemented in a 2-D geometry, revealed differential lengths of recirculation zones around the struts which affect the transfer of drug into the tissue. They then extended the steady state models to 3-D geometry of stents placed in a bifurcation [185], and evaluated the effect of the location of the stent in the main branch relative to the side branch.

In this chapter, a steady state, pure diffusion model in the tissue, and an advection-diffusion model in the lumen, with appropriate coupling is used to evaluate the drug-distribution patterns. The effect of reverse binding (endocytosis) and advection in the tissue is ignored [20].

8.4 Methodology

The following subsections describe the methodology adopted in this chapter.

8.4.1 Geometry

The geometry is obtained from the deformed shape of section 5.5.2. The plaque and the artery regions are combined to represent one homogeneous and isotropic tissue region. Figure 8.11 shows this assembly. The embedding of the stent into the tissue region, as shown in the figure, makes the stent both a part of the lumen and the tissue region.

8.4.2 Governing equations

The flow in the lumen significantly affects the drug distribution, as the distribution of drug in the lumen follows an advection-diffusion pattern. This presents a challenge – since the time-scales of the flow and drug-diffusion differ in orders of magnitude, the coupling of unsteady flow with the drug diffusion equations becomes computationally very expensive. Moreover, since the objective of this thesis to search for optimal designs for drug delivery which involves comparison of the drug distribution patterns for different stent geometries, a steady state drug-diffusion problem is considered in this chapter. Although an unsteady time-dependent release, without considering the luminal flow, can be modelled to gain insight on drug delivery patterns, a steady state problem, by including the flow, gives enough variation when comparing different designs. Transmural convection, due to the interstitial flow in the tissue, is neglected. This is based on the study by Kolachalama et. al. [16] where it is shown that transmural convection has an effect of less than 1% on volume weighted average concentrations of drug. This study also shows that steady state simulations are reflective of the instantaneous flux of drug through the interface, and hence can provide useful relevant information.

The following continuity, momentum, and diffusion, equations are solved in the lumen –

$$\nabla.\mathbf{v} = 0, \tag{8.1}$$

$$\rho\mathbf{v}.\nabla\mathbf{v} = -\nabla P + \nabla.(\mu\nabla\mathbf{v}), \tag{8.2}$$

$$\mathbf{v}.\nabla C_l = D_l\nabla^2 C_l, \tag{8.3}$$

where $\mathbf{v}$, ρ, P, and μ denote the velocity, density, pressure, and the viscosity of blood, respectively. C_l is the drug concentration in the lumen and D_l is the diffusivity of the drug in blood.

As in chapter 4, ρ and μ are assumed to be equal to 1.060×10^{-3} kg/m^{-3} and 3.7×10^{-3} Pa-s.

In the tissue region (comprising both the plaque and the artery), the following pure diffusion equation, similar to equations 7.1 and 7.2 used in chapter 7, is solved –

$$D_t\nabla^2 C_t = 0, \tag{8.4}$$

where C_t is the drug concentration in the tissue, and D_t is the diffusivity of the drug in tissue. The interface between the lumen and plaque is modelled in a similar manner as the plaque-artery interface in chapter 7 (equation 7.4). The flux across the lumen-plaque interface is defines as –

$$q(C_l, C_t) = k_p(C_l - C_t), \tag{8.5}$$

where q is the flux across the interface, k_p is the permeability, and C_l, C_t are drug concentrations in the lumen and drug respectively.

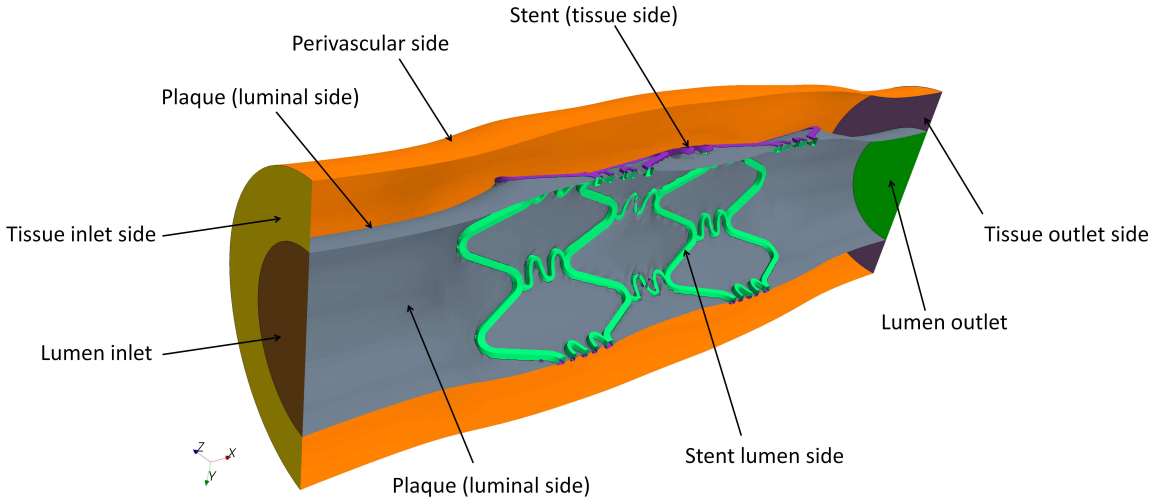


Figure 8.11: Section of the assembly for the drug release simulation – lumen and tissue

8.4.3 Boundary conditions

For the both the lumen and the tissue region, the stent is set to a unity concentration boundary condition. Such a Dirichlet boundary condition has also been used by Kolachalama et. al. [16] to model drug release. The 'lumen inlet' is set to an inlet velocity of 16.29 cm/s (mean velocity of the inlet profile shown in figure 4.4, and a zero concentration boundary. The 'lumen outlet' is set to a zero pressure boundary for flow and a zero flux boundary is specified for concentration. In the tissue region the 'tissue inlet side' is set to a zero concentration boundary, and both the 'tissue outlet side' and the 'perivascular side' are set to a zero flux boundary condition for the drug. The plaque interface is modelled using equation 8.5. The diffusivity of drug in the lumen (D_l) and tissue (D_t) are assigned a value of 1.5×10^{-4} mm^2/s and 7.7×10^{-6} mm^2/s respectively, and the permeability (k_p) is assumed to be 4×10^{-4} mm/s [20, 186]. Equations 8.1-8.4, coupled by Eqn. 8.5, are then solved over the domains of lumen and tissue. The equations are modelled as a heat transfer problem in Star-CCM+ 3.06.006. The similarity between the diffusion equations and the heat equation is utilized for this purpose [169, 171, 170].

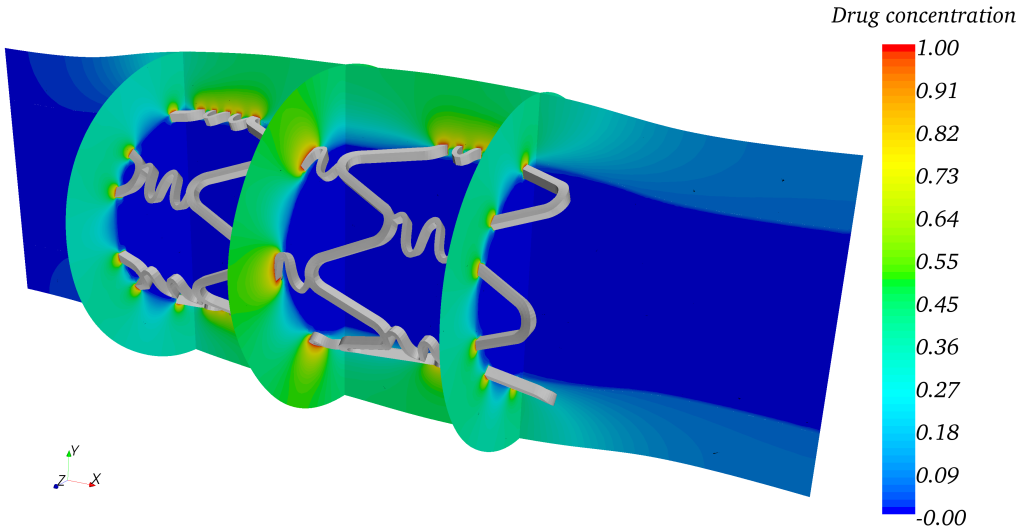


Figure 8.12: Drug distribution contours

8.5 Results & Discussion

Figure 8.12 shows the drug contours on various sections of the two domains. The highest drug concentrations are observed near the stent struts. This is not surprising as the stent is the source of the drug. As one moves away from the stent struts, the drug concentration decreases. In the axial cross sections it is observed that the drug concentration in the tissue region between any two consecutive struts in the circumferential direction reaches a minimum. This is attributed to the design of the stent. If one considers the entire tissue domain then the uniformity of drug distribution can be attributed to the areas where the source region (unity concentration boundary condition) lies. In other words, the stent can be seen as a set of curves on the plaque surface, and the topology of these curves determine the pattern of drug distribution. Areas which are far from such curves both in the circumferential and the axial directions are more likely to receive less drug relative to the areas which are nearer.

In the lumen region most of the drug is washed away because of the high convective transfer. It is the recirculation of the flow which retains some drug near the struts, and hence helps drug transport into the tissue. This is shown in figure 8.13. The only areas in the lumen which have a non-zero drug concentration are the areas of recirculation. Each recirculation zone, formed before and after each strut, arrests the washing-out of drug. Part of this drug is then transported to the tissue across the plaque interface. These recirculation zones, although unwanted while considering the haemodynamics, appear to be beneficial for drug transport. In many of the newer DES designs, the drug is coated only on the abluminal side of the stent struts. This prevents the drug loss due to the high convective transport in the lumen.

8.6 Formulation of objective functions

As mentioned in chapter 7, two factors are considered important relating to the drug response of a stent – first concerning the amount of drug in the tissue, and second concerning the uniformity of drug. The metric relating to the former, *VAD*, is formulated in section 7.2.1, chapter 7. In line with the discussion in section 7.3.3 (page 129), the standard deviation of the drug distribution in the tissue region is proposed to measure uniformity of drug delivered.

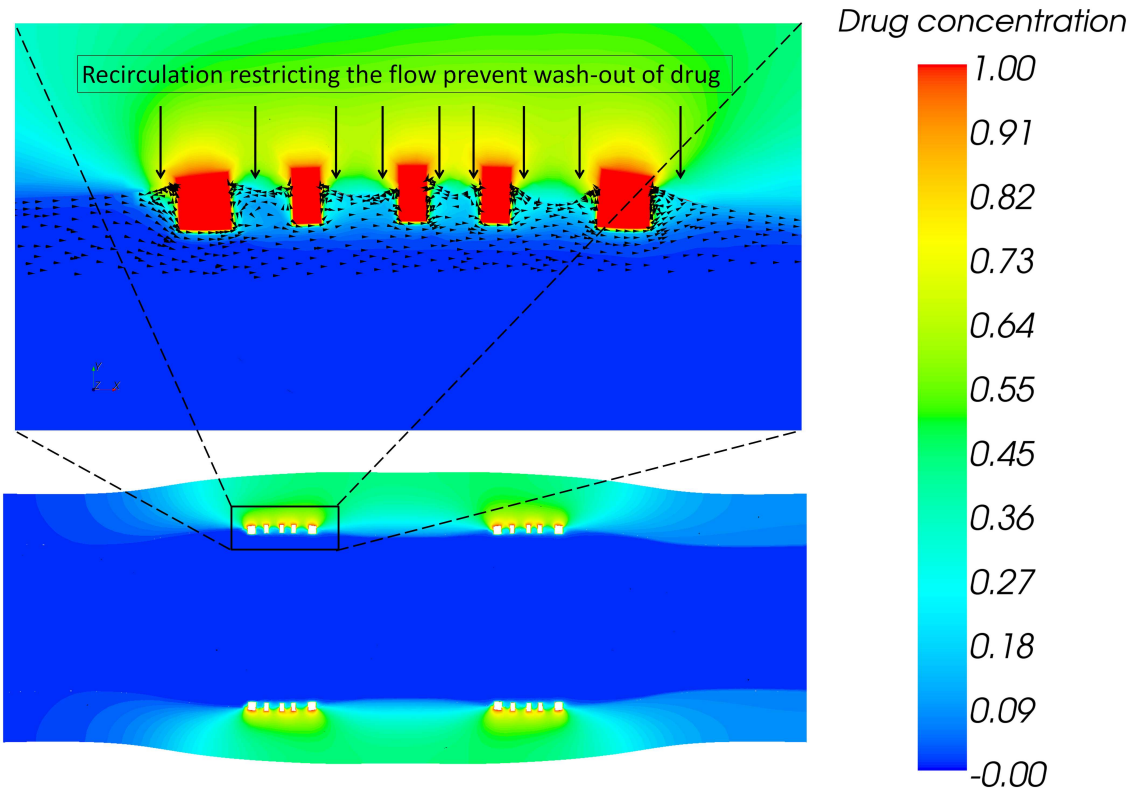


Figure 8.13: Drug distribution contours on a section plane: effect of flow

Let n_{cell} be the number of tissue cells lying in the axial domain shown in Figure 5.29 (page 97), and the set $D_d = \{c_1, c_2, c_3, ..., c_{n_{cell}}\}$ contain the drug concentrations, c_i, in each of these n_{cell} cells. Then, the standard deviation (D_{dev}) of this can be used as a measure of uniformity of drug-distribution around its mean.

$$D_{dev} = \sqrt{\frac{1}{n_{cell}} \sum_{i-1}^{i=n_{cell}} (c_i - \bar{c})^2}, \tag{8.6}$$

where $\bar{c}$ is the mean of elements in set D_d. A lower value of D_{dev} implies a more uniform distribution of the drug around its mean.

As discussed in section 7.3.3, it is important to node that D_{dev} could have a low value (implying a more uniform distribution) even when the mean is low (implying lesser drug). Hence, this measure should be used in conjunction with VAD in order to compare stents. The values for VAD and D_{dev} for the geometry used in this chapter are 0.3701 and 0.2525 units respectively.

8.7 Conclusions

In this chapter, haemodynamic evaluation of stented vessels, whose geometry is obtained by FEA expansion analysis, is performed. Recirculation zones, similar to those reported in chapter 4, are observed. However, considerable tissue prolapse is observed in the post-expansion geometry, which reduces the size of the recirculation zones. Furthermore, a CFD model, that accounts for luminal blood flow, to evaluate drug-distribution in the post-expansion stent-artery assembly is developed. High concentrations of drug are observed, both in the tissue

and lumen region, around the struts. It is concluded that recirculation zones, although undesirable from a haemodynamic point of view, promote drug transport from the lumen in to the tissue. Lastly, an objective function D_{dev}, which quantifies the uniformity of drug-distribution in a stented artery is formulated.

This chapter concludes the engineering analysis of stents. Now, the performance of any stent design can be measured by six figures of merit (objective functions) which represent

1. acute recoil, $Recoil$, defined by equation 5.10, which is a measure of both restenosis and structural strength of a stent (see section 5.7.1),

2. tissue stresses, VAS, defined by equation 5.12, which is a measure of injury caused to the artery by the stenting procedure (see section 5.7.2),

3. haemodynamic disturbance, $HLRFI$, defined by equation 4.9, which is a measure of stented artery area exposed to low shear stress and reverse flow (see section 4.4),

4. drug delivered to the tissue, VAD, which is a measure of the average amount of drug delivered to the tissue by the stent (see section 7.2.1),

5. uniformity of drug distribution, D_{dev}, defined by equation 8.6 (see section 8.6), and

6. stent flexibility, FM, defined by equation 6.1 (see section 6.5).

With the above metrics, the next chapter concerns with a multiobjective formulation of the stent design problem.

Chapter 9

Multiobjective optimisation study on the CYPHER stent

This chapter[1] combines the stent analysis performed in previous chapters to present the first multi-objective and multi-disciplinary coronary stent design optimization study of its kind. In line with the discussion on section 7.4 (page 133), in this chapter an improvement in all the metrics, used to measure a stent's efficacy, is sought simultaneously, as opposed to one at a time in chapter 7. The aims of this chapter are

1. to propose a parameterisation technique for the CYPHER (Cordis corporation, Johnson & Johnson co.) stent. The details of this parameterisation and a justification for this choice are discussed in section 9.1,

2. to perform a multiobjective optimisation study using the above proposed parameterisation technique,

3. to study trade-offs between various pairs of objectives that are in conflict,

4. to evaluate trends between the non-dominated solutions and parameters defining stent geometry,

5. to propose various paradigms, which can be used to choose ideal stent(s), from the set of non-dominated solutions, and

6. to demonstrate a link betweeen the results of the multiobjective study of this chapter and the constrained optimisation study presented in chapter 7.

The performance of each stent design is measured by six figures of merit (objectives/metrics): $Recoil$, VAS, $HLRFI$, VAD, D_{dev}, and FM. These metrics are obtained from computational simulations of

[1]The contents of this chapter are published in the below mentioned article with the author's supervisory team and Dr. N. Curzen. Dr. N. Curzen is a cardiac surgeon who actively advises the author regarding the clinical aspects of restenosis.
Pant, S., Limbert, G., Curzen, N. P., Bressloff, N. W. Multiobjective design optimisation of coronary stents. *Biomaterials*, 32:7755–7773, 2011.

1. structural deformation through balloon inflated expansion of a stent into contact with a stenosed vessel (see chapter 5),

2. pulsatile flow over the deformed stent embedded in the vessel wall (see chapters 4 and 8),

3. steady-state drug distribution into the tissue (see chapter 8), and

4. flexibility of a stent in response to an applied moment (see chapter 6).

Design improvement is obtained by a multi-objective surrogate modelling approach (see section 3.2.3) using a non-dominated sorting genetic algorithm (NSGA-II, see section 3.5.2) to search for an optimal family of designs. A number of trade-offs between the different objectives are identified. In particular a conflict between pairs of the following objectives are shown – (a) volume average stress vs recoil, (b) volume average drug vs. volume average stress, (c) flexibility vs volume average stress, (d) flexibility vs. haemodynamic disturbance, (e) volume average drug vs. haemodynamic disturbance, and (f) uniformity of drug vs. volume average stress.

Having presented a review of the optimisation studies in the area of stent design in chapter 7 (section 7.1), this chapter starts with the definition of the proposed parameterisation technique for the CYPHER stent. Then, the multiobjective problem is formulated and the solution methodology adopted to solve this problem is presented. Thereafter, various slices of the obtained Pareto front are discussed, and based on these results different paradigms to choose the optimal designs from the Pareto fronts are proposed. Finally, under each such paradigm, the optimal designs and their relative positions with respect to a representative CYPHER stent are shown.

9.1 Stent geometry parameterisation

The constrained optimisation study, chapter 7, showed that while the parameters W_{strut}, h_c, and p_1, had significant effect on the metrics used, the parameter p_2 had little effect on the efficacy of a stent (see sensitivity analysis: Figure 7.18, page 128). Hence, it is unreasonable to include two parameters for defining the links, thereby increasing the dimensionality of the design space, when one variable does not have significant effect on the evaluation metrics. With this view, an alternate parameterisation for the links is sought, while the parameterisation of circumferential rings is kept the same as in chapter 7. For the links, a CYPHER like link is chosen for two reasons. First, the link in CYPHER stent has a uniform 'n' shaped structure. Hence, only one parameter which specifies the height of the links (i.e. circumferential dimension) acts similar to the p_1 parameter used in chapter 7. Second, the FEA analysis used in this thesis have been validated for a representative CYPHER stent (against manufacturer data for expansion analysis, and against De Beule's study [11] for flexibility analysis). Hence, in the results, i.e. in the set of the non-dominated solutions, an assessment regarding the relative position of this representative CYPHER stent with respect to other solutions can be made.

Figure 9.1 shows the various parameters for the proposed parameterisation. The length of the stent, L_{stent}, and its semi-crimped state outer radius, R_{stent}, are fixed to be equal to 8.0 mm and 0.75 mm respectively. Four circumferential rings in the longitudinal direction are used

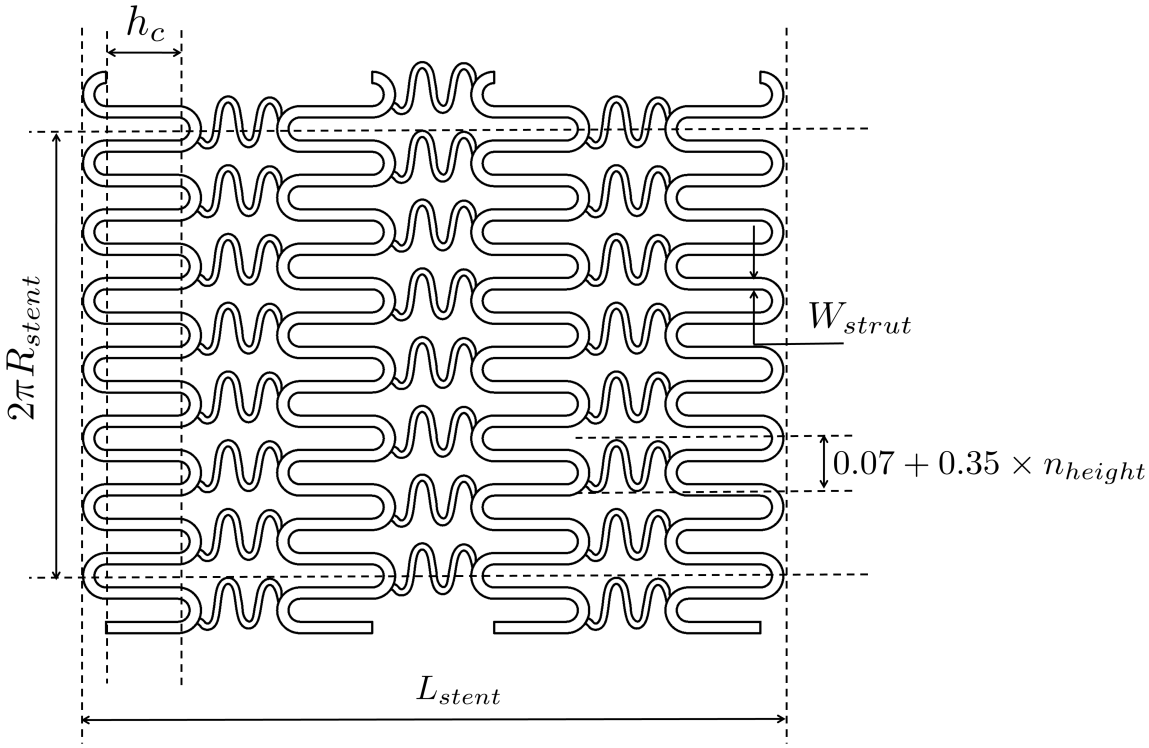


Figure 9.1: Stent parameterisation used in this chapter

and there are twelve curved parts (six peaks and six troughs) in each circumferential ring. The parameters which are allowed to vary are the longitudinal length of the circumferential rings, h_c, circumferential strut width, W_{strut}, and the height of the 'n' shaped links, n_{height}. Since L_{stent} is constant, h_c also controls the length of the links. The circumferential width of the links is kept constant at 0.07 mm. Strut thickness, i.e. the stent dimension in the radial direction, is also kept constant at 0.14 mm based on the value for the CYPHER stent [135]. The shape of the links is kept similar to the CYPHER stent using NURBS curves as shown in Fig. 9.2.

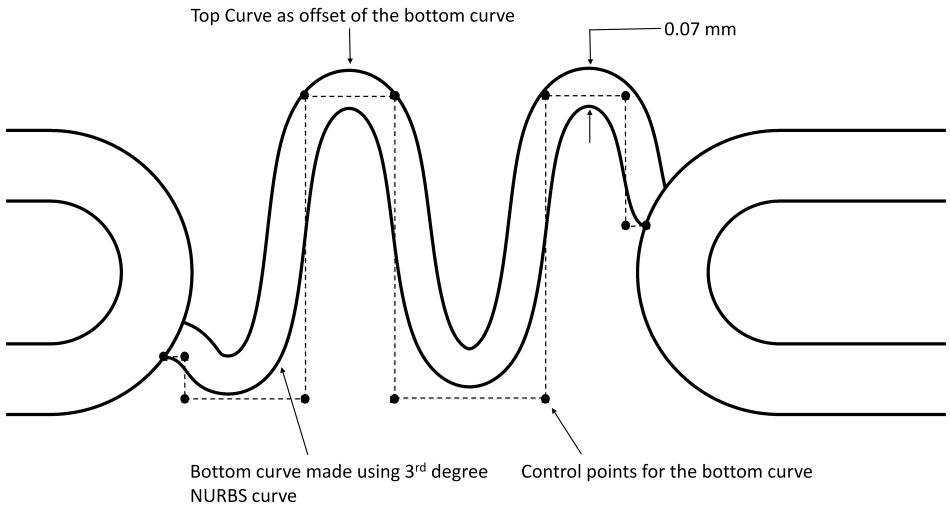


Figure 9.2: Stent link construction using NURBS control points

Table 9.1 shows the values of the three parameters for the representative CYPHER stent

and the bounds imposed on each parameter. For an explanation of the bounds on W_{strut} and h_c refer to section 7.2.2. The upper bound for n_{height} is chosen considering that a very high value of n_{height} results in contact of one link with the other. The lower bound for n_{height} is chosen considering the geometry of the 'n' links – with a constant circumferential width of the links a very small value of n_{height} results in almost straight links. Figure 9.3 shows some of the designs created using this parameterisation.

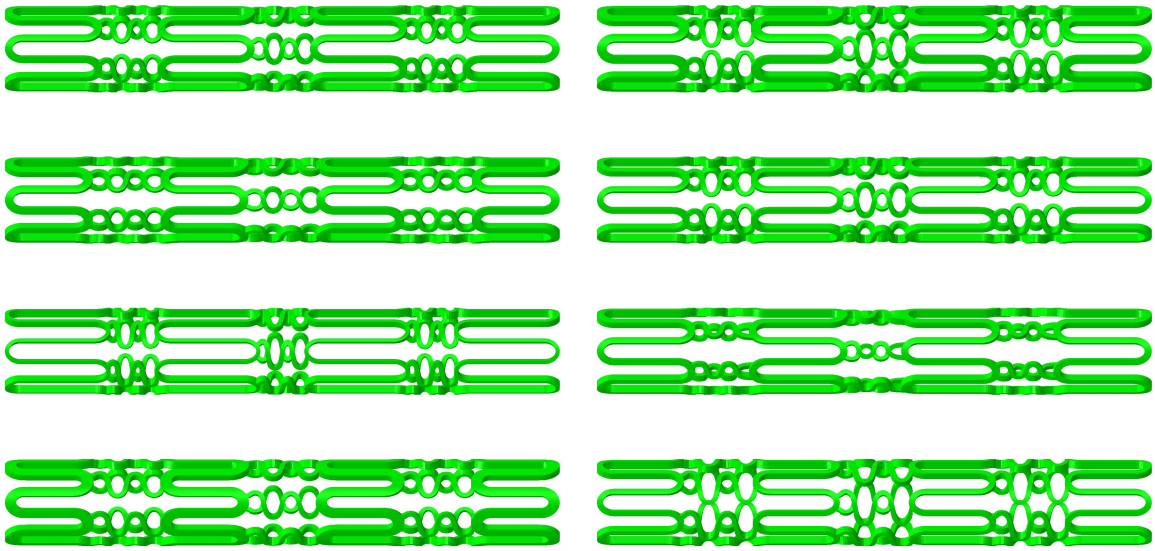


Figure 9.3: Sample designs created using the proposed parameterisation

9.2 Optimisation problem & solution methodology

The multi-objective optimisation problem is formulated as follows-

$$\begin{aligned}
&\textit{Minimise} && \textit{Recoil}\ (W_{strut}, h_c, n_{height}) \\
&\textit{Minimise} && \textit{VAS}\ (W_{strut}, h_c, n_{height}) \\
&\textit{Minimise} && \textit{HLRFI}\ (W_{strut}, h_c, n_{height}) \\
&\textit{Minimise} && -\textit{VAD}\ (W_{strut}, h_c, n_{height}) \\
&\textit{Minimise} && D_{dev}\ (W_{strut}, h_c, n_{height}) \\
&\textit{Minimise} && \textit{FM}\ (W_{strut}, h_c, n_{height})
\end{aligned} \tag{9.1}$$

Table 9.1: Limits imposed on the design parameters

	W_{strut} mm	h_c mm	n_{height} mm
CYPHER stent	0.130	0.825	1.600
Lower bound	0.050	0.700	0.700
Upper bound	0.170	1.100	1.900

such that,

$$0.050\ mm \leq W_{strut} \leq 0.170\ mm,$$
$$0.700\ mm \leq h_c \leq 1.100\ mm,$$
$$0.700\ mm \leq n_{height} \leq 1.900\ mm.$$

As discussed in section 7.4, some of the objectives in the above formulation could be specified as constraints.

The use of Kriging for the optimisation procedure is justified as the evaluation of all the metrics used in equation 9.1 for one design takes over 120 hours (wall time) using 8 parallel compute processes (Microsoft Windows 64-bit high performance computing platform, Intel quad core 2.8 GHz processor, 16 GB RAM). This high computational time and the limited licenses available for Abaqus/Exlpicit, Abaqus/Standard, and StarCCM+, limit the computational budget available for the analyses to 30 stent designs. Consequently, high-fidelity analysis is performed for 30 designs, which is 10 times the number of design variables used in this chapter. Based on this, the solution methodology adopted in this chapter is depicted in Figure 9.4. The process starts with the parametric definition of the stent geometry. A number of designs (15 in this chapter), uniformly distributed over the design search space, are created using the proposed parameterisation. This process of selecting the initial points in the design space is known as *sampling*. Each of these designs are then analysed for the six objectives defined in previous sections. The box with dashed lines in Fig. 9.4 illustrates the steps for analysis. For each design, CAD geometries are constructed, in a semi-crimped state, and exported to the FEA package to solve for the equations governing the stent expansion process and bending (for flexibility). The expanded geometries obtained as an output from the expansion analysis are then exported into the CFD package to model haemodynamics and drug distribution process. The objective functions from each of these four analyses are extracted and used to construct the initial surrogate model. The surrogate models are constructed using Gaussian Process modeling/Kriging (see section 3.2.3. These models (Krigs/GP models) are then searched using multi-objective non-dominated sorting genetic algorithm-II (NSGA-II) [74] (see section 3.5.2), resulting in an initial predicted Pareto front. It should be noted that the obtained Pareto front is a predicted one, and is only as good as the accuracy of the Krigs. To improve the quality of the Krigs the update process is carried out. The number of update points (five in this chapter), also known as the *infill points*, are selected uniformly along the predicted Pareto front. After the analysis of the new points, the Krigs are reconstructed and the NSGA-II search is performed on the updated Krigs to obtain an improved Pareto front. This process is repeated twice more, based on the computational budget available.

An LP_τ [172] based sampling plan is used to construct a 15-point design of experiments [46]. An LP_τ based sampling plan is suitable as it gives a uniform coverage of the design space. The distribution of 30 evaluations is divided into an intial sampling of 15 points, and three subsequent updates, each update adding five new points. Once the Kriging models are constructed the NSGA-II algorithm is run for 50 generations, each generation having a population of 50 members, to search the Krigs. The methodology in this chapter can be summarised in the following steps –

Step 1: Analyse initial 15 points → Construct Krigs for each objective → perform an NSGA-II search to obtain the Pareto front → identify five update points → analyse the new points.

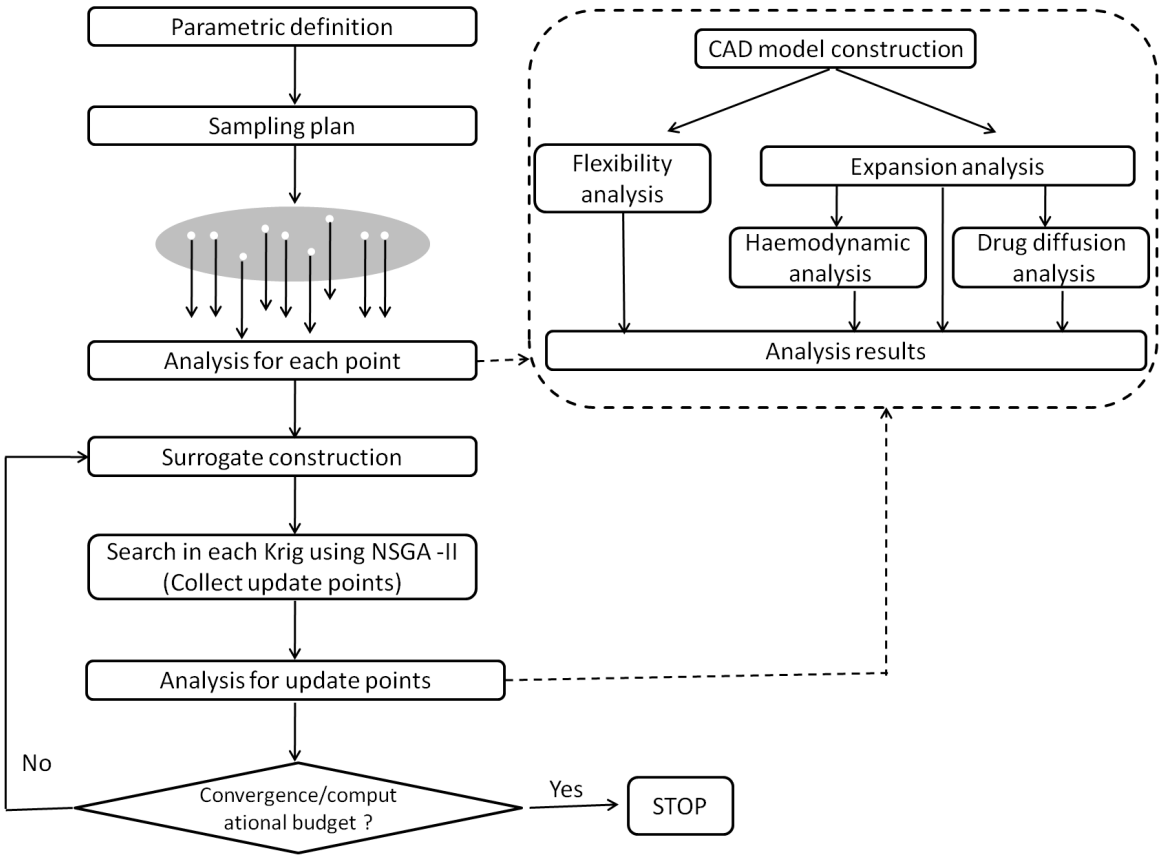


Figure 9.4: Flow chart detailing the optimisation methodology adopted in this chapter

Step 2: Construct Krigs for each objective using 15+5 points → perform an NSGA-II search to obtain the Pareto front → identify five update points → analyse the new points.

Step 3: Construct Krigs for each objective using 15+5+5 points → perform an NSGA-II search to obtain the Pareto front → identify five update points → analyse the new points → Stop.

Step 4: Appraise the optimal designs.

It is worth noting at this point that there are many methods proposed in the past to find update points. The method adopted in this chapter, i.e. to choose uniformly distributed update points on the predicted Pareto front, is the simplest of these. For single objective optimisation Jones et. al. [50] proposed an expected improvement criterion to find update points. This criterion was extended to multi-objective optimisation in many different algorithms [187, 188, 189, 190, 191, 192, 193]. A discussion on these algorithms is out of scope of this thesis. Nonetheless, the choice of a full exploitive approach adopted in this chapter is based on the following rationale: first, it is observed from chapter 7 that the objective functions modelled are not highly multimodal in nature, i.e. the physical responses show a relatively simple trend of increase or decrease in terms of a particular parameter in the search space modelled; and second the errors at the end of first update (shown in figure 9.5) are deemed acceptable in most of the search domain.

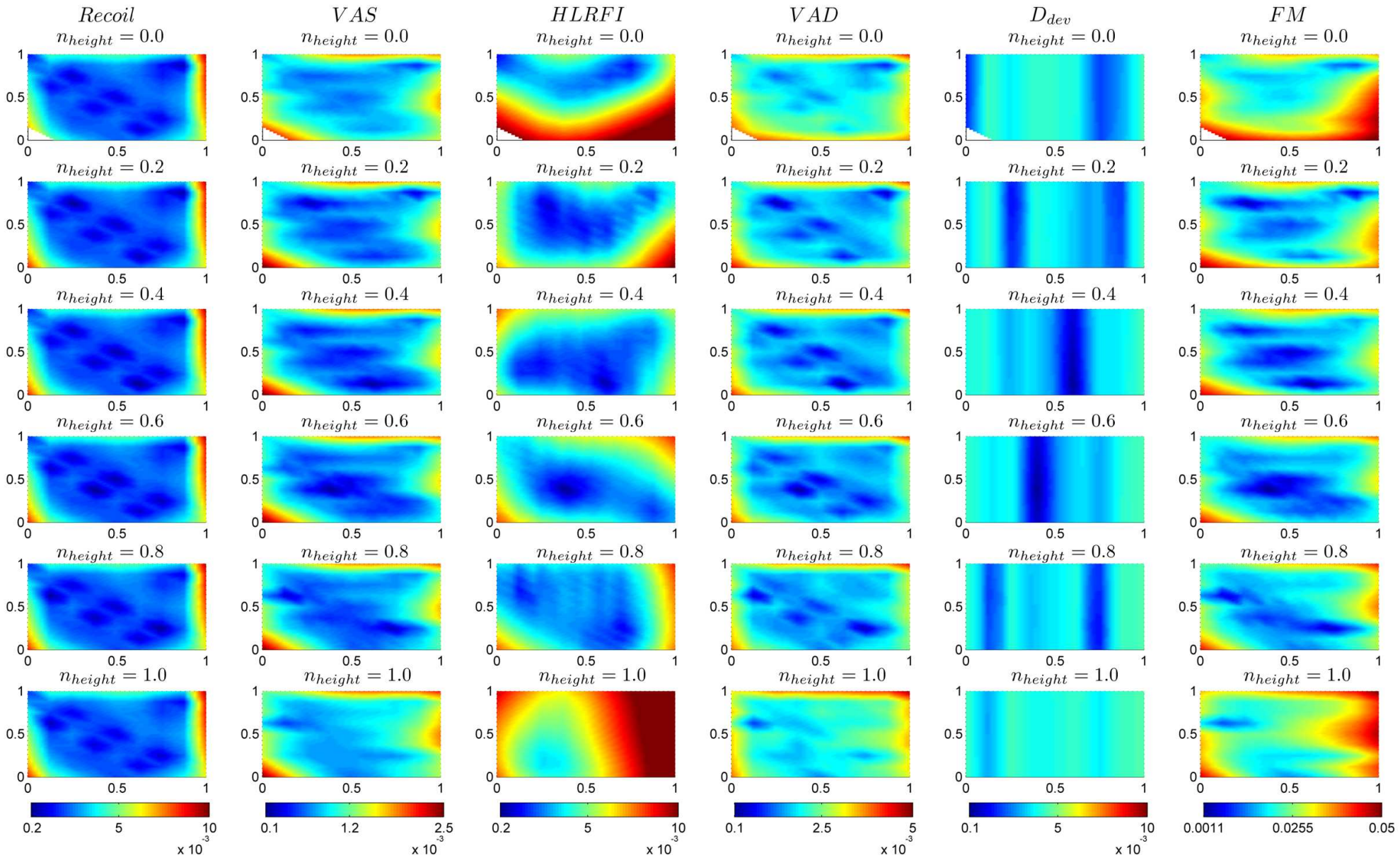


Figure 9.5: Errors in the Kriging models after first update: each subplot shows W_{strut} in x-axis, h_c on y-axis, while n_{height} is constant; and n_{height} increases vertically downwards in subplots

Table 9.2: Result matrix for the 15 point intial sampling and the three updates

Design	W_{strut}	h_c	n_{height}	*Recoil*	*VAS*	HLRFI	*VAD*	D_{dev}	*FM*
CYPHER	0.1300	0.8250	1.6000	0.1337	0.0464	27.40%	-0.3701	0.2525	0.1775
DOE_BX_1	0.1100	0.9000	1.3000	0.1716	0.0463	26.10%	-0.3605	0.2511	0.2094
DOE_BX_2	0.0800	1.0000	1.0000	0.2707	0.0398	24.42%	-0.3460	0.2528	0.2557
DOE_BX_3	0.1400	0.8000	1.6000	0.1146	0.0473	26.89%	-0.3792	0.2528	0.1690
DOE_BX_4	0.0650	0.9500	1.7500	0.2918	0.0400	26.43%	-0.3533	0.2648	0.2553
DOE_BX_5	0.1250	0.7500	1.1500	0.1226	0.0457	25.24%	-0.3632	0.2631	0.2241
DOE_BX_6	0.0950	0.8500	1.4500	0.1851	0.0449	27.39%	-0.3660	0.2610	0.1807
DOE_BX_7	0.1550	1.0500	0.8500	0.1405	0.0480	26.12%	-0.3696	0.2541	0.4149
DOE_BX_8	0.0575	1.0750	1.5250	0.3054	0.0376	26.34%	-0.3619	0.2564	0.1841
DOE_BX_9	0.1175	0.8750	0.9250	0.1548	0.0462	26.15%	-0.3528	0.2509	0.2808
DOE_BX_10	0.0875	0.7750	1.8250	0.1867	0.0446	27.90%	-0.3768	0.2512	0.1410
DOE_BX_11	0.1475	0.9750	1.2250	0.1372	0.0478	27.06%	-0.3645	0.2576	0.2557
DOE_BX_12	0.0725	0.8250	1.0750	0.2467	0.0398	24.56%	-0.3452	0.2551	0.2259
DOE_BX_13	0.1325	1.0250	1.6750	0.1639	0.0485	26.70%	-0.3787	0.2526	0.2089
DOE_BX_14	0.1025	0.9250	0.7750	0.1845	0.0448	24.56%	-0.3452	0.2509	0.3157
DOE_BX_15	0.1625	0.7250	1.3750	0.0901	0.0450	25.53%	-0.3772	0.2525	0.1907
DOE_BX_16	0.1395	1.0320	0.7367	0.1540	0.0473	24.80%	-0.3605	0.2519	0.4074
DOE_BX_17	0.0721	0.8256	1.1727	0.2529	0.0396	24.85%	-0.3482	0.2520	0.2101
DOE_BX_18	0.0503	1.0860	0.7232	0.3222	0.0345	23.34%	-0.3391	0.2596	0.2608
DOE_BX_19	0.1137	0.7068	1.7700	0.1439	0.0522	28.53%	-0.3758	0.2517	0.1440
DOE_BX_20	0.0890	1.0263	1.7776	0.2006	0.0435	26.65%	-0.3747	0.2544	0.1720
DOE_BX_21	0.1377	0.7051	1.8544	0.1098	0.0465	27.30%	-0.3886	0.2577	0.1403
DOE_BX_22	0.0500	1.0113	0.7010	0.3291	0.0340	23.31%	-0.3343	0.2566	0.2600
DOE_BX_23	0.0500	1.0576	1.0760	0.3090	0.0360	24.06%	-0.3475	0.2579	0.2162
DOE_BX_24	0.1677	1.0504	0.7021	0.1278	0.0479	24.76%	-0.3737	0.2536	0.3354
DOE_BX_25	0.0500	0.8121	1.7705	0.2922	0.0404	26.80%	-0.3771	0.2650	0.1359
DOE_BX_26	0.1175	0.7004	1.8919	0.1315	0.0472	27.76%	-0.3787	0.2493	0.1358
DOE_BX_27	0.0725	0.9891	0.7056	0.2542	0.0393	23.69%	-0.3427	0.2528	0.2969
DOE_BX_28	0.1550	0.9958	0.7010	0.1305	0.0475	24.76%	-0.3659	0.2520	0.4273
DOE_BX_29	0.0800	0.9997	0.7010	0.2710	0.0393	23.71%	-0.3387	0.2515	0.3137
DOE_BX_30	0.0632	0.7498	1.9000	0.2480	0.0403	26.71%	-0.3792	0.2611	0.1306

9.3 Results & Discussion

9.3.1 Model validation and response surfaces

For model validation, the same approach as used in chapter 7 is adopted. The leave-one-out and SCVR (see section 3.2.3.2) plots for each objective function after the end of three updates are shown in figures 9.6 – 9.11. These plots show that for each function the SCVR values after every update cycle, lie in the interval [-3,+3], thereby asserting the validity of the Kriging models for the chosen metrics. In each of these figures, every subplot represents W_{strut} on the x-axis and h_c on the y-axis; n_{height} increases vertically downwards through subplots.

The response surfaces for all the objective functions at the end of each update are shown in Figures 9.12 – 9.17. Figures 9.12 and 9.15 show that the response surfaces for *Recoil* and *VAD* do not change significantly through the three updates. On the other hand considerable

change, especially for n_{height} values lower than 0.6, is observed for the response of FM and $HLRFI$ as update points are added (see figures 9.14 and 9.17). Significant change in the response of VAS is also observed by the addition of update points (cf. figure 9.13). The plots for D_{dev} show extreme changes through the updates. This could be a result of many factors: first, it could be that the response of D_{dev} is too complicated to be modelled by a Kriging model (even though the SCVR plots suggest otherwise); second, it could be that the response of D_{dev} is governed completely by only one parameter, namely W_{strut}, which explains the vertical patches of colors in the last column of Figure 9.16 and insignificant change in D_{dev} by changes in h_c and n_{height}; third, it could be that D_{dev} is not a particularly good measure to differentiate stents based on uniformity of drug distribution; and fourth, it could be, looking down the second-last column of table 9.2, that it is the very similar D_{dev} response of even drastically different stent geometries that make this objective difficult to model. The last problem could potentially be remedied by choosing an appropriate transformation for the D_{dev} metric and constructing a Krig of the transformed metric instead of D_{dev}.

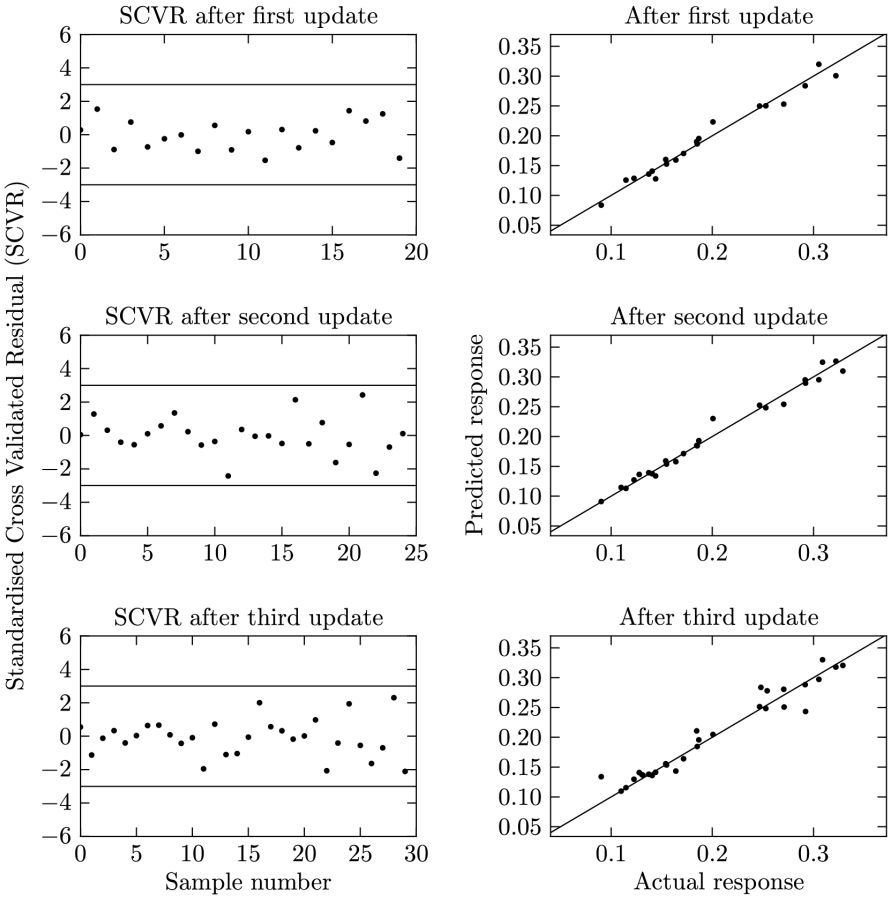


Figure 9.6: SCVR and leave-one-out plots for *Recoil* after each update

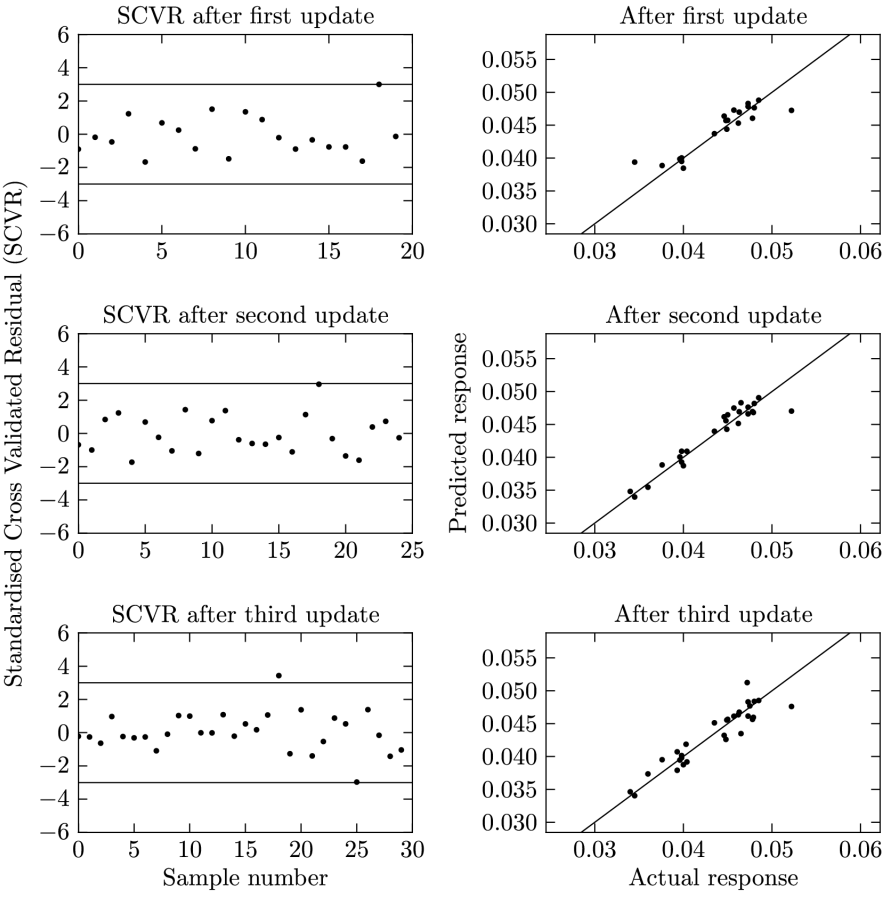


Figure 9.7: SCVR and leave-one-out plots for *VAS* after each update

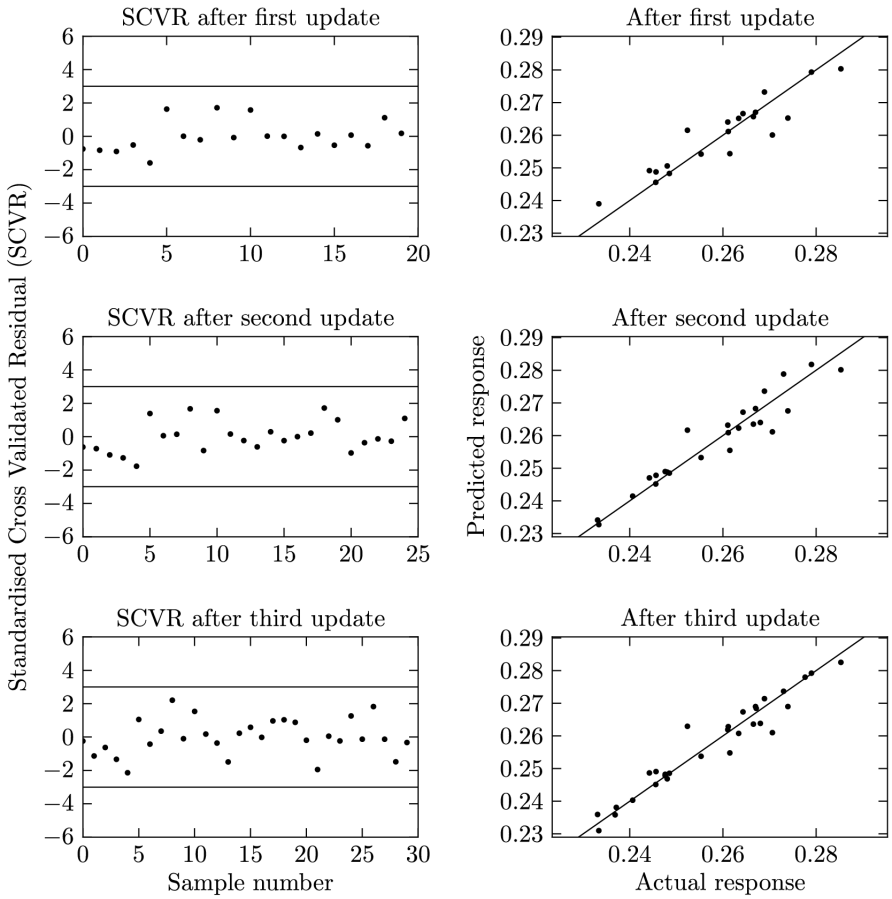


Figure 9.8: SCVR and leave-one-out plots for *HLRFI* after each update

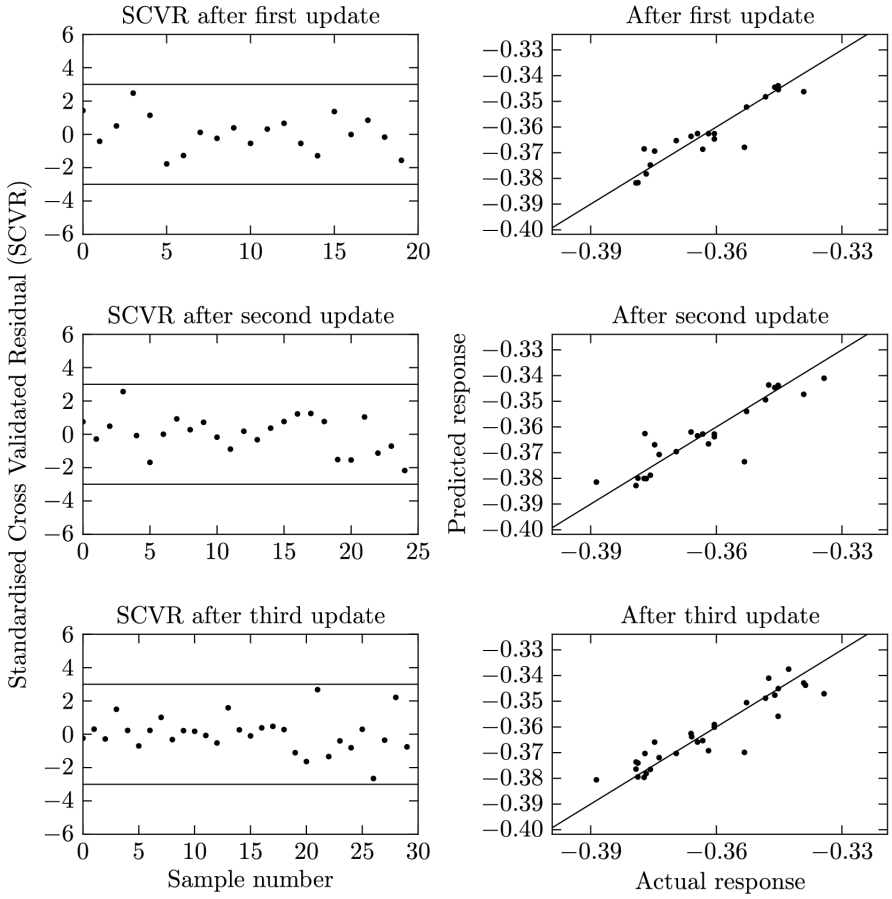


Figure 9.9: SCVR and leave-one-out plots for *−VAD* after each update

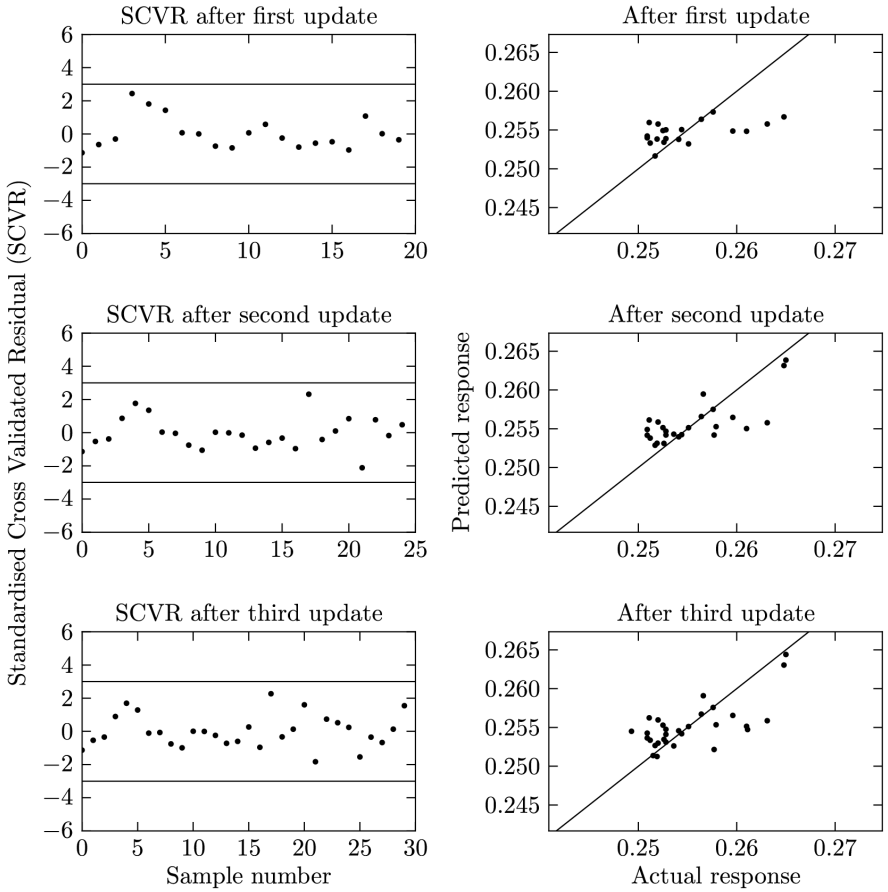


Figure 9.10: SCVR and leave-one-out plots for D_{dev} after each update

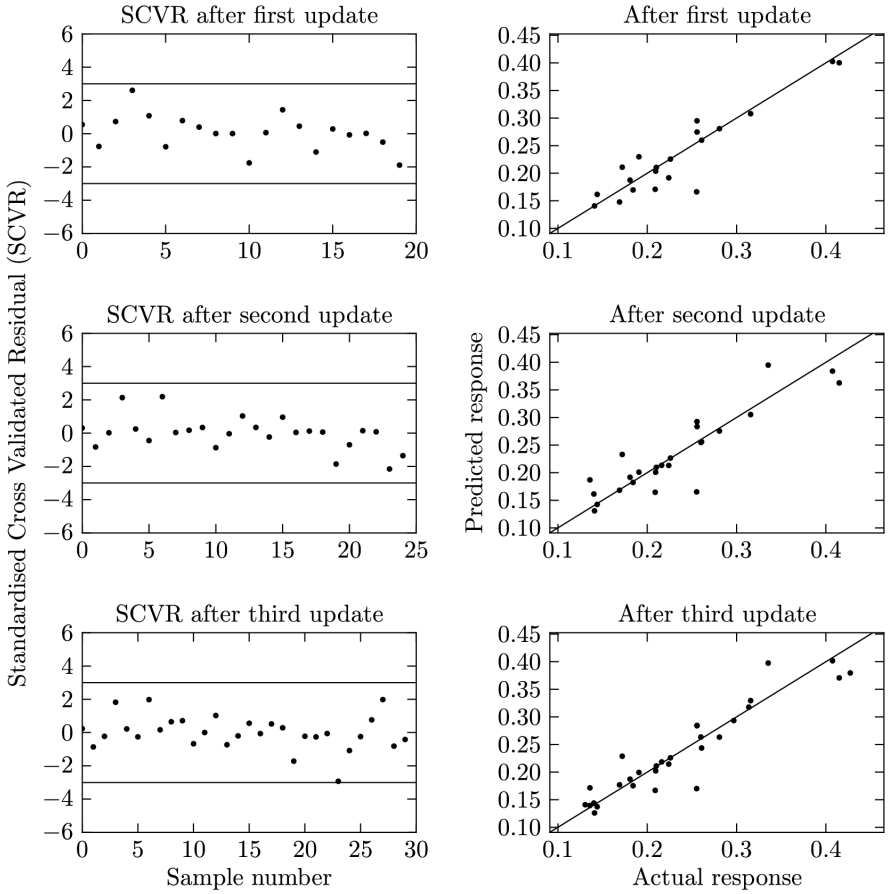


Figure 9.11: SCVR and leave-one-out plots for FM after each update

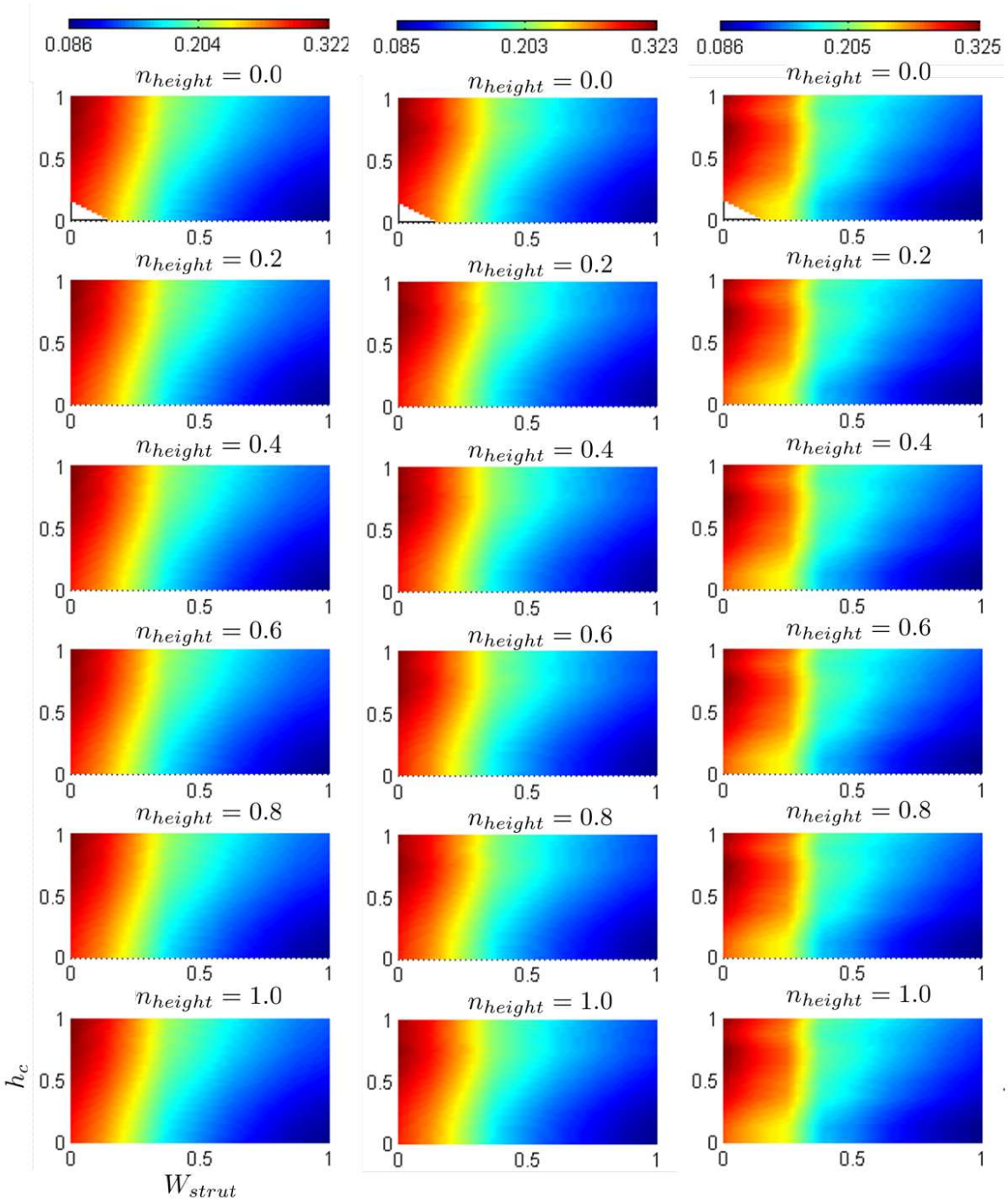


Figure 9.12: Colour plot for *Recoil* after the three updates; first column: update-1; second column: update-2; third column: update-3; each subplot represents W_{strut} on the x-axis, h_c on the y-axis, and n_{height} increases vertically downwards

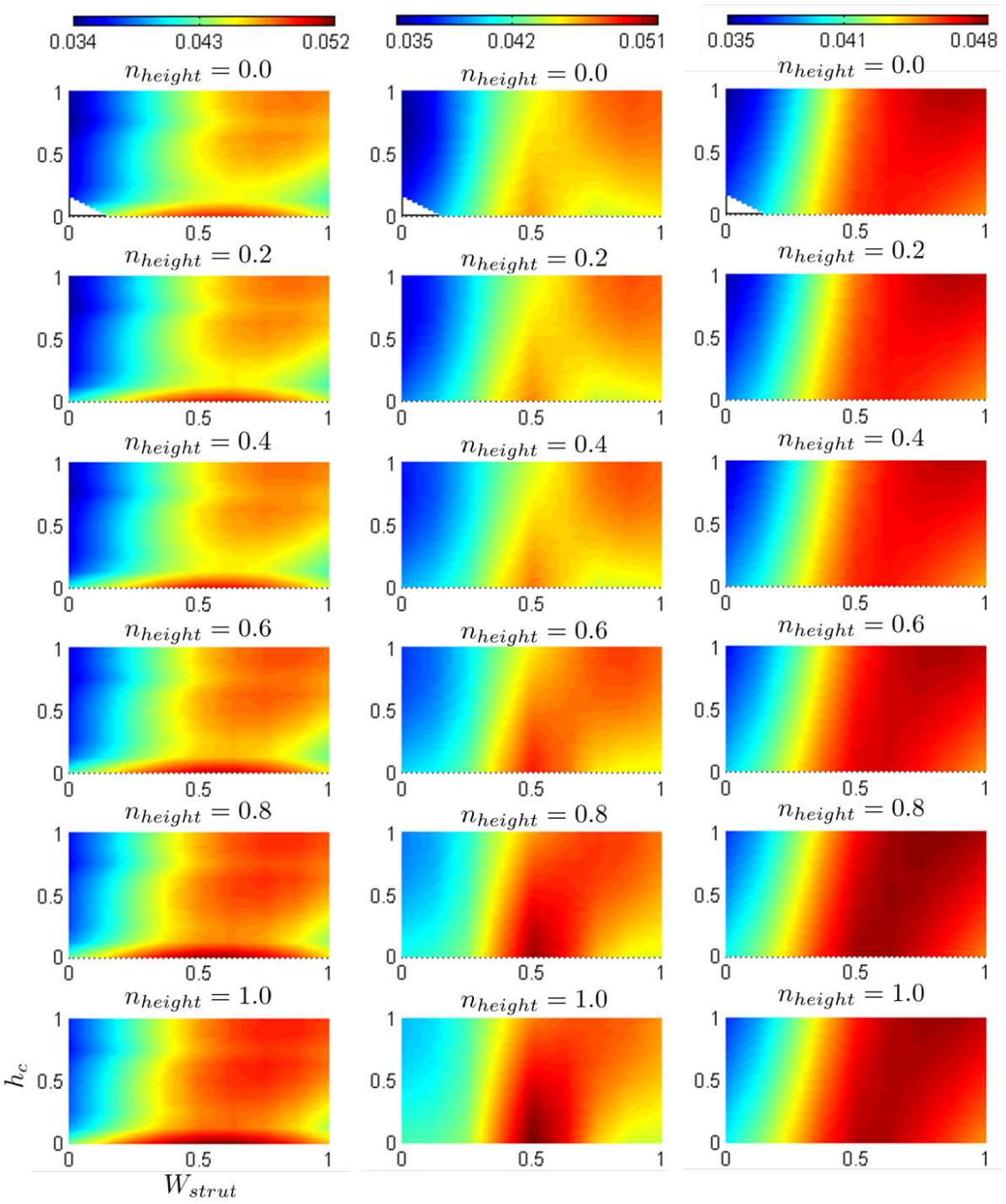


Figure 9.13: Colour plot for *VAS* after the three updates; first column: update-1; second column: update-2; third column: update-3; each subplot represents W_{strut} on the x-axis, h_c on the y-axis, and n_{height} increases vertically downwards

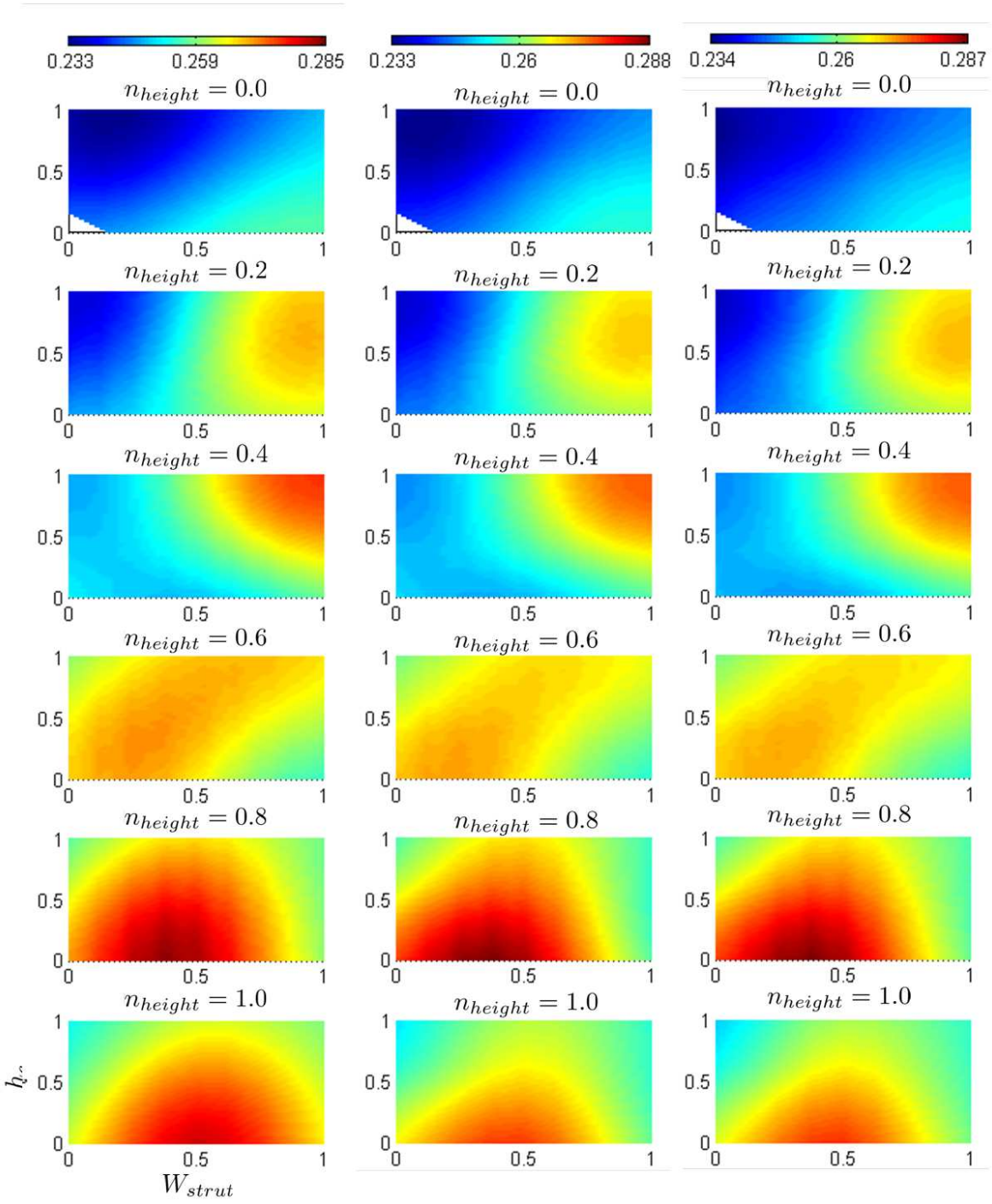


Figure 9.14: Colour plot for *HLRFI* after the three updates; first column: update-1; second column: update-2; third column: update-3; each subplot represents W_{strut} on the x-axis, h_c on the y-axis, and n_{height} increases vertically downwards

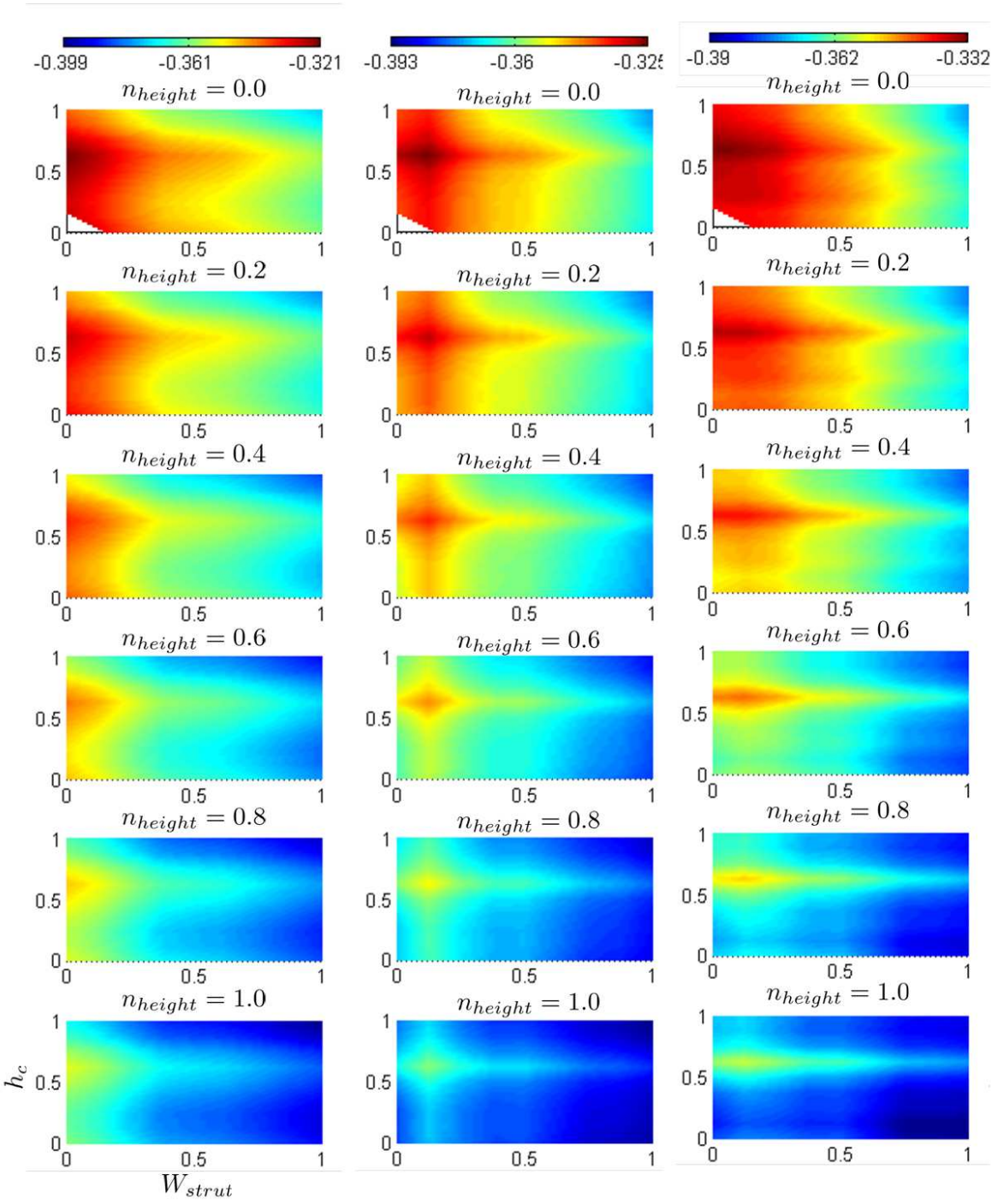


Figure 9.15: Colour plot for $-VAD$ after the three updates; first column: update-1; second column: update-2; third column: update-3; each subplot represents W_{strut} on the x-axis, h_c on the y-axis, and n_{height} increases vertically downwards

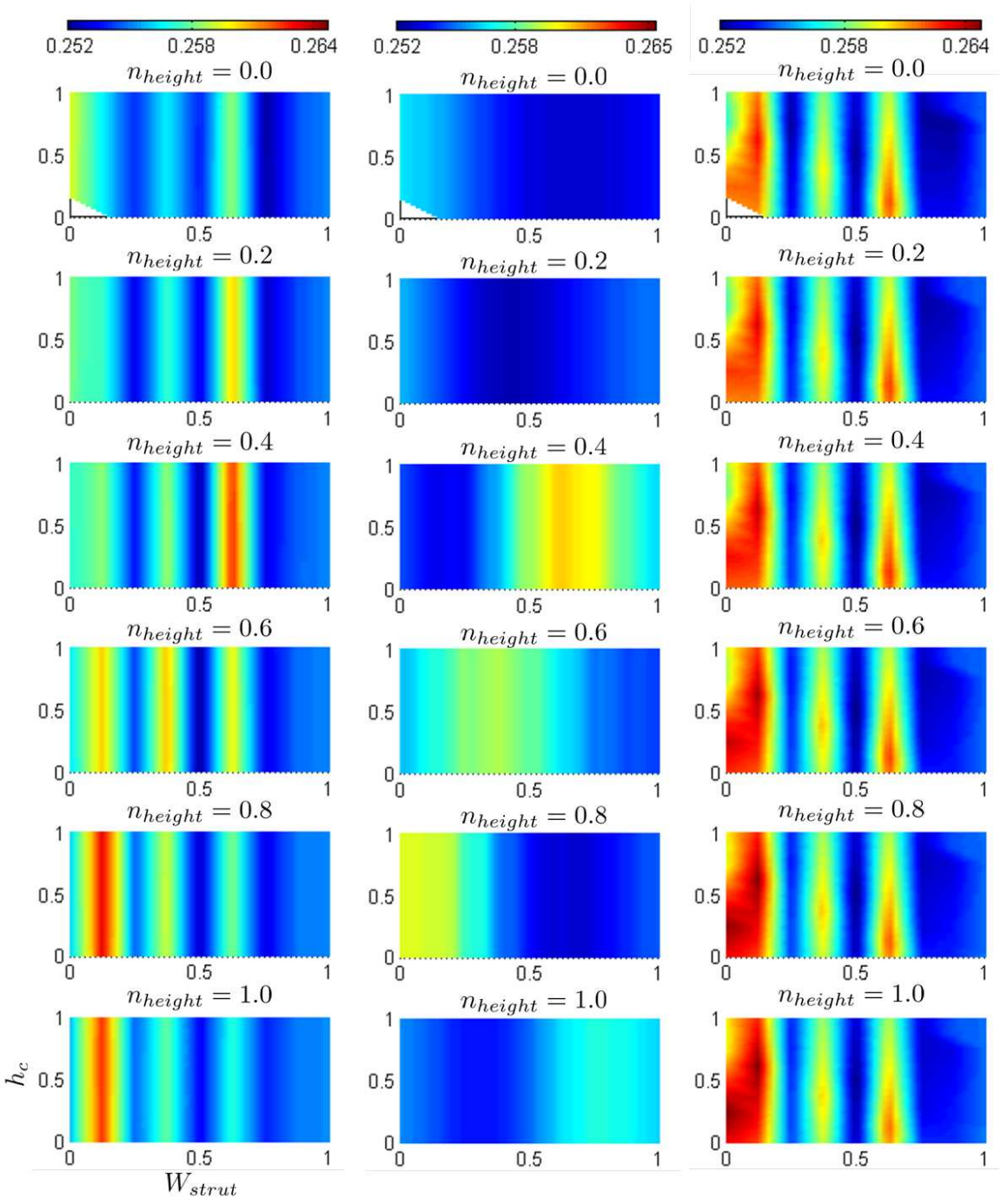


Figure 9.16: Colour plot for D_{dev} after the three updates; first column: update-1; second column: update-2; third column: update-3; each subplot represents W_{strut} on the x-axis, h_c on the y-axis, and n_{height} increases vertically downwards

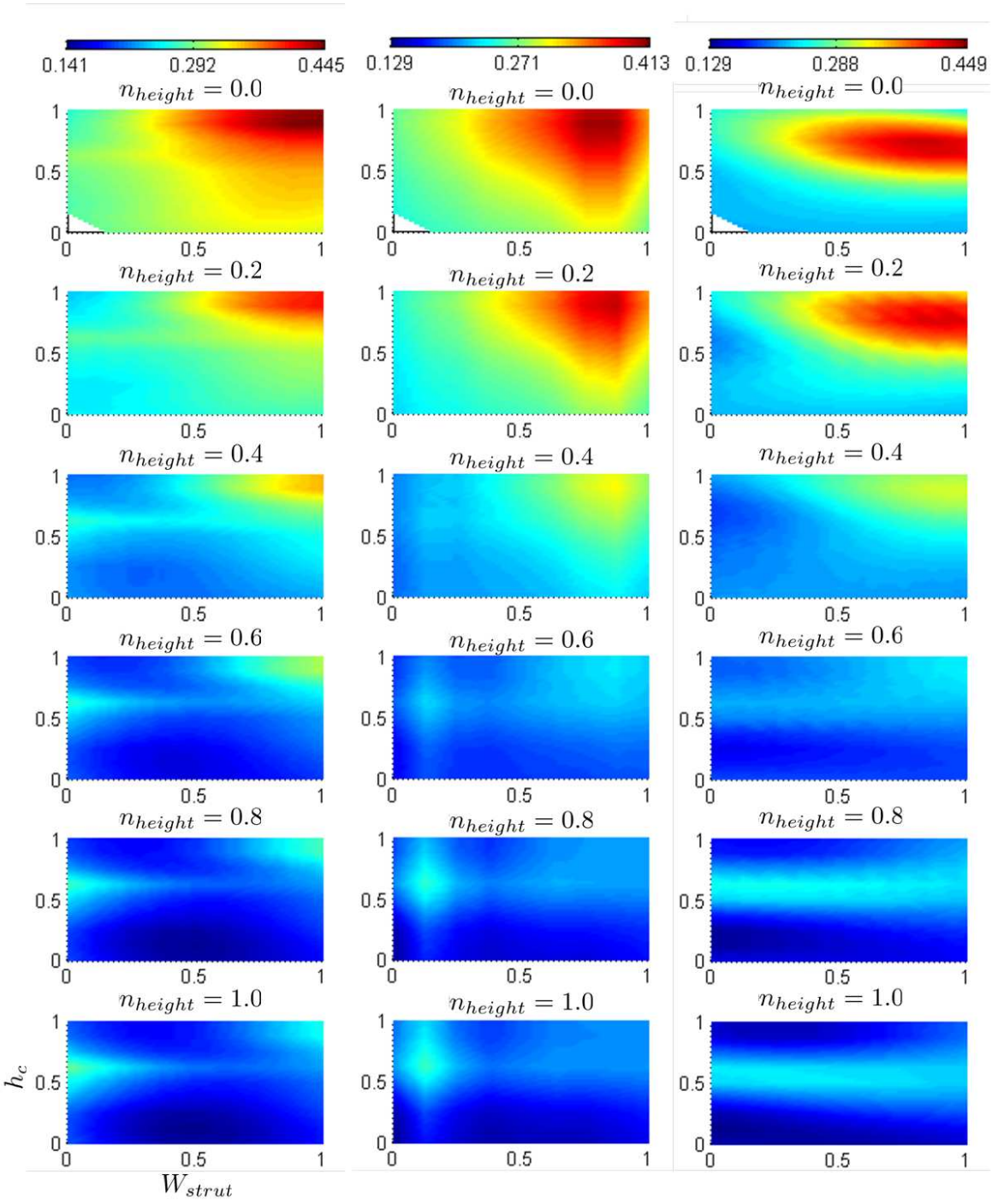


Figure 9.17: Colour plot for FM after the three updates; first column: update-1; second column: update-2; third column: update-3; each subplot represents W_{strut} on the x-axis, h_c on the y-axis, and n_{height} increases vertically downwards

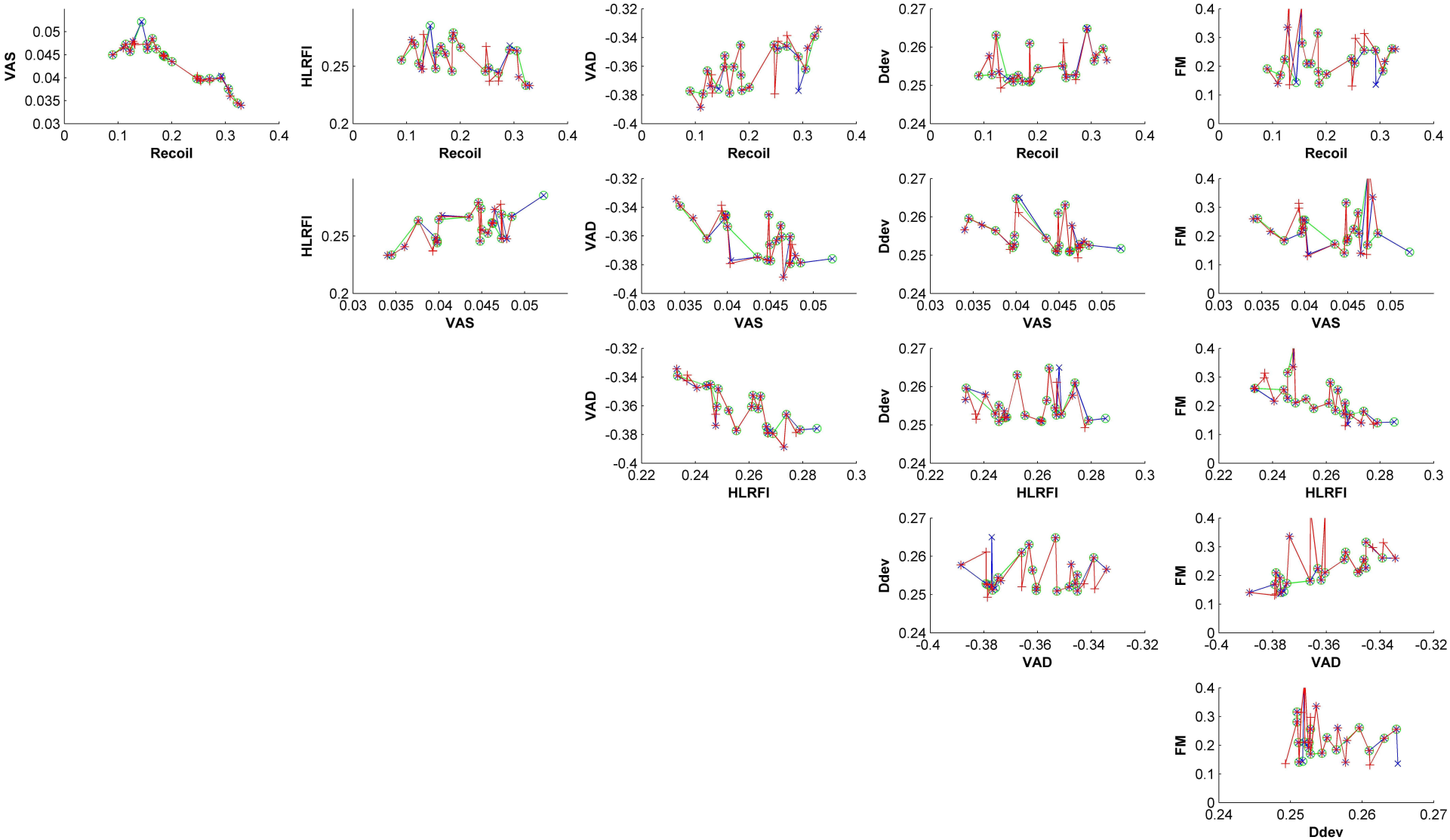


Figure 9.18: Trade-off curves for all combinations of the six metrics: green front indicates Pareto front after fist sample + first update; blue front indicates Pareto front after second update; and red front indicates Pareto front after the end of the third update

9.3.2 Pareto fronts

Figure 9.18 shows slices from the Pareto fronts after each update cycle – green circles showing the pareto front after the initial sample and first update, the blue circles after second update, and the red pluses after the final update. Figure 9.19 shows one of the plots (*VAD* vs. *VAS*) and marks the update points in each update cycle. As can be observed in this figure the update points serve two purposes - first to fill in each successive Pareto front and second to push it towards the bottom left corner.

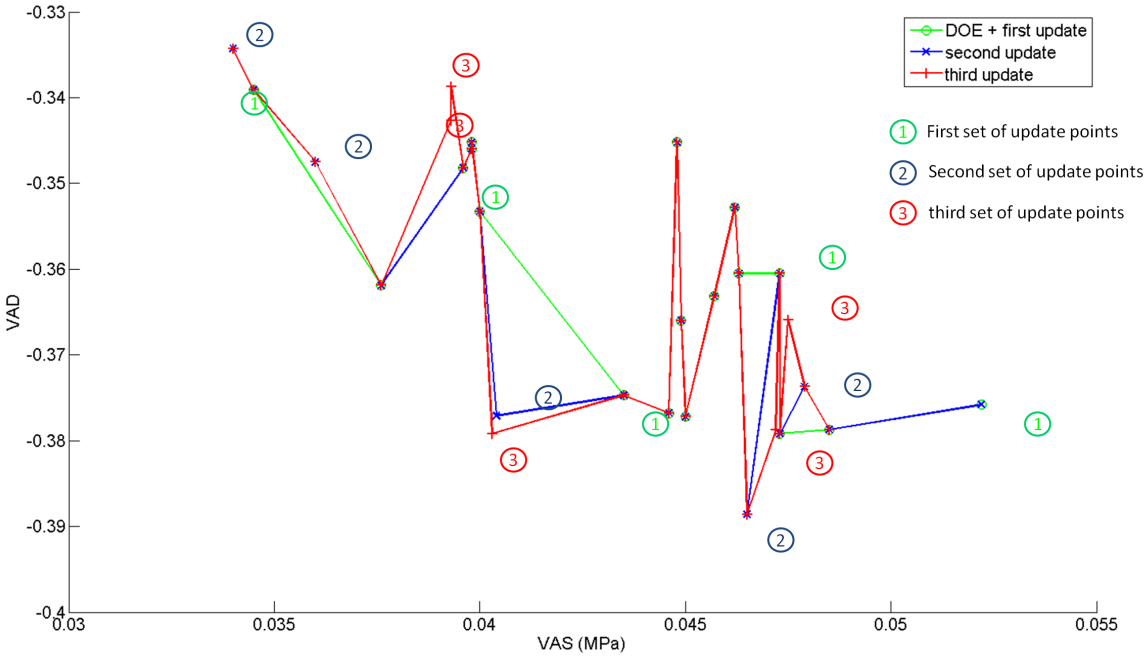


Figure 9.19: The update process - green front indicates the initial sample + first update, the blue front indicates the second update, and the red front indicates the third update

Clear trade offs are observed for the following set of objectives in Fig. 9.18 –

- *VAS* vs. *Recoil*
- *VAD* vs. *VAS*
- *FM* vs. *VAS*
- *FM* vs. *HLRFI*
- *VAD* vs. *HLRFI*
- D_{dev} vs. *VAS*

The following subsections discuss the individual trade-offs in detail–

9.3.2.1 Stress vs. Recoil

Figure 9.20 shows a slice from the final pareto front depicting the trade-off between volume average stress and acute recoil. Designs with lower value of acute recoil show a higher value of *VAS*. This inverse relationship between average stresses and recoil is inevitable. Lower recoil implies a higher lumen area which leads to higher circumferential strains in the artery/tissue

and consequently higher stresses. Although such an inverse relationship is bound to appear in every family of stent designs, it is interesting to note how the stent design (relative to the three design parameters) changes along the Pareto front. As one moves from top-left towards the bottom-right of this front the stent design changes significantly – in particular, the value of h_c increases while the value of W_{strut} decreases. The third parameter, n_{height}, does not follow any particular trend in this plot. This can be explained by the fact that changes in n_{height} result in insignificant changes in both *Recoil* and *VAS* (see last columns of Figures 9.12 and 9.13, respectively). The effect of W_{strut} can be explained by the reasoning that a higher value of W_{strut}, other parameters being equal, implies a higher metal-to-artery ratio (consequently a higher contact area) and hence results in higher stresses. For acute recoil, increasing the value of W_{strut} has the opposite effect. The curved parts of circumferential rings with wider struts undergo higher plastic deformation and consequently resist recoil. Thus an increase in the value of W_{strut} is better in terms of reducing recoil but comes at a price of increased average stresses. The effect of the parameter h_c is not as straightforward

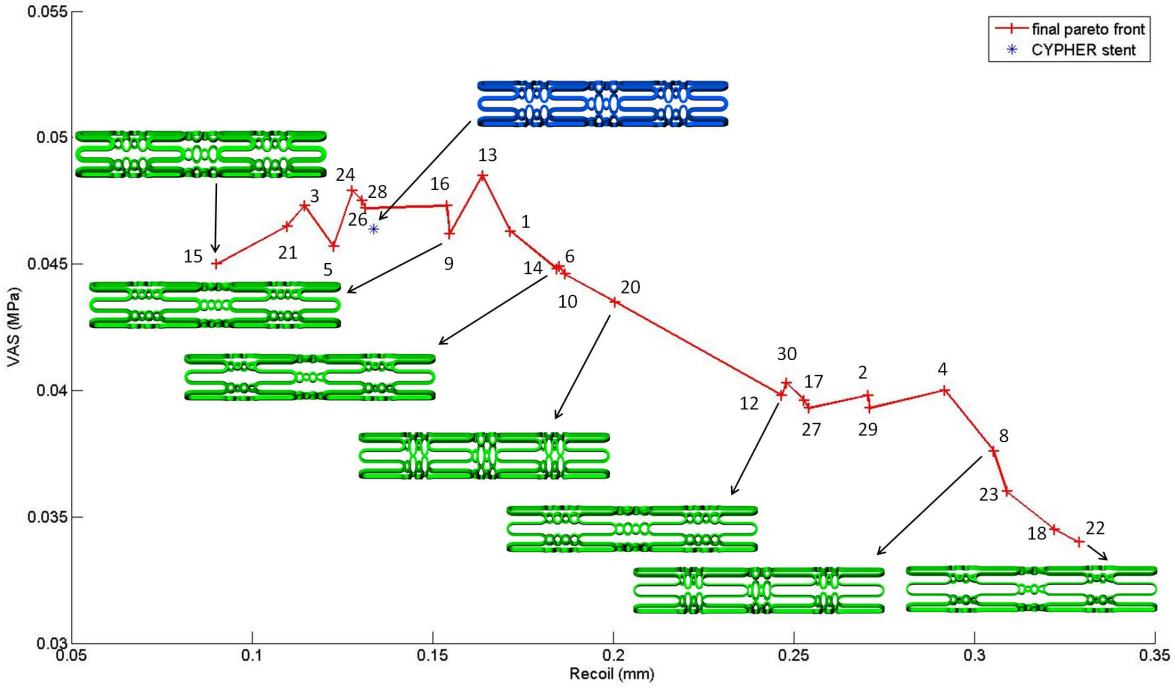


Figure 9.20: Final Pareto front slice showing the trade-off between volume average stress (*VAS*) and acute recoil (*Recoil*)

as the effect of W_{strut}. In terms of recoil a lower value of h_c results in larger plastic strains in the curved regions of the circumferential rings. This can be understood by imagining the unfolded stent on a flat plane and viewing the expansion process as the stretching of these rings (c.f. Figure 7.19). Consequently designs with a lower value of h_c have lower recoil. In terms of the metal-to-artery ratio (contact area between stent and plaque) the effect of h_c depends on n_{height}. Since the circumferential width of the links is constant (0.07 mm) whether a decrease in h_c results in an increase or decrease of the metal-to-artery ratio depends on the net change in area brought by the two competing factors: decrease of area due to the shortening of circumferential rings and increase in area due to the higher axial length that the connectors/links occupy. In general, if the value of W_{strut} is much lower than 0.07 mm and the value of n_{height} is relatively large, then a decrease in h_c will result in a net increase

in the contact area of the stent. However, if the value of W_{strut} is much higher than 0.07 mm and n_{height} is small, a decrease in the value of h_c will result in a net decrease in area. This relationship also explains the behaviour in Figure 9.13 as follows. For the last subplot of the last column, i.e. the plot of *VAS* for a high n_{height} value, consider the region where W_{strut} is lower than 0.07 mm (0.16 in the normalised plot). Here, *VAS* is observed to increase with a decrease in h_c. Contrast this figure with the first plot of the same column; i.e. the plot with a low n_{height} value, and consider the region for of high W_{strut}. Here, a decrease in *VAS* is observed with a decrease in h_c. This effect is reflected in all the columns of the plot: for lower values of W_{strut} decreasing h_c results in increase of *VAS*, while for higher values of W_{strut} decreasing h_c results in a decrease of *VAS*. This complicated relationship is reflected in the observed Pareto front, Figure 9.20, in the following way – when moving from left to right, the initial design changes reflect a decrease in W_{strut} with little change in h_c. This happens until the value of W_{strut} is close to 0.07 mm, after which the design change is largely due to an increase in h_c as opposed to a decrease in the value of W_{strut}.

9.3.2.2 Drug vs. Stress

Figure 9.21 shows a slice from the final Pareto front depicting the trade-off between volume average drug and volume average stress. The designs which lead to a higher volume average drug value (lower value of the *VAD* metric) are the designs associated with higher values of volume average stress. As a general trend along the Pareto front, the designs for better drug response (and consequently worse stress response) have high W_{strut} values, high n_{height} values, and low h_c values. Since both *Recoil* and *VAD* are in competition with *VAS*, intuitively it is

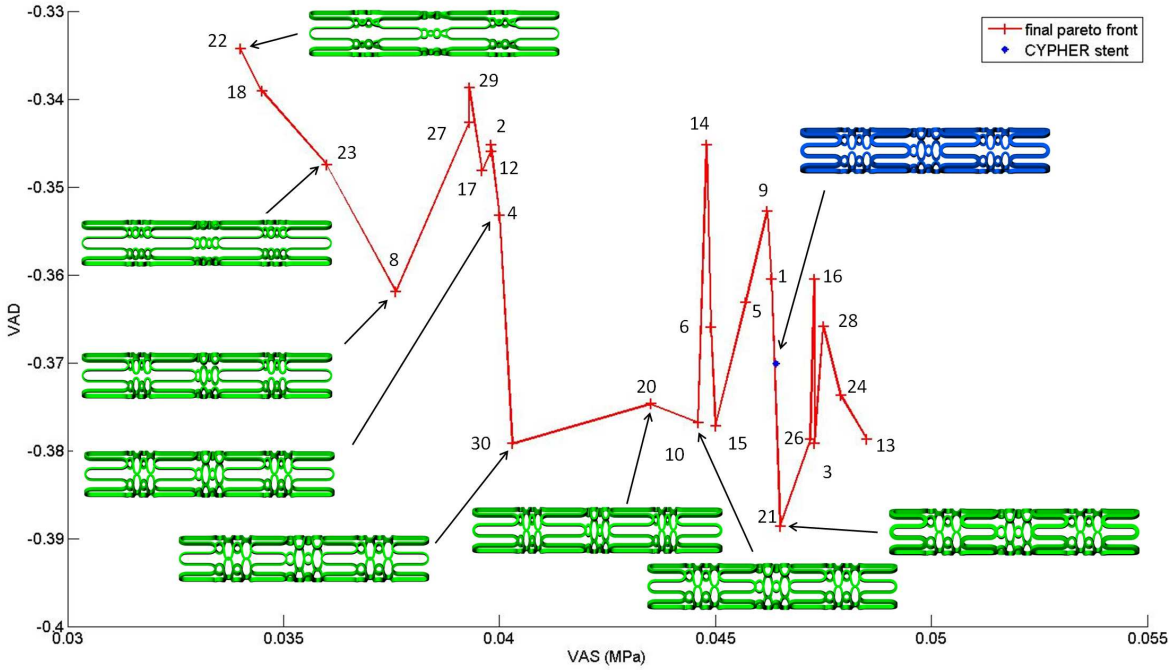
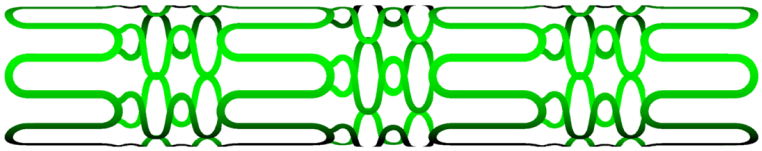


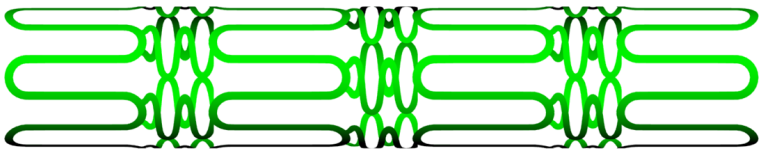
Figure 9.21: Final Pareto front slice showing the trade-off between volume average drug (*VAD*) and volume average stress (*VAS*)

expected that *Recoil* and *VAD* would be positively correlated. This can be seen in Fig. 9.18. As expected *Recoil* and *VAD* follow a general trend of positive correlation. However, there are significant deviations from the general trend. This is primarily due to the fact that while the link height does not play any particular role in determining *Recoil*, it significantly affects

volume average drug (see Figures 9.12 and 9.15). This phenomenon is demonstrated in the best designs for *Recoil*, i.e. DOE_BX_15 and *VAD*, i.e. DOE_BX_21, on the pareto front. The best design for *VAD* has a lower value of W_{strut} and a higher n_{height} value when compared to the best design for *Recoil* on the Pareto front. A longer length of the link provides a better coverage of the artery wall, and since the links provide a source of the drug, the concept of better wall coverage becomes a key contributor in determining *VAD*. One is inclined to relate the contact area of the stent (metal-to-artery ratio) as one of the factors determining *VAD*. Although this is an important factor, the distribution of the struts can not be ignored. This is illustrated by a comparison of designs DOE_BX_10 and DOE_BX_20 in figure 9.22. While DOE_BX_20 has roughly 9% higher contact area when compared to DOE_BX_10, it does not perform better in terms of *VAD*. The distribution of struts in DOE_BX_10 provides a relatively uniform wall coverage and hence has a similar response to DOE_BX_20 in terms of *VAD* despite its lower surface area.



Design: DOE_BX_10
Outer Surface area = 9.02 mm^2
Volume average drug (VAD) = 0.377 units
D_{dev} = 0.251 units



Design: DOE_BX_20
Outer Surface area = 9.86 mm^2
Volume average drug = 0.375 units
D_{dev} = 0.254 units

Figure 9.22: Contact area of a stent and its effect on *VAD*

9.3.2.3 Flexibility vs. Stress

Figure 9.23 shows a slice from the final Pareto front depicting the trade-off between the flexibility metric and volume average stress. The conflict between *FM* and *VAS* is primarily a result of the cross-flow (circumferential) length of the links. While thinner struts are preferred both in terms of volume average stress and flexibility, the competition results from the length of the links, n_{height}. Two desirable characteristics are observed for designs which are relatively more flexible – h_c has lower values and n_{height} has larger values. Smaller values of h_c ensure that the links occupy a larger axial length which in turn delays self contact between the 'n' shaped links as there is more free space between the links. The parameter n_{height} on the other hand gives more length to the links. A longer length of the link is desirable as the bending of the stent can be seen as a stretching and compression of the links. If the links are straight or the value for n_{height} is small, there is little allowance for the links to unfold (while stretching) and squeeze (while compressing), which leads to decreased flexibility. Consequently, the best design in terms of flexibility (DOE_BX_30) has very low value of h_c and the maximum allowed

value (upper bound) for n_{height}. The effect of h_c in Fig. 9.23 can be explained along the same lines as in section 9.3.2.1. Lower values of h_c are not preferred by VAS when W_{strut} is low (owing to the relative effect of area change of the links versus the area of the circumferential rings). In such regions, i.e. when W_{strut} is low, a lower value of h_c is preferred by FM, as it leads to a larger axial length for the links (which results in delayed self contact and hence improved flexibility).

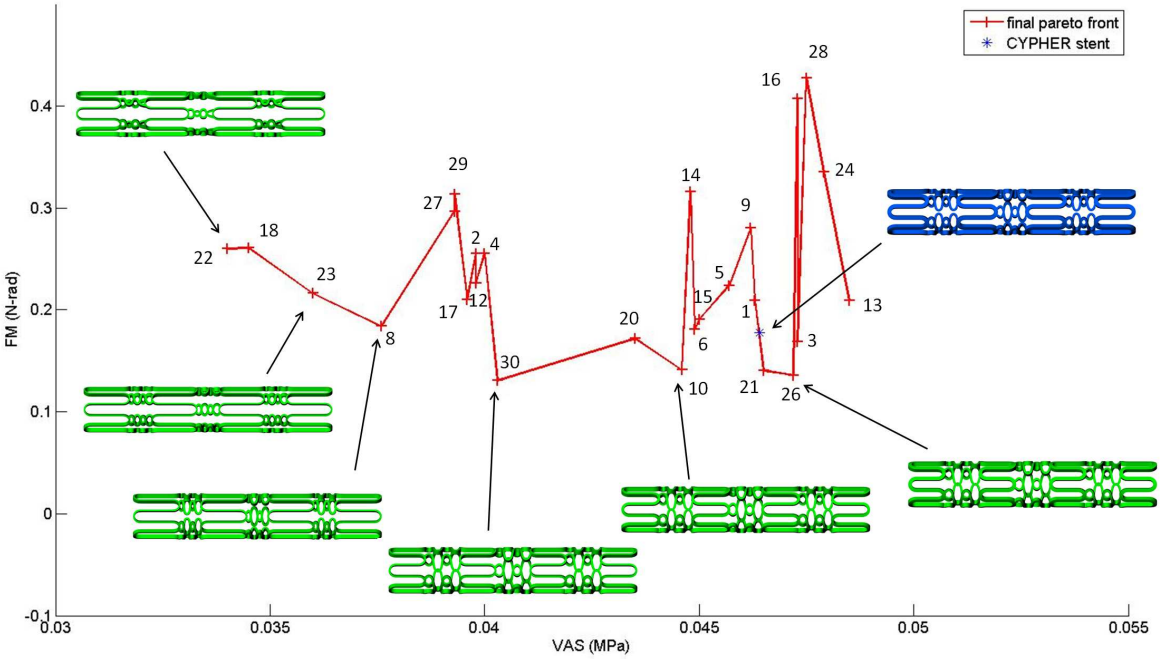


Figure 9.23: Final Pareto front slice showing the trade-off between the flexibility metric (FM) and volume average stress (VAS)

9.3.2.4 Flexibility vs. flow

Figure 9.24 shows a slice from the final Pareto front depicting the trade-off between the flexibility metric and the flow index ($HLRFI$). As a general trend more flexible designs tend to show worse response in terms of haemodynamics. This can be attributed primarily to the length of the links. It was shown in chapter 4 (see section 4.4) that the length of the links in the cross-flow direction (n_{height}) is key in determining the flow response. A direct dependence of HLRFI on n_{height} was shown and it was observed that higher values of n_{height} lead to higher alteration of the haemodynamic features relevant to restenosis, i.e. high $HLRFI$. As discussed in the previous section larger values of n_{height} improve the response in terms of flexibility. This conflict in the parameter n_{height} is the primary reason for the observed trade-off between FM and $HLRFI$ in the Pareto front. In terms of W_{strut}, lower values are preferred by both flexibility and $HLRFI$. Since recirculation zones are formed in between the links, the designs with good $HLRFI$ response tend to have large values of h_c. This minimises the axial length covered by the links and hence leads to a relatively better flow response. However, as discussed in the section 9.3.2.3 this is not favourable in terms of flexibility, as lower axial length for the links implies sooner self-contact with increase in curvature index.

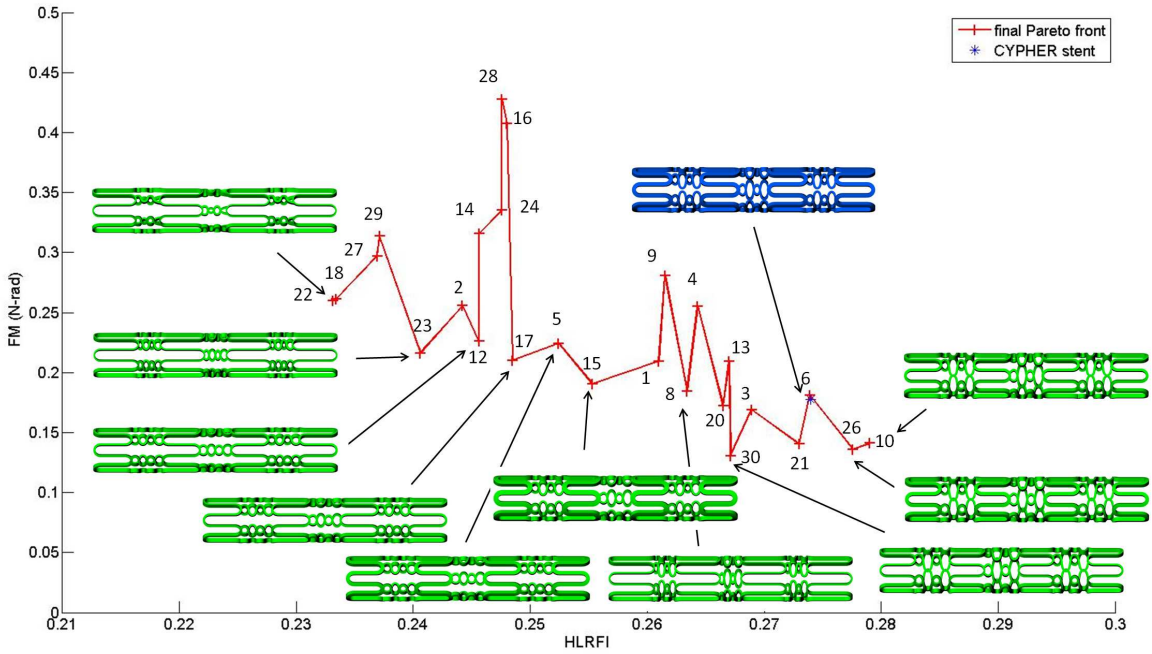


Figure 9.24: Final Pareto front slice showing the trade-off between the flexibility metric (FM) and the flow index ($HLRFI$)

9.3.2.5 Drug vs. flow

Figure 9.25 shows a slice from the final Pareto front depicting the trade-off between the volume average drug and the flow index ($HLRFI$). As one moves from top-left to bottom-right along the front the design changes from low W_{strut}, high h_c, and low n_{height} values to high W_{strut}, low h_c, and high n_{height} values. In terms of W_{strut}, a high value is favourable for VAD so that more drug can be delivered, but the flow metric favours thinner struts as they cause less haemodynamic alteration. Minimisation of flow disturbances within the links drive h_c to higher values so that the axial length occupied by the links is less. Similarly n_{height} is driven to lower values too for improved $HLRFI$. However, these changes in h_c and n_{height} are not preferred from the drug perspective as these result in a) disturbing the uniformity of strut distribution and b) lowering the overall contact area, which is important for drug transport. A strong conflict occurs in these two objectives as all the three parameters drive the objectives in opposite directions.

9.3.2.6 Drug standard deviation vs. stress

Figure 9.26 shows a slice from the final Pareto front depicting the trade-off between the standard deviation of the drug concentration and the volume average stress. The standard deviation of drug concentration is primarily governed by the uniformity of the metal distribution in the stent. The designs in the bottom right corner of the Pareto front have a relatively uniform distribution of struts when compared to the designs in the top-left corner. The conflict in these two objectives is a result of the fact that a design which covers the artery wall in a uniform manner generally has a higher metal-to-artery ratio, which leads to a higher contact area and consequently higher contact stresses.

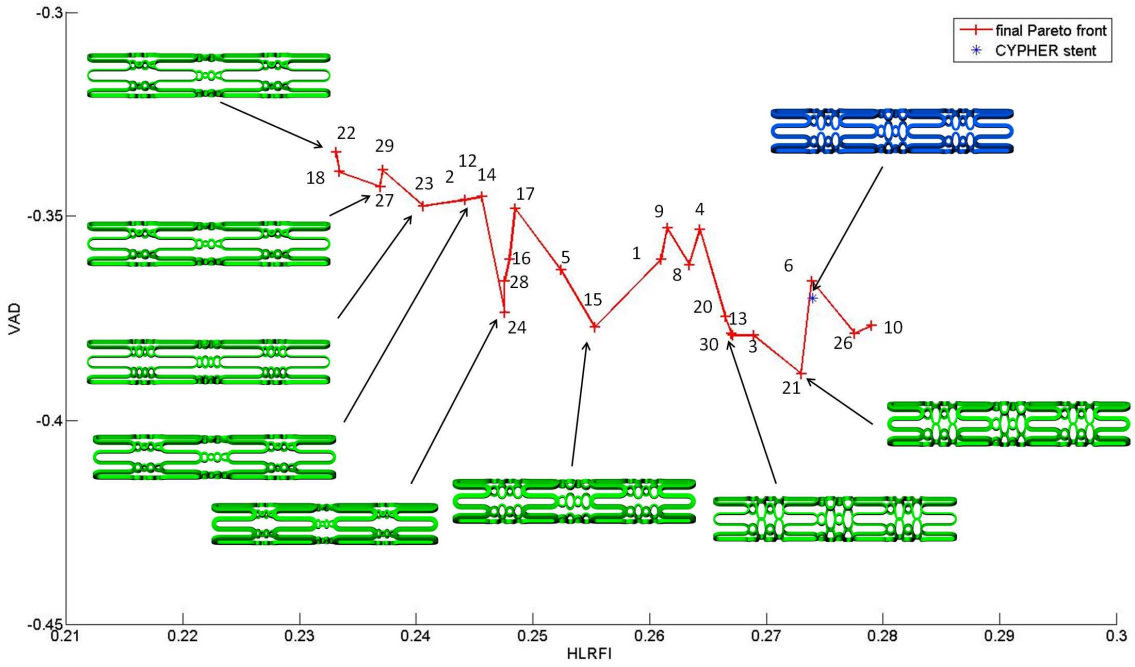


Figure 9.25: Final Pareto front slice showing the trade-off between volume average drug (*VAD*) and the flow index (*HLRFI*)

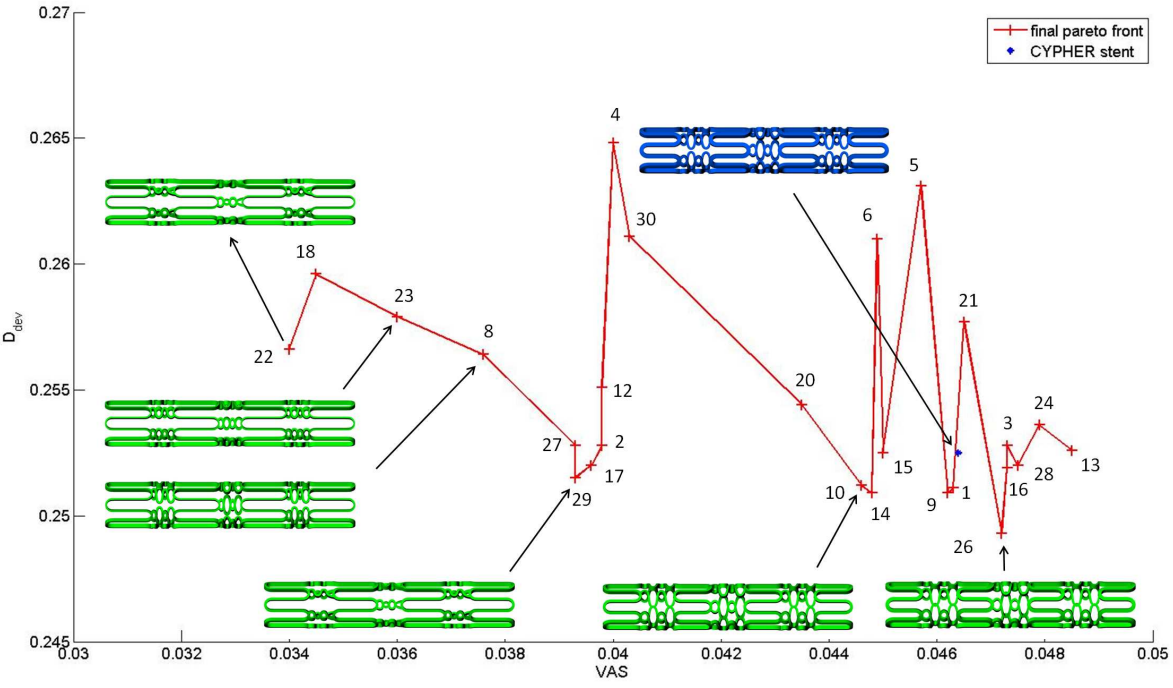


Figure 9.26: Final Pareto front slice showing the trade-off between standard deviation of drug concentration (D_{dev}) and the volume average stress (*VAS*)

9.3.2.7 Other plots

The other sub-plots in Figure 9.18 which have not been discussed in the above sections do not display any conclusive trade-off between the objectives. The behaviour of most of these plots can be partly explained by the reasoning given in the above sections. For instance, section 9.3.2.3 explains the trade-off between FM and VAS, and section 9.3.2.4 explains the trade-off between FM and $HLRFI$. Since both VAS and $HLRFI$ both are in conflict with FM, they are bound to show some positive correlation. Similar reasoning holds for the other plots too.

The plots with D_{dev} as one of the objectives do not show any particular trend and hence lead to no particular conclusions (except for the plot discussed in section 9.3.2.6). As discussed in section 9.3.1, this could be because D_{dev}, as defined by Eq. 8.6, is not a particularly good measure for the uniformity of the drug distribution in a stented artery.

9.4 Choosing the ideal stent

The choice of an ideal stent, even after obtaining the Pareto front, is not trivial. Owing to the multiple number of desirable characteristics, computational studies can only lead to the non-dominated (Pareto) designs. Several approaches can be taken to identify designs which could potentially be considered ideal –

9.4.1 A conservative paradigm

A conservative design approach, assuming no other information (clinical or otherwise) is available to judge individual designs in the non-dominated family, is to remain in the middle region of the Pareto front, thereby avoiding poor performance in any of the objectives. Using this approach designs DOE_BX_20 and DOE_BX_30 stand out (cf. Figs 9.20, 9.21, 9.23, 9.24, 9.25, and 9.26). When compared with the CYPHER stent, which also lies on the Pareto front but is usually skewed towards one end in all the Pareto slices, these designs provide a more balanced trade-off between the various merits.

Of particular interest, besides the designs DOE_BX_20 and DOE_BX_30, are designs DOE_BX_10 and DOE_BX_15. While both these designs perform well, under the conservative paradigm discussed above, in almost all objectives, DOE_BX_10 performs relatively poorly in terms of $HLRFI$ and DOE_BX_15 in terms of VAS.

9.4.2 A constraint based paradigm

This approach utilises additional information based on clinical guidance. For instance, a maximum required value of FM (minimum flexibility of a stent) could be assigned based on the maximum curvature of the stent deployment path. Similarly, a maximum value for the recoil could be assigned based on the required minimum lumen area; and dependent on the biochemistry of the drug being delivered, a minimum value of VAD could be ascertained. Once such limits are determined, the design choice can be narrowed down to a handful of designs satisfying such constraints. However, the determination of such limits is not an easy task and further statistical research is required, to determine them especially when account-

ing for patient variability.

A constraint optimisation approach for stent design, where the concept of assigning one or more of objectives as constraints was demonstrated in chapter 7. It was assumed that the limits of *Recoil*, *VAS*, *VAD*, and *FM* were set by the values of the baseline design, and improvement in design was sought for one objective at a time. Since the two parameters, W_{strut} and h_c, are similar to the ones used in that chapter, a comparison can be made between the results of chapter 7 and the findings of this chapter. Figure 9.27 is used for this purpose. First, consider the problem of minimizing the *VAS*, relative to baseline (DOE_BX_1 in this case), without any decrease in *Recoil*, *VAD*, and *FM*. In the first plot of Fig. 9.27 one can draw a vertical line passing through DOE_BX_1, Line-1. All designs to the left of this line, viz. designs 3, 5, 9, 13, 15, 16, 21, 24, 26, and 28, are designs which do not have worse recoil relative to design 1. Similarly, a horizontal line through DOE_BX_1, Line-2, can be drawn. All designs below this line do not have worse *VAS* values relative to DOE_BX_1. Hence, from this plot, it can be seen that only designs 15, 5, and 9, show an improvement in *VAS* without compromising *Recoil*. Similar vertical and horizontal lines can be drawn in the second and third plots of Figure 9.27 where designs in the bottom-left quadrant are feasible designs. Of the designs selected from the first plot, listed above, only design 15 satisfies the feasibility criteria in the other two plots. A similar exercise can be performed for the other two objectives of *VAD* and *FM*. For minimizing *VAD* and *FM*, designs 21 and 26 are identified to be the best, respectively, even though they are marginally on the wrong side of the vertical lines in the second and third plots. This is not unreasonable as the parameters p_1 and p_2 in the chapter 7 are not equivalent to just one n_{height} parameter used in this chapter. However, the qualitative agreement of designs 15, 21, and 26 for the corresponding optimum solutions obtained for minimising *VAS*, *VAD*, and *FM* in chapter 7 is convincing – high values of W_{strut} combined with low values of h_c for minimising *VAS* and *VAD*, and mid-range values of W_{strut} and low values of h_c for maximising *FM*.

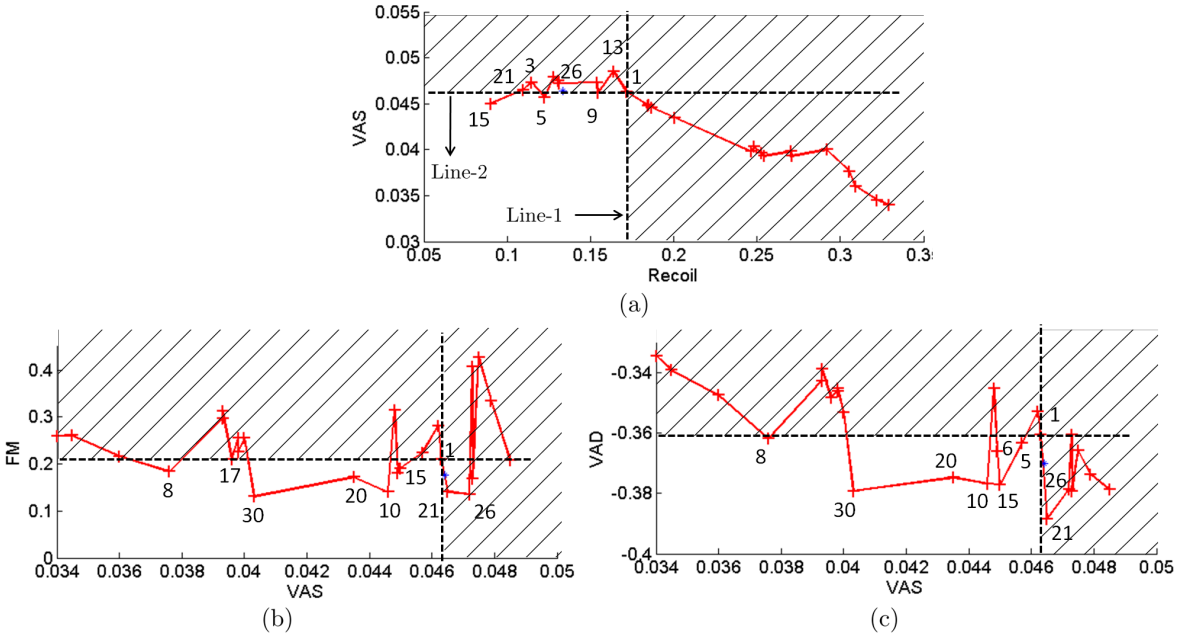


Figure 9.27: Constraint based approach to pick designs from the Pareto plots; comparison with the study of chapter 7

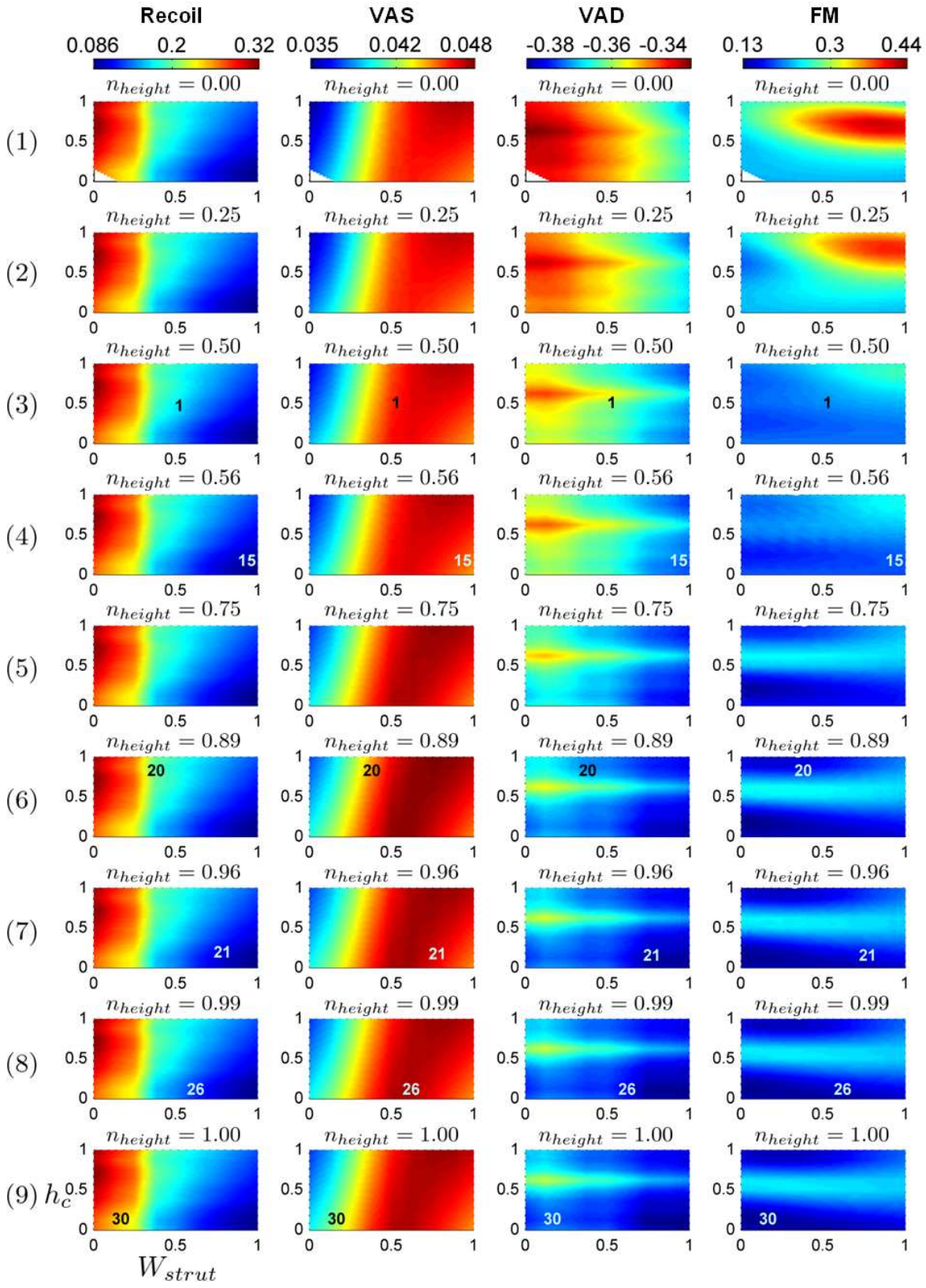


Figure 9.28: Response surface contour plots for the four objective functions, *Recoil*, *VAS*, *VAD*, and FM, after the third update (all design parameters are normalised): each subplot has W_{strut} on x-axis, h_c on y-axis, and the parameter n_{height} increases vertically downwards

The above stated mechanics of finding optimal designs in consideration of constraints can also be seen in the response surfaces for various objectives, as show in Figure 9.28. This figure shows contour slices, of the four objective functions viz. *Recoil*, *VAS*, *VAD*, and *FM* against W_{strut} (on the x-axis of each sub-plot) and h_c (on the y-axis of each sub-plot) at different values of n_{height} (aligned vertically). The n_{height} values chosen include two sets – one at a constant interval of 0.25 units between 0 and 1 (in the non-dimensional space) to illustrate the general trend of the response surfaces, and the other to include slices for the designs discussed before, i.e. 1, 20, 30, 15, 21, and 26 (the positions of the numbers correspond to the values of the design parameters). For constrained optimisation of *VAS* with respect to the baseline geometry, consider rows 3 and 4 (numbered from 1) of the figure which show designs 1 and 15, respectively. From the *VAS* plot of row 3, it can be observed that major improvement in *VAS*, with respect to design 1, can be achieved by moving to the left side of the plot (i.e. going from the red region to the blue region). The *FM* plots favour this response as the left-top and left-bottom regions of the plots ($n_{height} > 0.5$) show no worse value of *FM* relative to design 1. However, the *Recoil* and *VAD* plots reveal that movement to the left will compromise *Recoil* and *VAD*. Hence a move to the right remains the only choice. Moreover, a move to the right-bottom corner, region where design 15 lies, is the only region which shows *VAS* improvement without compromising other objectives. Similar reasoning can be applied for designs such as 21 and 26 which closely resemble the constrained optimal designs for *VAD* and *FM* found in chapter 7.

The contour plots of Figure 9.28 also show the location of the optimal designs in the conservative paradigm, viz. designs 20 and 30 in rows 6 and 9, respectively. These designs lie in the yellow/blue regions of all the objectives, thereby not taking extreme values for any of the objective under the proposed paradigm.

9.4.3 The experimental approach

This approach requires experimental guidance. A few designs across the Pareto front can be picked and tested *ex-vivo* in a laboratory, known as *bench-testing*. Such experimental results can then guide the process of choosing the optimal stent. Furthermore, leaving the ethics and morality of the process aside, testing in animal models can also be used to guide such an experimental process.

Notwithstanding the merit of the above discussion, on a philosophical note it seems rather unreasonable to think of one stent design as 'ideal' given the differences in the lesion geometry, morphology, and other inter-patient variability found inherently in the human population. A more logical approach, though extremely challenging in terms of computational, technical, and practical aspects, would be to perform patient-specific optimisation studies using magnetic resonance imaging (MRI)/ angiography/intravascular ultrasound (IVUS) data to obtain both the geometry and morphology of the specific lesions. Another approach, midway between the one-size-fits-all approach and patient-specific optimal designs, could be to classify lesions into manageable number of classes, and find optimal designs, most likely to be different, for each class separately.

A final note should be mentioned here with regards to the Krig-update methodology adopted in this chapter. As mentioned in section 9.2, various other update strategies, can be

adopted for choosing update points. However, on account of the extremely high evaluation times needed for evaluation of all the objectives, a comparison of the performance of such methods on the stent design problem is out of the scope of this thesis, and forms one area of future work. Similarly, the results of this chapter show that not all pairs of objectives are in conflict with each other. The apparent correlation between some objectives can be used to decrease the number of objectives to less than six and consequently result in a more efficient optimisation study.

9.5 Conclusions

This chapter proposes a three-parameter technique to vary the design of the widely known CYPHER stent. Six figures of merit (numeric indicators of a stent's efficacy) viz. acute recoil, volume average stress, flow index, volume average drug, drug uniformity, and flexibility, are formulated. A surrogate modelling technique coupled with NSGA-II is employed to obtain the Pareto front showing the trade-off between different sets of the six figures of merit. The effect of the three parameters on these metrics is also studied. It is demonstrated that a change in one parameter that leads to an improvement in one of the objectives often leads to a compromise in one or more of the other objectives. It is found that while strut width and the length of the circumferential rings most affect volume average stress and recoil, the length of the links in the cross-flow direction significantly affects volume average drug, flexibility, and the flow index. The complex interplay between stent design (distribution of struts, link design, strut thickness, and circumferential ring design) and stent performance, from the perspective of the various conflicting/desirable properties, is clearly shown. Despite this complex interplay, the non-dominated solutions, which represent a potentially optimum family of CYPHER like stents, for the proposed parameterisation are obtained and discussed. Moreover, several approaches for selecting optimal designs are identified and a parallel has been shown between the constraint optimisation study presented in chapter 7. In particular, designs 20 and 30 are identified as optimal in terms of all the objectives in a conservative paradigm, and their relative position with respect to a representative CYPHER stent is shown. Finally, in a constraint based approach, designs 15, 21, and 26 are identified as designs showing maximal improvement, from the baseline geometry, in the corresponding chosen objectives. In essence, a methodology to perform design optimisation studies on stents and the process of choosing different stent designs appropriate to different needs is presented.

This chapter culminates the process of finding an optimal family of coronary stent designs, by combining all the stent analysis methods developed in previous chapters in a single multiobjective design study. The next chapter is not specific to the problem of coronary stent design. It looks at Kriging-assisted optimisation at a more fundamental level, and explores how derivative information in a Krig can be used for efficient search of the Krig. This underlying idea emerged during the process of learning and applying the Gaussian Process models as described so far.

Chapter 10

An Optimisation algorithm that exploits derivative information in Kriging

In this chapter, the use of derivative information in a Kriging predictor, which is available analytically, but not hitherto used in any known optimisation method, is explored to propose an optimisation algorithm that can be used for both global and multimodal optimisation of Kriging predictors. This chapter has stemmed from the process of learning and applying Gaussian process models for the problem of stent design, and the first principles belief that if the analytical form for a function is available and differentiable, then it makes sense to equate the analytical expression of the derivative to zero, and explore if any useful information can be obtained from such an expression.

10.1 Motivation

The motivation for this chapter stems from the realisation that Kriging predictors can be differentiated, and the fact that, to the best of author's knowledge, this gradient information has not been explored to be used in optimisation algorithms. Moreover, the exciting aspect of using derivative information of a Krig is that no matter what underlying phenomenon is being modelled by the Kriging predictor, the mathematical form of the Kriging predictor remains the same. Hence, if an algorithm were to use this mathematical form to calculate the derivatives, the need to code different derivatives for different functions is eliminated. Lastly, the derivative information, if derived analytically, is available in exact form and need not be approximated by finite differencing methods.

Based on these motivations, in this chapter the following tasks are performed:

1. An expression for the derivative of a Kriging predictor is derived.

2. The above expression is manipulated to yield a fixed-point iterative sequence in order to find the stationary points of the Kriging predictor.

3. The convergence of such a fixed-point iterative sequence is explored.

4. Modifications to this fixed-point iterative sequence and a Newton-Raphson equivalent of the sequence in order to improve convergence is proposed.

5. Based on the results of the above, a new algorithm for finding stationary points of a Kriging predictor is proposed.

6. The proposed algorithm is compared for global and multimodal optimisation with a standard genetic algorithm, a fitness sharing genetic algorithm, and a dynamic hill climbing algorithm.

It must be noted that this chapter does not concern methods which improve the accuracy of the Krig during the process of optimisation, such as the EGO algorithm [50]. This chapter deals with the case when a reasonably accurate Krig has been constructed, although it is applicable to any Krig regardless of its accuracy, and there is a need to search this Krig, either for global optimisation, or for multimodal optimisation.

10.2 Derivative of the Kriging predictor

The formulation of a Gaussian Process predictor is presented in section 3.2.3.1. This section continues from there, and uses the same notation. To summarise, the equations for the correlation function used, GP predcitor (i.e. posterior mean), and posterior variance are given by equations 3.22, 3.35/3.38, and 3.37, respectively.

In this section the derivative of the Kriging predictor is calculated. Henceforth, the discussion is limited to cases where the parameter m_j in equation 3.22 is equal to 2, implying the prior belief that the engineering functions involved are infinitely differentiable.

If $\{X(t), t \geq 0\}$ is a stochastic process, then the derivative of the process, $X'(t)$, is defined as

$$X'(t) = \lim_{h \to 0} \frac{X(t+h) - X(t)}{h}. \tag{10.1}$$

The above limit exists in a mean square sense if the derivative of the mean function of the stochastic process, $m(t)$, exists and the mixed second derivative, i.e.

$$\frac{\partial^2}{\partial s\, \partial t} Cov[X(s), X(t)] \tag{10.2}$$

exists and is continuous [49]. When these conditions are satisfied, as in the case of the Kriging predictor, the derivative is a linear operator. This implies that the 'mean of the derivative' and the 'derivative of the mean' are equal and can be interchanged [194] , i.e.

$$E[X'(t)] \;=\; E\left[\frac{d}{dt}X(t)\right] \;=\; \frac{d}{dt}E[X(t)] \;=\; m'(t), \tag{10.3}$$

where $E[\cdot]$ is the expectation operator.

Applying the above to the Kriging predictor of equation 3.35, the mean of the derivative at a point, $\mathbf{x}$, can be written as

$$\hat{y}'(\mathbf{x}) \;=\; \frac{d}{d\mathbf{x}}\left(\beta + \mathbf{r}(\mathbf{x})^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\beta)\right) \tag{10.4}$$

or, following equation 3.38, since $\mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\beta)$ does not depend on $\mathbf{x}$,

$$\hat{y}'(\mathbf{x}) \;=\; \frac{d}{d\mathbf{x}}\left(\beta + \mathbf{r}(\mathbf{x})^T \mathbf{w}\right). \tag{10.5}$$

Since $\mathbf{x} \in \mathbb{R}^n$, the above equation represents a set of n equations, i.e.

$$\frac{\partial \hat{y}(\mathbf{x})}{\partial x_i} = \frac{\partial}{\partial x_i}\left(\beta + \mathbf{r}(\mathbf{x})^T \mathbf{w}\right) \quad ; \quad i = 1 \ldots n. \tag{10.6}$$

Next, the i^{th} equation from the above set is expanded, to yield

$$\frac{\partial \hat{y}(\mathbf{x})}{\partial x_i} = \frac{\partial}{\partial x_i}\left(\mathbf{r}(\mathbf{x})^T \mathbf{w}\right) \tag{10.7}$$

$$\frac{\partial \hat{y}(\mathbf{x})}{\partial x_i} = \left(\frac{\partial}{\partial x_i}\mathbf{r}(\mathbf{x})\right)^T \mathbf{w} \tag{10.8}$$

$$\frac{\partial \hat{y}(\mathbf{x})}{\partial x_i} = \left[\frac{\partial R(\mathbf{x}, \mathbf{x^{(1)}})}{\partial x_i} \quad \frac{\partial R(\mathbf{x}, \mathbf{x^{(2)}})}{\partial x_i} \quad \cdots \quad \frac{\partial R(\mathbf{x}, \mathbf{x^{(p)}})}{\partial x_i}\right]\mathbf{w}. \tag{10.9}$$

Since $m_j = 2$, using equation 3.22, the above equation can be written as

$$\frac{\partial \hat{y}(\mathbf{x})}{\partial x_i} = \begin{bmatrix} -2\theta_i\left(x_i - x_i^{(1)}\right) R(\mathbf{x}, \mathbf{x^{(1)}}) \\ -2\theta_i\left(x_i - x_i^{(2)}\right) R(\mathbf{x}, \mathbf{x^{(2)}}) \\ \vdots \\ -2\theta_i\left(x_i - x_i^{(p)}\right) R(\mathbf{x}, \mathbf{x^{(p)}}) \end{bmatrix}^T \mathbf{w} \tag{10.10}$$

$$\frac{\partial \hat{y}(\mathbf{x})}{\partial x_i} = -2\theta_i \begin{bmatrix} \left(x_i - x_i^{(1)}\right) \\ \left(x_i - x_i^{(2)}\right) \\ \vdots \\ \left(x_i - x_i^{(p)}\right) \end{bmatrix}^T \begin{bmatrix} R(\mathbf{x}, \mathbf{x^{(1)}}) & 0 & \cdots & 0 \\ 0 & R(\mathbf{x}, \mathbf{x^{(2)}}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R(\mathbf{x}, \mathbf{x^{(p)}}) \end{bmatrix} \mathbf{w} \tag{10.11}$$

$$\frac{\partial \hat{y}(\mathbf{x})}{\partial x_i} = -2\theta_i \, \mathbf{D_i}(\mathbf{x}) \, \mathbf{Z}(\mathbf{x}) \, \mathbf{w} \quad ; \quad i = 1 \ldots n, \tag{10.12}$$

where

$$\mathbf{D_i}(\mathbf{x}) = \begin{bmatrix} \left(x_i - x_i^{(1)}\right) & \left(x_i - x_i^{(2)}\right) & \cdots & \left(x_i - x_i^{(p)}\right) \end{bmatrix} \in \mathbb{R}^{1 \times p}, \tag{10.13}$$

$$\mathbf{Z}(\mathbf{x}) = \begin{bmatrix} R(\mathbf{x}, \mathbf{x^{(1)}}) & 0 & \cdots & 0 \\ 0 & R(\mathbf{x}, \mathbf{x^{(2)}}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R(\mathbf{x}, \mathbf{x^{(p)}}) \end{bmatrix} \in \mathbb{R}^{p \times p} \tag{10.14}$$

and

$$\mathbf{w} = \mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\beta) \quad \in \mathbb{R}^{p \times 1}. \tag{10.15}$$

This completes the expression for the derivative of the Kriging predictor, where the n partial derivatives are given by equation 10.12. In the following sections, the use of this derivative in optimisation is explored.

10.3 Fixed point (FP) iterative scheme to find stationary points of a Krig

One of the tasks in the realm of optimisation is to locate the stationary points of a function. These are points where the derivative vanishes, and hence are either a local maximum, local minimum, or a saddle point, depending on the behaviour of the second derivative at that

point. In order to find the stationary points of a Krig, one can set equation 10.12 to zero and solve for $\mathbf{x}$, i.e. solve the following equation

$$-2\theta_i \, \mathbf{D_i(x)} \, \mathbf{Z(x)} \, \mathbf{w} \;=\; 0 \quad ; \quad i = 1 \ldots n. \tag{10.16}$$

It is clear that an analytical expression for the above set of equations does not exist. However, it is noticed that if this set of equations is written in a suitable form, then a fixed point iterative scheme can be employed to find a solution. In the following two sections, first the principles behind a fixed point iterative scheme are presented, and then such a scheme for solving equation 10.16 is deduced.

10.3.1 Fixed point iteration

A *fixed point* (FP) of a function, $f(x)$ is a point that maps the function onto itself, i.e. the value of f evaluated at the fixed point, x_p, is equal to x_p. For example, $x = 0.739085133$ is a fixed point for the function $\mathrm{Cos}(x)$ as $\mathrm{Cos}(0.739085133) = 0.739085133$. Mathematically, a fixed point of the function $f(x)$ is a solution, x^*, if it exists, to the following equation [195]

$$x = f(x). \tag{10.17}$$

If a fixed point is *aymptotically stable* (explained in section 10.3.3), then starting from a point, say x_0, close enough to the solution, x^*, the iterative sequence

$$x_0 \, , \; f(x_0) \, , \; f(f(x_0)) \, , \; f(f(f(x_0))) \, , \; \ldots \tag{10.18}$$

converges to x^*. This iterative sequence is called the *fixed point iterative* sequence, also known as nonlinear Richardson iteration, Picard iteration, or the method of successive substitution [195], and can be written as follows

$$x_{\kappa+1} = f(x_\kappa) \quad ; \quad \text{starting from a guess point } x_0. \tag{10.19}$$

In case of a solution of n non-linear equations in n unknowns, $\mathbf{x} = [x_1, x_2, x_3, \ldots, x_n]^T$ as represented by the following system

$$x_1 = f_1(\mathbf{x}) \tag{10.20}$$
$$x_2 = f_2(\mathbf{x}) \tag{10.21}$$
$$\vdots \tag{10.22}$$
$$x_n = f_n(\mathbf{x}), \tag{10.23}$$

the iterative sequence can be written as

$$x_1^{\kappa+1} = f_1(\mathbf{x}^\kappa) \tag{10.24}$$
$$x_2^{\kappa+1} = f_2(\mathbf{x}^\kappa) \tag{10.25}$$
$$\vdots \tag{10.26}$$
$$x_n^{\kappa+1} = f_n(\mathbf{x}^\kappa), \tag{10.27}$$

starting from an initial guess $\mathbf{x}^0 = [x_1^0, x_2^0, x_3^0, \ldots, x_n^0]^T$.

10.3.2 Fixed point iterative form of the Krig derivative

In order to solve equation 10.16, the matrix $\mathbf{D_i(x)}$, defined in equation 10.13, is written as

$$\mathbf{D_i(x)} = \begin{bmatrix} x_i & x_i & \cdots & x_i \end{bmatrix} - \begin{bmatrix} x_i^{(1)} & x_i^{(2)} & \cdots & x_i^{(p)} \end{bmatrix} \in \mathbb{R}^{1\times p}, \tag{10.28}$$

$$\mathbf{D_i(x)} = x_i \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} - \begin{bmatrix} x_i^{(1)} & x_i^{(2)} & \cdots & x_i^{(p)} \end{bmatrix} \in \mathbb{R}^{1\times p}, \tag{10.29}$$

$$\mathbf{D_i(x)} = x_i \mathbf{1_p} - \mathbf{X}_i, \tag{10.30}$$

where

$$\mathbf{1_p} = [1, 1, \ldots, 1] \in \mathbb{R}^{1\times p}, \tag{10.31}$$

and

$$\mathbf{X}_i = \begin{bmatrix} x_i^{(1)} & x_i^{(2)} & \cdots & x_i^{(p)} \end{bmatrix}. \tag{10.32}$$

The components of $\mathbf{X}_i$ represent the i^{th} dimensions of each of the p points used to construct the Kriging model. Substituting the above equation for $\mathbf{D_i(x)}$ in equation 10.33 yields

$$-2\theta_i \left[x_i \mathbf{1_p} - \mathbf{X}_i \right] \mathbf{Z(x)\, w} = 0 \quad ; \quad i = 1 \ldots n. \tag{10.33}$$

When $\theta_i \neq 0$, this yields

$$\left[x_i \mathbf{1_p} - \mathbf{X}_i \right] \mathbf{Z(x)\, w} = 0 \quad ; \quad i = 1 \ldots n \tag{10.34}$$

$$x_i \left[\mathbf{1_p\, Z(x)\, w} \right] - \left[\mathbf{X}_i\, \mathbf{Z(x)\, w} \right] = 0 \quad ; \quad i = 1 \ldots n \tag{10.35}$$

$$x_i = \frac{\mathbf{X}_i\, \mathbf{Z(x)\, w}}{\mathbf{1_p\, Z(x)\, w}} = \mathbf{G_i(x)} \quad ; \quad i = 1 \ldots n \tag{10.36}$$

or

$$x_i = \mathbf{G_i(x)} \quad ; \quad i = 1 \ldots n, \tag{10.37}$$

where

$$\mathbf{G_i(x)} = \frac{\mathbf{X}_i\, \mathbf{Z(x)\, w}}{\mathbf{1_p\, Z(x)\, w}}. \tag{10.38}$$

The above equation shows that the stationary points of the Kriging predictor function are the fixed points of the set of functions represented by $\mathbf{G_i(x)}$, $i = 1 \ldots n$. Consequently, starting from an initial guess, $\mathbf{x}^0 = [x_1^0, x_2^0, x_3^0, \ldots, x_n^0]^T$, the following sequence can be used to reach a solution

$$x_i^{\kappa+1} = \frac{\mathbf{X}_i\, \mathbf{Z(x^\kappa)\, w}}{\mathbf{1_p\, Z(x^\kappa)\, w}}. \tag{10.39}$$

For computational efficiency, the entire $\mathbf{x}^\kappa \in \mathbb{R}^{\kappa\times 1}$ vector can be calculated using the following iterative equation

$$\mathbf{x}^{\kappa+1} = \mathbf{G(x}^\kappa\mathbf{)} \tag{10.40}$$

and

$$\mathbf{G(x)} = \frac{\mathbf{X\, Z(x)\, w}}{\mathbf{1_p\, Z(x)\, w}} \tag{10.41}$$

where $\mathbf{X} \in \mathbb{R}^{n \times p}$ is given by

$$\mathbf{X} = \begin{bmatrix} x_1^{(1)} & x_1^{(2)} & x_1^{(3)} & \cdots & x_1^{(p)} \\ x_2^{(1)} & x_2^{(2)} & x_2^{(3)} & \cdots & x_2^{(p)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n^{(1)} & x_n^{(2)} & x_2^{(3)} & \cdots & x_n^{(p)} \end{bmatrix}. \tag{10.42}$$

10.3.3 Convergence conditions

In this section the conditions under which a fixed point iterative sequence converges to a fixed point are outlined. Fixed points can be classified into three categories [196]

- *Asymptotically stable points*: These are fixed points with a property that all nearby points converge to the fixed point.
- *Stable points*: These are fixed points with a property that all nearby points stay nearby.
- *Unstable points*: These are fixed points with a property that almost all nearby points diverge away from the fixed point.

For a scalar function, $g(u), g : \mathbb{R} \to \mathbb{R}$, the following theorem for convergence of the fixed point iterative scheme holds

Theorem 10.3.1. *"Let g(u) be a continuously differentiable scalar function. Suppose* $u^* = g(u^*)$ *is a fixed point. If* $|g'(u^*)| < 1$*, then* u^* *is an asymptotically stable fixed point, and hence any sequence of iterates* $u^{(\kappa)}$ *which starts out sufficiently close to* u^* *will converge to* u^**. On the other hand, if* $|g'(u^*)| > 1$*, then* u^* *is an unstable fixed point, and the only iterates which converge to it are those that land exactly on it, i.e.,* $u^{(\kappa)} = u^*$ *for some* $\kappa \geq 0$*."* [196]

Similarly, for a set of functions , $\mathbf{g(u)}, \mathbf{g} : \mathbb{R}^n \to \mathbb{R}^n$, the following theorem for convergence of the fixed point iterative scheme holds

Theorem 10.3.2. *"Let* $\mathbf{u}^*$ *be a fixed point for the discrete dynamical system* $\mathbf{u}^{(\kappa+1)} = \mathbf{g}(\mathbf{u}^{(\kappa)})$*. If the Jacobian matrix norm* $\|\mathbf{g}'(\mathbf{u}^*)\| < 1$*, then* $\mathbf{g}$ *is a contraction at* $\mathbf{u}^*$*, and hence the fixed point* $\mathbf{u}^*$ *is asymptotically stable."* [196]

In other words, the above theorem implies that if the *spectral radius* (supremum of the set containing absolute values of the eigen values) of $\mathbf{g}'(\mathbf{u}^*)$ is less than one, then $\mathbf{u}^*$ is an asymptotically stable point [196], i.e.

$$\rho(\mathbf{g}'(\mathbf{u}^*)) < 1, \tag{10.43}$$

where $\rho(\cdot)$ represents the spectral radius. Furthermore, the convergence rate of a fixed point iteration is directly related to the spectral radius. A smaller spectral radius of the Jacobian matrix at a stationary point implies faster convergence of nearby iterates to the stationary point [196].

Having laid out the convergence conditions, the term *basin of attraction* can be defined as

Let $\mathbf{u}^*$ *be a fixed point for the discrete dynamical system* $\mathbf{u}^{(\kappa+1)} = \mathbf{g}(\mathbf{u}^{(\kappa)})$*. Then the set of all points,* $\mathbf{u}^\gamma$*, for which the iterative sequence* $\mathbf{u}^{(\kappa+1)} = \mathbf{g}(\mathbf{u}^{(\kappa)})$*, starting from* $\mathbf{u}^\gamma$*, converges to the*

solution $\mathbf{u}^*$ *is called the basin of attraction of* $\mathbf{u}^*$.

Hence, from the view of finding the stationary points through a fixed point iterative sequence, it is desirable to have a large basin of attraction for all stationary points.

Applying theorem 10.3.2 to the iterative sequence represented by equation 10.40, the iterative sequence converges to a stationary point, $\mathbf{x}^*$ of the Kriging predictor if

$$\|\mathbf{G}'(\mathbf{x}^*)\| < 1, \tag{10.44}$$

where $\|\cdot\|$ denotes a norm and $\mathbf{G}'(\mathbf{x}^*)$ is the Jacobian matrix of $\mathbf{G(x)}$ evaluated at $\mathbf{x}^*$, and is given by

$$\mathbf{G}'(\mathbf{x}^*) = \begin{bmatrix} \frac{\partial \mathbf{G_1(x)}}{\partial x_1} & \frac{\partial \mathbf{G_1(x)}}{\partial x_2} & \cdots & \frac{\partial \mathbf{G_1(x)}}{\partial x_n} \\ \frac{\partial \mathbf{G_2(x)}}{\partial x_1} & \frac{\partial \mathbf{G_2(x)}}{\partial x_2} & \cdots & \frac{\partial \mathbf{G_2(x)}}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \mathbf{G_n(x)}}{\partial x_1} & \frac{\partial \mathbf{G_n(x)}}{\partial x_2} & \cdots & \frac{\partial \mathbf{G_n(x)}}{\partial x_n} \end{bmatrix}_{\text{at } \mathbf{x}^*} \in \mathbb{R}^{n \times n}. \tag{10.45}$$

The ij^{th} term of the above Jacobian matrix is evaluated next

$$\frac{\partial \mathbf{G_i(x)}}{\partial x_j} = \frac{\left(\mathbf{1_p}\ \mathbf{Z(x)}\ \mathbf{w}\right)\ \frac{\partial}{\partial x_j}\left(\mathbf{X_i}\ \mathbf{Z(x)}\ \mathbf{w}\right) - \left(\mathbf{X_i}\ \mathbf{Z(x)}\ \mathbf{w}\right)\frac{\partial}{\partial x_j}\left(\mathbf{1_p}\ \mathbf{Z(x)}\ \mathbf{w}\right)}{\left(\mathbf{1_p}\ \mathbf{Z(x)}\ \mathbf{w}\right)^2} \tag{10.46}$$

$$\frac{\partial \mathbf{G_i(x)}}{\partial x_j} = \frac{\left(\mathbf{1_p}\ \mathbf{Z(x)}\ \mathbf{w}\right)\left(-2\theta_j\right)\left(\mathbf{X_i}\ \mathbf{Z(x)}\ \mathbf{K_j}\ \mathbf{w}\right) - \left(\mathbf{X_i}\ \mathbf{Z(x)}\ \mathbf{w}\right)\left(-2\theta_j\right)\left(\mathbf{1_p}\ \mathbf{Z(x)}\ \mathbf{K_j}\ \mathbf{w}\right)}{\left(\mathbf{1_p}\ \mathbf{Z(x)}\ \mathbf{w}\right)^2}, \tag{10.47}$$

where

$$\mathbf{K_j} = \begin{bmatrix} \left(x_j - x_j^{(1)}\right) & 0 & \cdots & 0 \\ 0 & \left(x_j - x_j^{(2)}\right) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \left(x_j - x_j^{(p)}\right) \end{bmatrix} \in \mathbb{R}^{p \times p}. \tag{10.48}$$

Equation 10.47 represents the general partial derivative at any point $\mathbf{x}$. At a stationary point, $\mathbf{x}^*$, we have an additional condition represented by equation 10.36, i.e.

$$x_i^*\left(\mathbf{1_p}\ \mathbf{Z(x^*)}\ \mathbf{w}\right) = \mathbf{X}_i\ \mathbf{Z(x^*)}\ \mathbf{w}; \tag{10.49}$$

substituting this in equation 10.47, yields

$$\frac{\partial \mathbf{G_i(x^*)}}{\partial x_j} = \frac{\left(\mathbf{1_p}\ \mathbf{Z(x^*)}\ \mathbf{w}\right)\left(-2\theta_j\right)\left(\mathbf{X_i}\ \mathbf{Z(x^*)}\ \mathbf{K_j}\ \mathbf{w}\right) - \left(x_i^*\mathbf{1_p}\ \mathbf{Z(x^*)}\ \mathbf{w}\right)\left(-2\theta_j\right)\left(\mathbf{1_p}\ \mathbf{Z(x^*)}\ \mathbf{K_j}\ \mathbf{w}\right)}{\left(\mathbf{1_p}\ \mathbf{Z(x^*)}\ \mathbf{w}\right)^2}. \tag{10.50}$$

Now, when $\left(\mathbf{1_p}\ \mathbf{Z(x^*)}\ \mathbf{w}\right) \neq 0$, the above equation becomes

$$\frac{\partial \mathbf{G_i}(\mathbf{x}^*)}{\partial x_j} = \frac{\left(-2\theta_j\right)\left(\mathbf{X_i}\ \mathbf{Z}(\mathbf{x}^*)\ \mathbf{K_j}\ \mathbf{w}\right) - \left(x_i^*\right)\left(-2\theta_j\right)\left(\mathbf{1_p}\ \mathbf{Z}(\mathbf{x}^*)\ \mathbf{K_j}\ \mathbf{w}\right)}{\left(\mathbf{1_p}\ \mathbf{Z}(\mathbf{x}^*)\ \mathbf{w}\right)} \tag{10.51}$$

$$\frac{\partial \mathbf{G_i}(\mathbf{x}^*)}{\partial x_j} = \left(-2\theta_j\right)\frac{\left(\mathbf{X_i}\ \mathbf{Z}(\mathbf{x}^*)\ \mathbf{K_j}\ \mathbf{w}\right) - \left(x_i^*\ \mathbf{1_p}\ \mathbf{Z}(\mathbf{x}^*)\ \mathbf{K_j}\ \mathbf{w}\right)}{\left(\mathbf{1_p}\ \mathbf{Z}(\mathbf{x}^*)\ \mathbf{w}\right)} \tag{10.52}$$

$$\frac{\partial \mathbf{G_i}(\mathbf{x}^*)}{\partial x_j} = \left(2\theta_j\right)\frac{\left(x_i^*\mathbf{1_p} - \mathbf{X_i}\right)\left(\mathbf{Z}(\mathbf{x}^*)\ \mathbf{K_j}\ \mathbf{w}\right)}{\left(\mathbf{1_p}\ \mathbf{Z}(\mathbf{x}^*)\ \mathbf{w}\right)}. \tag{10.53}$$

Using this equation, the Jacobian matrix $\mathbf{G'(x^*)} \in \mathbb{R}^{n\times n}$ (from equation 10.45) can be calculated as follows

$$\mathbf{G'}(\mathbf{x}^*) = \frac{2\ \mathbf{D}(\mathbf{x}^*)\ \mathbf{Z}(\mathbf{x}^*)\ \mathbf{W}\ \mathbf{D^T}(\mathbf{x}^*)\ \mathbf{\Theta}}{\mathbf{1_p}\ \mathbf{Z}(\mathbf{x}^*)\ \mathbf{w}} \quad \in \mathbb{R}^{n\times n}, \tag{10.54}$$

where

$$\mathbf{D}(\mathbf{x}^*) = \begin{bmatrix} \left(x_1^* - x_1^{(1)}\right) & \left(x_1^* - x_1^{(2)}\right) & \cdots & \left(x_1^* - x_1^{(p)}\right) \\ \left(x_2^* - x_2^{(1)}\right) & \left(x_2^* - x_2^{(2)}\right) & \cdots & \left(x_2^* - x_2^{(p)}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \left(x_n^* - x_n^{(1)}\right) & \left(x_n^* - x_n^{(2)}\right) & \cdots & \left(x_n^* - x_n^{(p)}\right) \end{bmatrix} \in \mathbb{R}^{n\times p}, \tag{10.55}$$

$$\mathbf{W} = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_p \end{bmatrix} \in \mathbb{R}^{p\times p} \tag{10.56}$$

or

$$\mathbf{w} = \mathbf{W}\ \mathbf{1_p}, \tag{10.57}$$

,

$$\mathbf{\Theta} = \begin{bmatrix} \theta_1 & 0 & \cdots & 0 \\ 0 & \theta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \theta_n \end{bmatrix} \in \mathbb{R}^{n\times n}, \tag{10.58}$$

and $\mathbf{D^T}$ represents the transpose of $\mathbf{D}$. Hence, if the norm of the Jacobian matrix at a stationary point, $\|\mathbf{G'(x^*)}\|$, as calculated by equation 10.54 is less than one, then the stationary point is asymptotically stable, and an iterative sequence starting from a point near the stationary point converges to the stationary point.

10.3.4 Application to Branin function

The Branin function [197], as shown in Figure 10.1, is a widely used two variable test function for global optimisation. It is defined as follows

$$f(x_1, x_2) = a(x_2 - bx_1^2 + cx_1 - d)^2 + e(1 - f)\cos(x_1) + e, \tag{10.59}$$

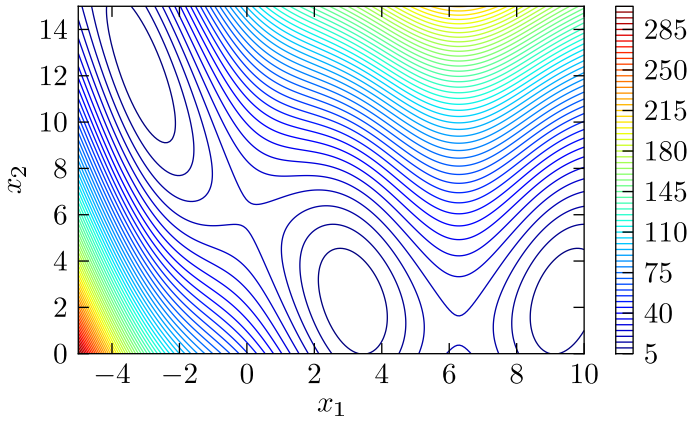


Figure 10.1: Contours of the original Branin Function

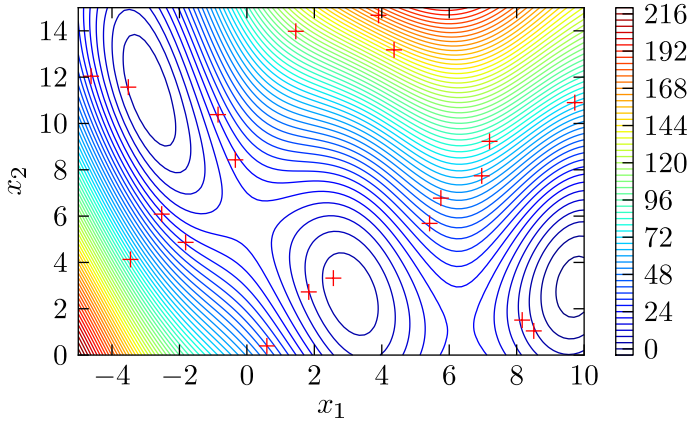


Figure 10.2: Contours of the Kriging predictor (generated from sampling 20 points shown in as red '+') for the Branin Function

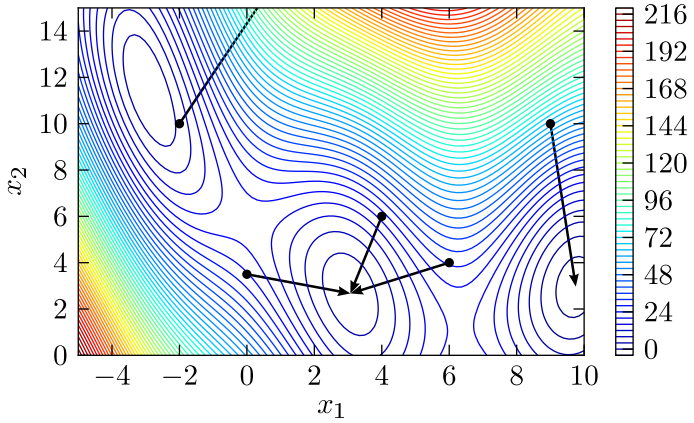


Figure 10.3: Equation 10.40 applied to the Kriging predictor of Branin Function starting from five different points

where $-5 \le x_1 \le 10$, $0 \le x_2 \le 15$, $a = 1$, $b = \frac{5}{4\pi^2}$, $c = \frac{5}{\pi}$, $d = 6$, $e = 10$, and $f = \frac{1}{8\pi}$. The Branin function has three global minima, $f(x_1, x_2) = 0.397887$, at $(x_1, x_2) = (-\pi, 12.275)$, $(\pi, 2.275)$, and $(9.42478, 2.475)$. A Krig of this function is created using 20 points generated using a semi-optimised Latin hypercube [18, 198] sampling plan (100 Latin hypercube samples are generated and the one with lowest metric proposed by Audze and Eglais [199] is chosen). The function as predicted by the Kriging model is shown in Figure 10.2. A fixed point iterative scheme represented by equation 10.40 is applied to this Krig starting from five points, viz. $[(4, 6), (9, 10), (-2, 10), (0, 3.5), (6, 4)]$. The five points are chosen such that each point is closest to one of the stationary points of the Branin function Krig. The results of this are shown in Figure 10.3; the black dots represent the starting points of the fixed point iteration and the arrows show the points to which the iteration converge. It is observed that all iterations converge to one of the stationary points of the Krig function. However, the locations to which these iterations converge are far from ideal. The Branin function has five stationary points, three minima and two saddle points. The five chosen starting points lie close to one of these stationary points and hence, in an ideal case, iterations starting from each of these points should converge to the respective stationary points (see Figure 10.5 for such an ideal distribution). The observed behaviour of the points, especially point $(-2, 10)$, can be understood by looking at the spectral radius of the minimum closest to it, i.e. $(-2.933, 11.456)$. Note that this is slightly different from $(-\pi, 12.275)$, the minimum of the real Branin function, as we are working with the current Kriging predictor, the accuracy of which can be improved, if desired, by adding more points at suitable locations to construct the Krig. The spectral radius at this minimum is 1.519, see Table 10.1 (normal case), while the spectral radius at the other two minima are 0.638 and 0.740. This explains why a fixed point iteration starting from any point, except that which starts from the minimum itself, does not converge to this minimum as the spectral radius at this point is greater than one. Table 10.1 also explains why the two saddle points are missed as the spectral radii there are 1.763 and 1.984, both greater than one.

The above discussion leads to the question as to whether equation 10.40 can be posed in another way such that the spectral radius at all/more stationary points could be reduced to a quantity less than one in order to ensure that all fixed point iterative sequences starting from points closer to the stationary points converge to the respective stationary points. Such transformations form the discussion of the next sub-section.

10.3.5 Modifying the form of the fixed point Krig equation

Since the spectral radius of the Jacobian matrix, equation 10.45 is directly related to the magnitude of entries in the matrix, any transformation sought should be based on the underlying concept to decrease the magnitude of these entries. Example transformations for a general equation, $x = f(x)$, are:

$$x = \sqrt{(xf(x))} \tag{10.60}$$

$$x = \frac{x}{f(x)} + x - 1 \tag{10.61}$$

$$x = \frac{f(x)}{x} + x - 1 \tag{10.62}$$

$$x = \frac{1}{2}\left(x + f(x)\right). \tag{10.63}$$

Table 10.1: Spectral radii for the stationary points of the Branin function Kriging Predictor

Stationary Point	Spectral radii				
	normal	$n_d = 2$	$n_d = 10$	$n_d = 50$	$n_d = 1000$
$(-2.933, 11.456)$	1.519	0.706	0.941	0.988	0.999
$(3.070, 2.672)$	0.638	0.665	0.933	0.986	0.999
$(9.756, 2.963)$	0.740	0.467	0.893	0.978	0.998
$(-0.008, 5.008)$	1.763	1.381	1.076	1.015	1.0007
$(6.342, 1.121)$	1.984	1.492	1.098	1.019	1.0009

Clearly, one can write the same equation in many other forms, each still representing a fixed point iterative scheme. However, following the underlying idea of keeping the entries of the Jacobian matrix of the RHS of the above equations low, a general form for the final equation (equation 10.63) shown above is chosen:

$$x = \frac{1}{n_d}\Big((n_d - 1)x + f(x)\Big) \quad ; \quad n_d \geq 1 \text{ and } n_d \in \mathbb{Z}, \tag{10.64}$$

where n_d is an integer greater than or equal to one. The derivative of the RHS in such a form, $x = g(x)$ is

$$g'(x) = \frac{n_d - 1}{n_d} + \frac{f'(x)}{n_d}, \tag{10.65}$$

where $f'(x)$ would be the derivative of the RHS of the original form, $x = f(x)$. From the above equation one can observe that the original derivative term is now divided by the factor n_d, and a term of $(n_d - 1)/n_d$ has been added. It can be concluded that if $f'(x^*)$ is finite, then

$$\lim_{n\to\infty} g'(x^*) = 1. \tag{10.66}$$

The above method works well in cases where the original derivative, $f'(x^*)$, is negative, in which case the above limit tends to one from the left. For example, if $f'(x^*) = -1.5$, then when $n_d = 2$ from the equation $g'(x^*) = -0.25$, with a magnitude less than one, and hence x^* becomes asymptotically stable, i.e. fixed point iterative sequences starting from points close to it converge to it. The form of equation 10.64 when applied to equation 10.40 for the Kriging multidimensional case becomes

$$\mathbf{x}^{\kappa+1} = \frac{1}{n_d}\left((n_d - 1)\,\mathbf{x}^{\kappa} + \mathbf{G}(\mathbf{x}^{\kappa})\right) \tag{10.67}$$

with the Jacobian matrix, $\mathbf{J}$ for such a fixed point iteration being

$$\mathbf{J} = \left(\frac{n_d - 1}{n_d}\right)\mathbf{I} + \frac{1}{n_d}\mathbf{G}'(\mathbf{x}), \tag{10.68}$$

where $\mathbf{I}$ is an identity matrix, $\mathbf{I} \in \mathbb{R}^{n\times n}$.

Figure 10.4 shows the results of the five test points when the iteration represented by equation 10.67, with $n_d = 2$, is applied to the Branin function Krig (c.f. figure 10.2). Comparing this with Figure 10.3, it is observed that now the point $(-2, 10)$ converges to the closest (desired) minimum. This is also apparent by calculating the spectral radius of the right hand side of the fixed point equation. As shown in Table 10.1 the spectral radius for this minima is 0.706 (less than one) as opposed to the original case where the spectral radius was 1.519 (greater than one), and hence the minima has now become asymptotically stable. However, as can be seen from Figure 10.4, the saddle points of the Branin function are still not found.

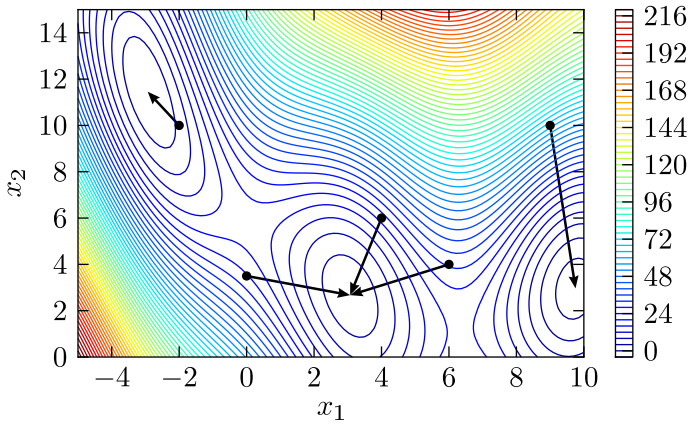


Figure 10.4: Equation 10.67 applied to the Kriging predictor of Branin Function starting from five different points

This is also revealed by the spectral radii in Table 10.1 where it is found that the spectral radii for the saddle points approach one as n_d increases from the right. Consequently, no finite value of n_d, no matter how large, makes the spectral radius of the Jacobian matrix less than one. This is primarily because the largest entries of the Jacobian matrix for these saddle points are positive in which case it is inevitable that the spectral radius approaches the limit of 1 from the right, as n_d increases. Lastly, it was mentioned in section 10.3.3 that the convergence of a fixed point iteration is related to the spectral radius of the Jacobian matrix; a smaller spectral radius implies faster convergence. With an increase in n_d, the spectral radii, as shown above, tends to one either from the left or the right. For the latter case, there is no convergence to the stationary point, and for the former case, the increase in spectral radius, as n_d increases leads to slower convergence.

In order to resolve the issues of spectral radii being less than one, and identifying a universal form for the fixed point iteration which ensures convergence, the Newton-Raphson (NR) scheme for the Krig equations is explored in the following section.

10.4 Newton-Raphson (NR) scheme to find stationary points of a Krig

The Newton-Raphson (NR) method is a fixed point iterative method to find the roots of an equation, $f(x) = 0$. The iterative scheme is defined as follows

$$x^{\kappa+1} = x^{\kappa} - \frac{f(x^{\kappa})}{f'(x^{\kappa})}. \tag{10.69}$$

It can readily be observed that the derivative of the RHS of the fixed point equation, i.e. $g(x) = x - f(x)/f'(x)$, is

$$g'(x) = \frac{f(x)\ f''(x)}{f'(x)^2}. \tag{10.70}$$

At the root, i.e. when $f(x^*) = 0$, $g'(x^*) = 0$, implying that every stationary point is asymptotically stable. When applied to finding a solution to equation 10.37, this procedure ensures that every stationary point of the Krig is asymptotically stable. However, this comes at a cost of calculating the derivative (i.e. the Jacobian matrix) in each iteration. The set of equations

represented by Equation 10.40 can be writtn as

$$\mathbf{F(x)} \;=\; \mathbf{x} - \mathbf{G(x)} \;=\; 0. \tag{10.71}$$

Hence, the Newton Raphson iteration can be written as

$$\mathbf{x}^{\kappa+1} \;=\; \mathbf{x}^{\kappa} - \frac{\mathbf{F}(\mathbf{x}^{\kappa})}{\mathbf{F}'(\mathbf{x}^{\kappa})} \tag{10.72}$$

$$\mathbf{x}^{\kappa+1} \;=\; \mathbf{x}^{\kappa} - \mathbf{F}'(\mathbf{x}^{\kappa})^{-1}\mathbf{F}(\mathbf{x}^{\kappa}), \tag{10.73}$$

where $\mathbf{F}'(\mathbf{x})^{-1}$ is the inverse of the Jacobian matrix of $\mathbf{F(x)}$, obtainable from

$$\mathbf{F}'(\mathbf{x}) \;=\; \mathbf{I} - \mathbf{G}'(\mathbf{x}), \tag{10.74}$$

where $\mathbf{I}$ is an identity matrix, $\mathbf{I} \in \mathbb{R}^{n\times n}$, and $\mathbf{G}'(\mathbf{x})$ is given by equation 10.45. Using equation 10.47, $\mathbf{F}'(\mathbf{x})$ can be written as

$$\mathbf{F}'(\mathbf{x}) \;=\; \mathbf{I} \;-\; \frac{2\,\left(\mathbf{X}\;\mathbf{Z(x)}\;\mathbf{w}\;\mathbf{1_p}\right)\;\left(\mathbf{Z(x)}\;\mathbf{W}\;\mathbf{D^T(x)}\;\mathbf{\Theta}\right)}{\left(\mathbf{1_p}\;\mathbf{Z(x)}\;\mathbf{w}\right)^2} \;+\; \frac{2\;(\mathbf{X})\;\left(\mathbf{Z(x)}\;\mathbf{W}\;\mathbf{D^T(x)}\;\mathbf{\Theta}\right)}{\left(\mathbf{1_p}\;\mathbf{Z(x)}\;\mathbf{w}\right)}, \tag{10.75}$$

where $\mathbf{X}$, $\mathbf{Z(x)}$, $\mathbf{w}$, $\mathbf{W}$, $\mathbf{D(x)}$, and $\mathbf{\Theta}$ are defined by equations 10.42, 10.14, 10.15, 10.56, 10.55, and 10.58, respectively.

Figure 10.5 shows the results of the five test points when the iteration represented by equation 10.73 is applied to the Branin function Krig. Comparing this with Figures 10.3 and 10.4, it is readily observed that, in this scheme of fixed point iterations, each stationary point, i.e. the three minima and the two saddle points, is asymptotically stable and, hence, iterative sequences starting from points in their neighbourhood converge to the stationary points. This, however, comes at a cost of computing the Jacobian matrix and its inverse at each iteration.

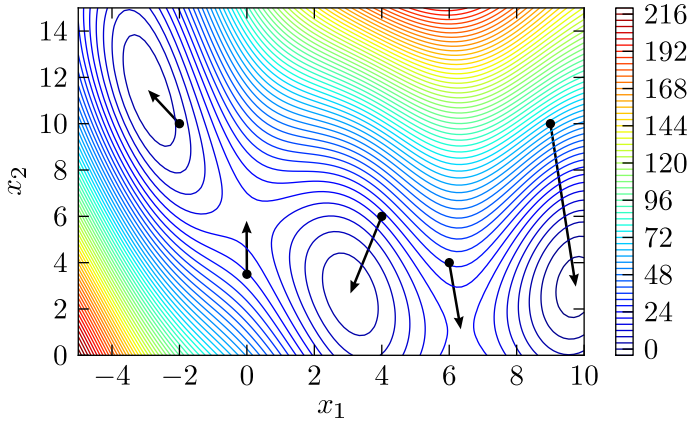


Figure 10.5: Newton-Raphson equation 10.73 applied to the Kriging predictor of the Branin Function starting from five different points

Henceforth, throughout this chapter, the three fixed point iterative sequences are referred as

1. *Standard fixed point* iteration (SFP): The standard fixed point iteration derived from the Kriging predictor (equation 10.40).

Table 10.2: Sample size to construct Krigs and grid-size to evaluate the MFP and NR iterative sequences

Function	Sample size	grid size	n_d (MFP)
Branin function	20	21×21	2
De Jong's function	20	10×10	2
Rosenbrock's function	50	9×9	2
Rastrigin's function	150	21×21	10
Bump function	200	21×21	6
Ackley's function	150	21×21	2
Six-hump camel back function	100	21×21	1
Dropwave function	150	21×21	2
Easom's function	200	21×21	2
Goldstein-Price's function	150	21×21	2

2. *Modified fixed point* iteration (MFP): The modified form of the standard fixed point iteration (equation 10.67).

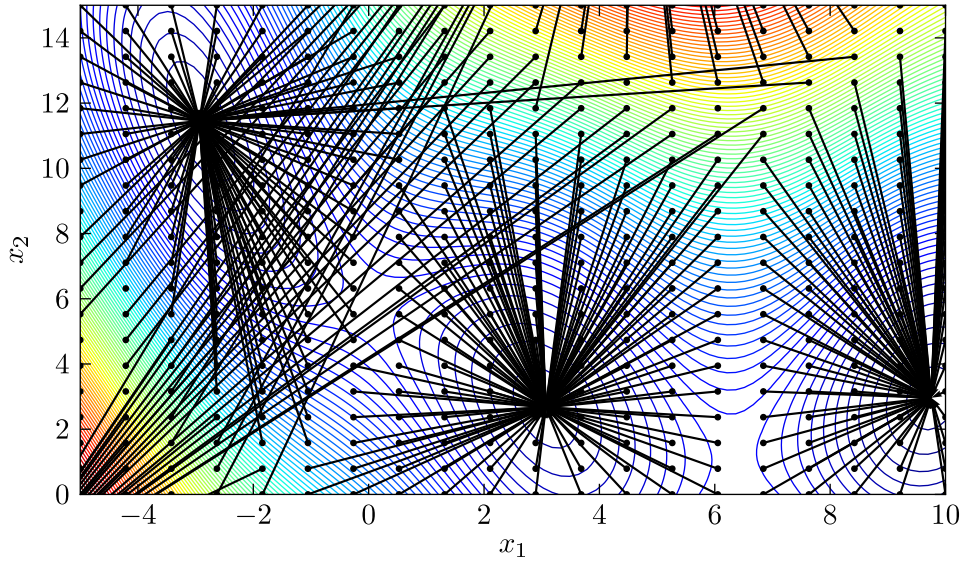
3. *Newton-Raphson* iteration (NR): The Newton-Raphson form of the fixed point iteration (equation 10.73).

10.5 Example Applications

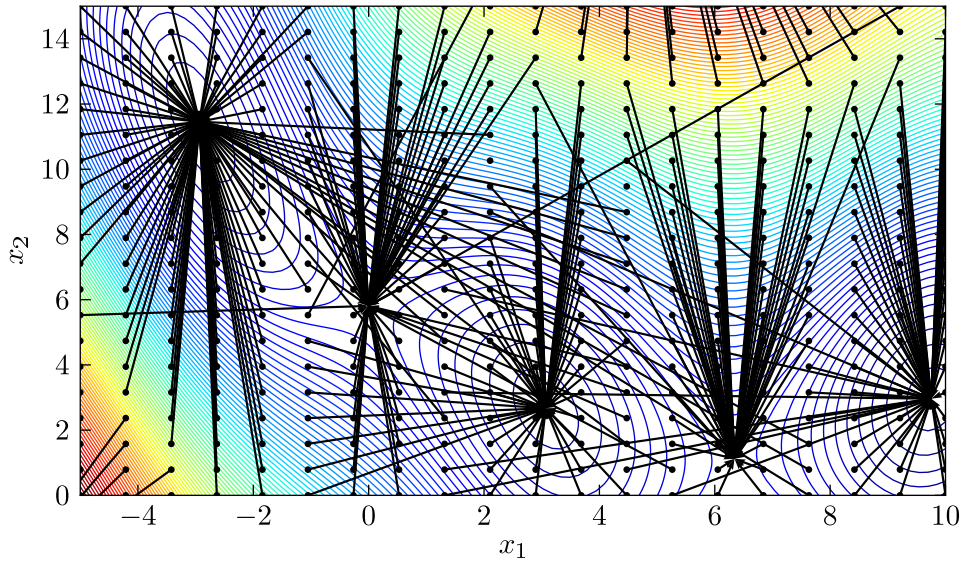
In this section the above formulated *fixed point iterative sequences* are applied to find the stationary points of a Krig of several test functions. These functions are used widely for testing optimisation algorithms and, unless otherwise stated, have been taken from the list provided by Molga et. al. [200] and Yang [201]. For each of the functions, a suitable sample size, that which ensures a low error (posterior variance) as predicted by equation 3.37, is used to construct the Krigs. Thereafter, a suitable grid size, i.e. a grid size that sufficiently depicts the behaviour of the two iterative sequences, is used to generate starting points for the MFP and NR iterative sequences. The results of these are presented by arrow plots, where each arrow starts from the starting point of the sequences and points to the point where that sequence converged. For each of the following functions, the sample size, p, used to construct the Krigs, the grid size to evaluate the iterative sequences, and the value of n_d used in equation 10.67 for the MFP sequence, are shown in Table 10.2.

10.5.1 Branin function

The Branin function has already been the exemplary function so far. Figures 10.6a and 10.6b show the behaviour of the MFP and NR sequences for a 21×21 grid of starting points. The MFP sequence, as can be seen from Figure 10.6a does not converge to the two saddle points. The basins of attraction for the three minima are relatively large when compared to the NR sequence. However, the NR sequence does find all the stationary points of the function, with a considerable sized basin of attraction for each of the stationary points.



(a) MFP iterations with $n_d = 2$



(b) NR fixed point iterations

Figure 10.6: Iterative schemes for finding stationary points of Branin function Krig; grid size $= 21 \times 21$

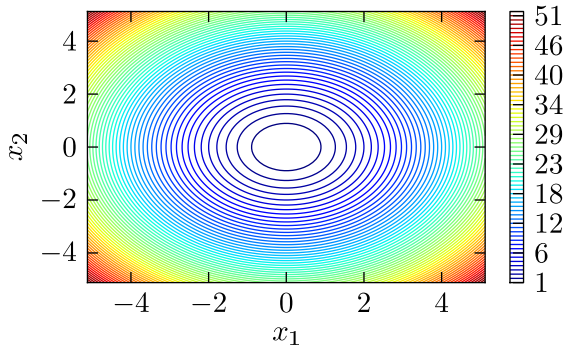
10.5.2 De Jong's function

The simple De Jong's sphere function is defined as

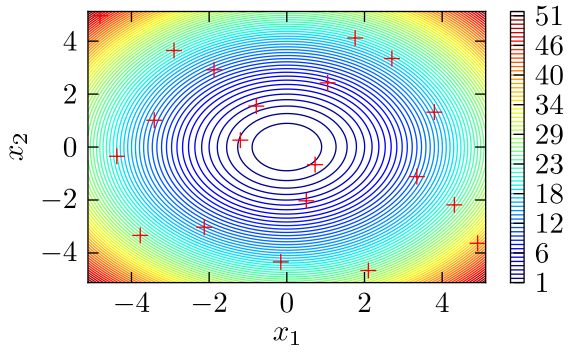
$$f(\mathbf{x}) = \sum_{i=1}^{n} x_i^2 \quad ; \quad -5.12 \leq x_i \leq 5.12 \tag{10.76}$$

This unimodal and convex function has its global minimum, $f(\mathbf{x}^*) = 0$, at $x_i^* = 0,\ i = 1 \ldots n$.

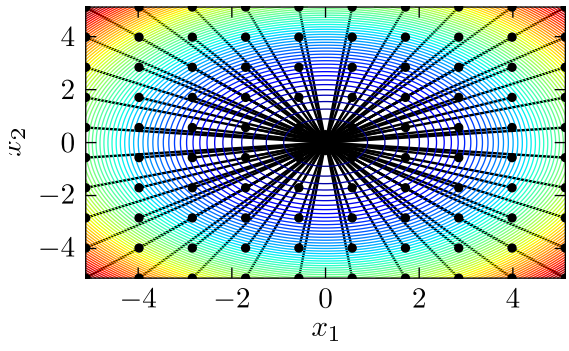
Figures 10.7a and 10.7b show the original De Jong's function and its Krig created by sampling 20 points, respectively. Figures 10.7c and 10.7d show the behaviour of the MFP and NR sequences for a 10×10 grid of starting points. Being a unimodal convex quadratic function, the observed behaviour of the sequences, i.e. all points converge to the global minimum, is not surprising. Such behaviour, for this simple function, is expected out of every gradient based optimiser.



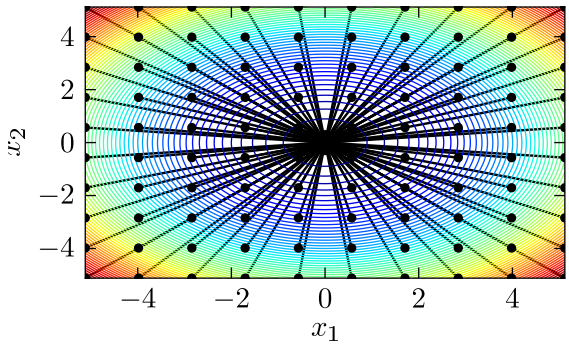
(a) Original De Jong's function



(b) Krig of De Jong's function (20 sample points shown in red 'plus')



(c) MFP iterations with $n_d = 2$



(d) NR fixed point iterations

Figure 10.7: Iterative schemes for finding stationary points of De Jong's function Krig; grid size $= 10 \times 10$

10.5.3 Rosenbrock's function

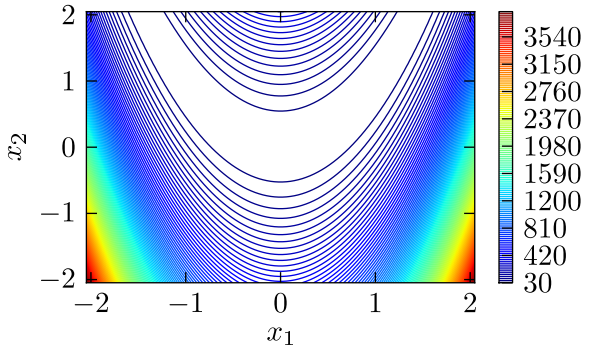
The Rosenbrock's function, also known as the *Banana function*, is defined as

$$f(\mathbf{x}) = \sum_{i=1}^{n-1} \left[(x_i - 1)^2 + 100(x_{i+1} - x_i^2)^2 \right] \quad ; \quad -2.048 \leq x_i \leq 2.048. \tag{10.77}$$

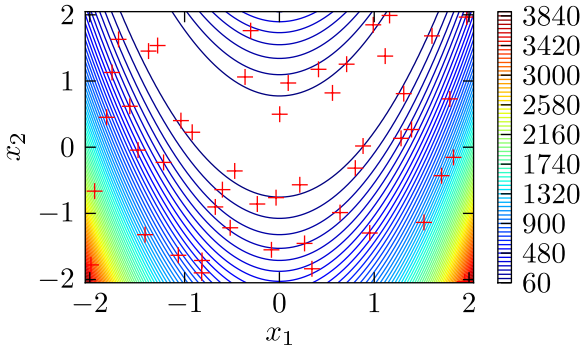
The global optimum of the above function lies in a narrow valley of parabolic shape. Most algorithms easily locate the valley, however the challenge to any algorithm lies in converging

to the global optimum in the valley. The global minimum, $f(\mathbf{x}^*) = 0$, for the Rosenbrock's function is located at $x_i^* = 1,\ i = 1 \ldots n$.

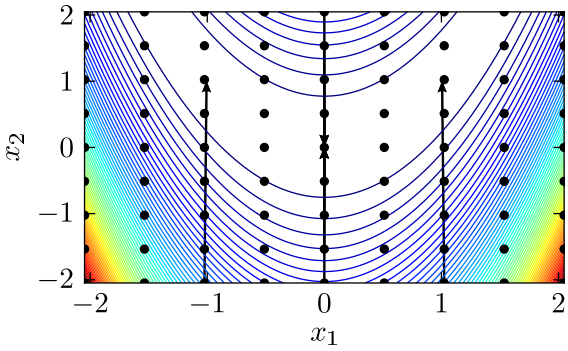
Figures 10.8a and 10.8b show the original Rosenbrock's function and its Krig created by sampling 50 points, respectively. Figures 10.8c and 10.8d show the behaviour of the MFP and NR sequences for a 9×9 grid of starting points. This choice of grid is intentional, as for the MFP sequence it is found that only those sequences that start from within an extremely narrow window around the x coordinate of the minimum converge to the desired minimum; a sequence starting from every other point diverges. This does not seem to be the case with the NR sequence, for which the basins of attraction for each stationary point are relative large.



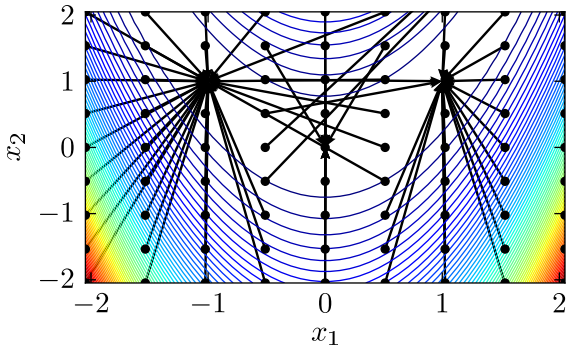
(a) Original Rosenbrock's function



(b) Krig of Rosenbrock's function (50 sample points shown in red 'plus')



(c) MFP iterations with $n_d = 2$



(d) NR fixed point iterations

Figure 10.8: Iterative schemes for finding stationary points of Rosenbrock's function Krig; grid size $= 9 \times 9$

10.5.4 Rastrigin's function

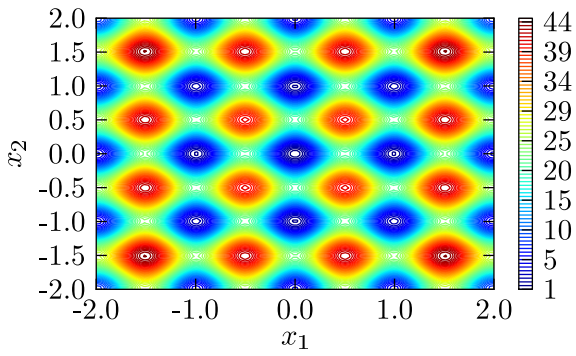
The Rastrigin's function is a highly multi-modal problem with regularly distributed minima. It is defined as

$$f(\mathbf{x}) \;=\; 10n + \sum_{i=1}^{n} \left[x_i^2 - 10\cos(2\pi x_i) \right] \quad ; \quad -2.0 \le x_i \le 2.0. \tag{10.78}$$

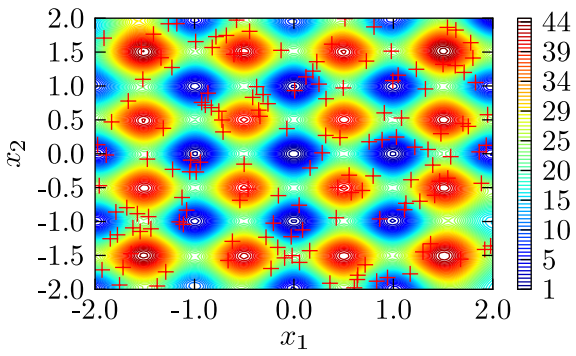
The presence of the cosine term in the function definition leads to the high multi-modality of this function. Its global minimum, $f(\mathbf{x}^*) = 0$ is located at $x_i^* = 0,\ i = 1 \ldots n$.

Figures 10.9a and 10.9b show, respectively, the original Rastrigin's function and its Krig created by sampling 150 points. Figures 10.9c and 10.9d show the behaviour of the MFP and NR sequences for a 21×21 grid of starting points. A finer grid is used here owing to the high

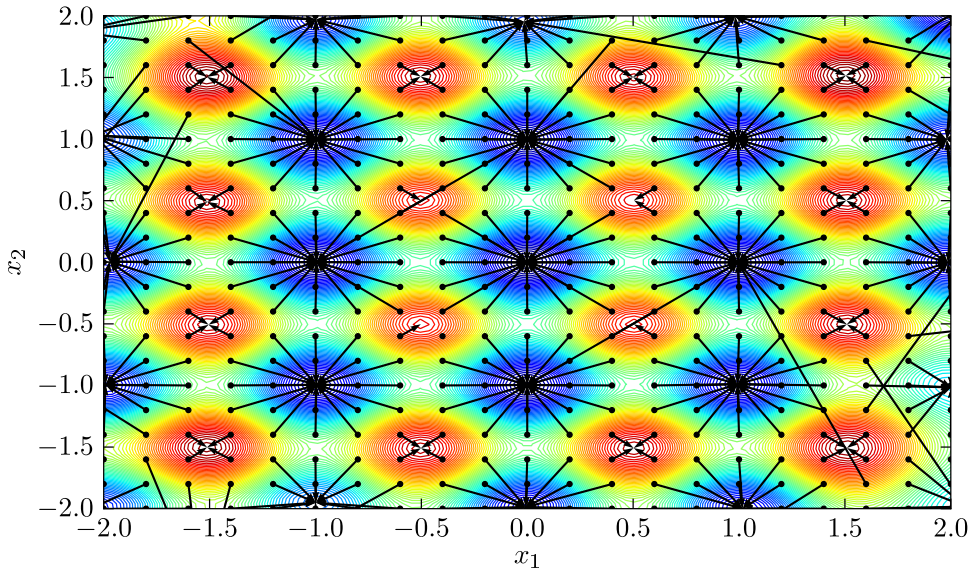
multimodality of the Rastrigin's function. As can be seein in Figure 10.9c, the MFP sequence is able to find most minima, albeit with a high value of $n_d = 10$. However, the MFP sequence misses all the saddle points, unlike the NR sequence, for which the basins of attraction for all the stationary points are well defined. This, if the goal of the procedure to find only the minima and maxima, lies in favour of the MFP method since size of the basins of attraction of the saddle points for the NR sequence, is gained by the basins of attraction of the minima for the MFP sequence. Nevertheless, if the goal was to find all the stationary points, then clearly NR method is superior.



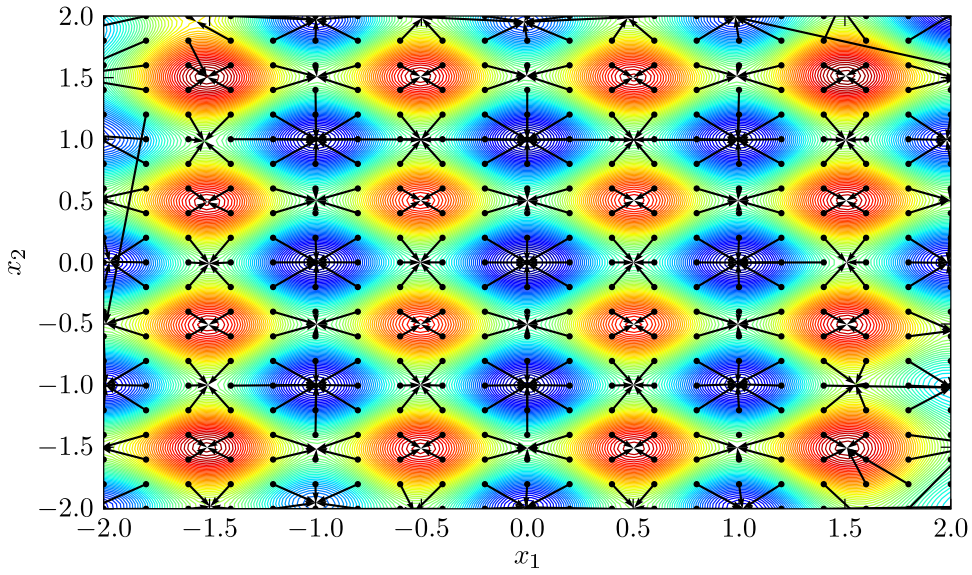
(a) Original Rastrigin's function



(b) Krig of Rastrigin's function (150 sample points shown in red 'plus')



(c) MFP iterations with $n_d = 10$



(d) NR fixed point iteration

Figure 10.9: Iterative schemes for finding stationary points of Rastrigin's function Krig; grid size $= 21 \times 21$

10.5.5 Bump Function

The Bump function, proposed by Keane [18], is a multi-modal function with several peaks of decreasing heights. It is defined as

$$f(\mathbf{x}) = \frac{-\text{abs}\left[\sum_{i=1}^{n}\cos^4(x_i) - 2\prod_{i=1}^{n}\cos^2(x_i)\right]}{\sqrt{\sum_{i=1}^{n} i x_i^2}} \quad ; \quad 0.0 \le x_i \le 10.0. \tag{10.79}$$

The bump function is widely used to test constrained optimisation algorithms, however its high multimodality makes it suitable to test the iterative sequences developed in this chapter. Figures 10.10a and 10.10b show, respectively, the original Bump function and its Krig created by sampling 200 points. Figures 10.10c and 10.10d show the behaviour of the MFP and NR sequences for a 21×21 grid of starting points. For the Bump function as well, it is observed that the MFP method is effective in finding the minima, with a high value of $n_d = 6$, but is not able to find the saddle points.

10.5.6 Ackley's Function

Ackley's function is also a widely used highly multi-modal test function. It is defined as

$$f(\mathbf{x}) = -a\exp\left[-b\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2}\right] - \exp\left[\frac{1}{n}\sum_{i=1}^{n}\cos(cx_i)\right] + a + \exp(1) \quad ; \; -2.0 \le x_i \le 2.0. \tag{10.80}$$

Its global minimum, $f(\mathbf{x}^*) = 0$ is located at $x_i^* = 0,\ i = 0 \ldots n$.

Figures 10.11a and 10.11b show, respectively, the original Ackley's function and its Krig created by sampling 150 points. Figures 10.11c and 10.11d show the behaviour of the MFP and NR sequences for a 21×21 grid of starting points. The behaviour of the two sequences for Ackley's function is very similar to that for Rastrigin's function. The difference between Ackley's function and Rastrigin's function is that the height of the peaks decreases considerably faster, further away from the centre for Ackley's function. This can effect the behaviour of population based optimisers, such as a fitness sharing genetic algorithm, but does not affect the performance of a gradient based optimiser, used to formulate the MFP and NR sequences.

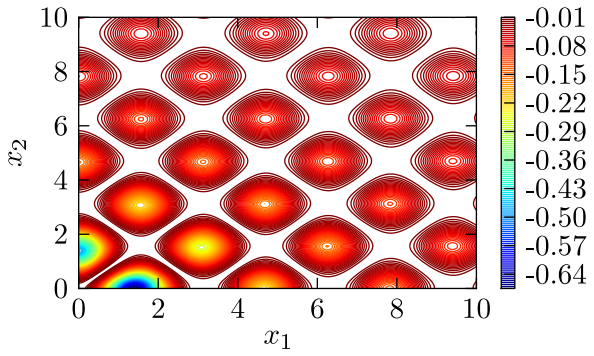
10.5.7 Six-hump camel back function

The Six-hump camel back function is a two variable function with six minima (two global and four local). It is defined as

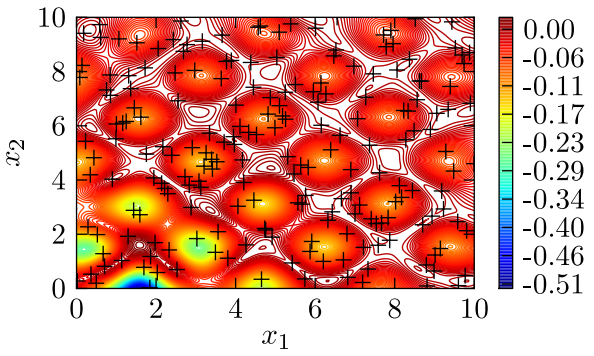
$$f(x,y) = \left(4 - 2.1x^2 + \frac{1}{3}x^4\right)x^2 + xy + 4(y^2-1)y^2 \quad ; \; -3 \le x \le 3 \; ; \; -2 \le y \le 2. \tag{10.81}$$

The two global minima, $f(x^*,y^*) \approx 1.0316$, are located at $(x^*,y^*) = (0.0898, -0.7126)$ and $(-0.0898, 0.7126)$.

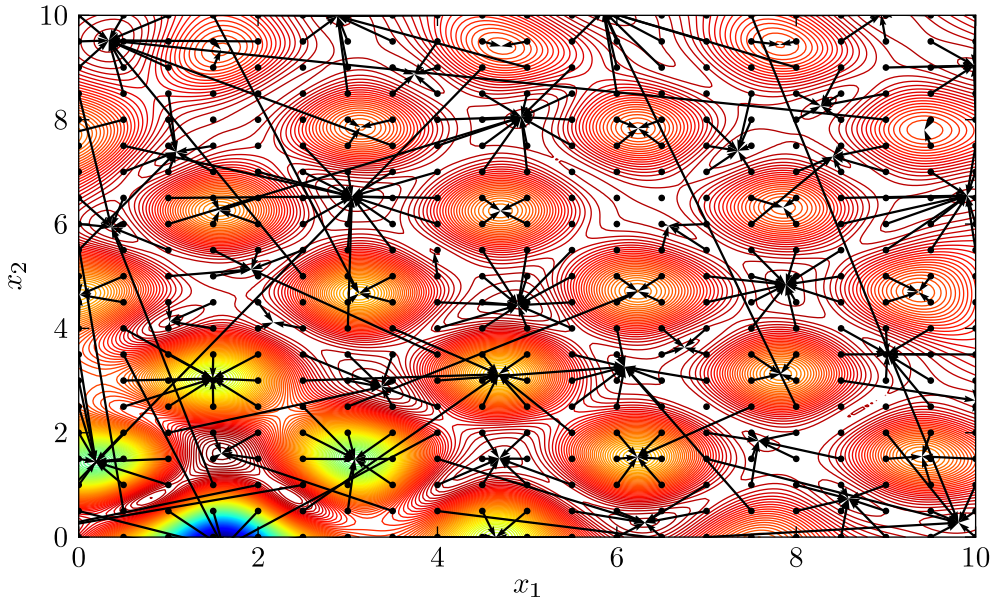
Figures 10.12a and 10.12b show, respectively, the original Six-hump camel back function and its Krig created by sampling 150 points. Figures 10.12c and 10.12d show the behaviour of the MFP and NR sequences for a 21×21 grid of starting points. For this function, it is found that an MFP sequence converges only for $n_d = 1$, i.e. for $n_d > 1$ all the sequences starting



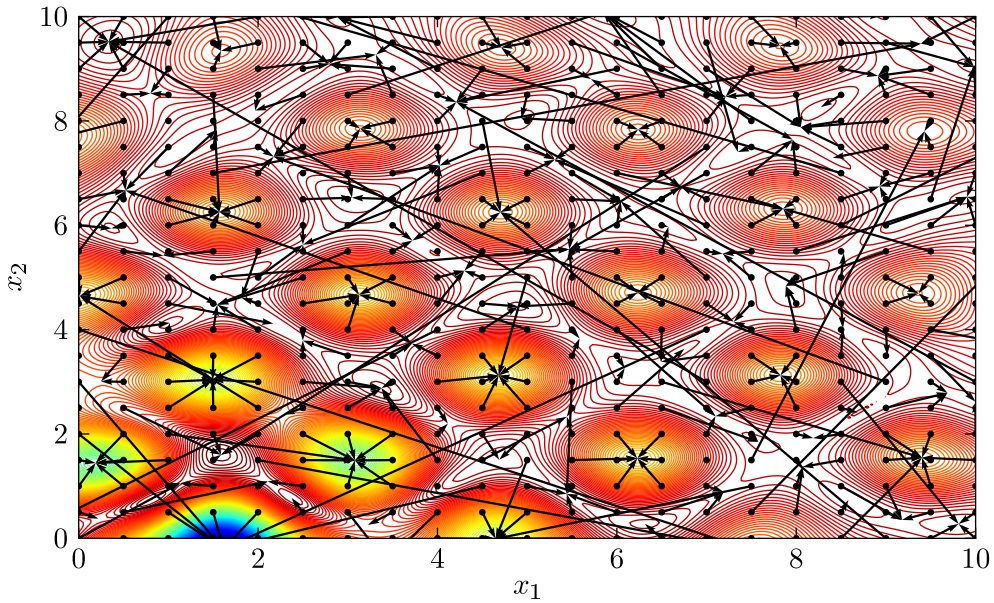
(a) Original Bump's function



(b) Krig of Bump's function (200 sample points shown in black 'plus')

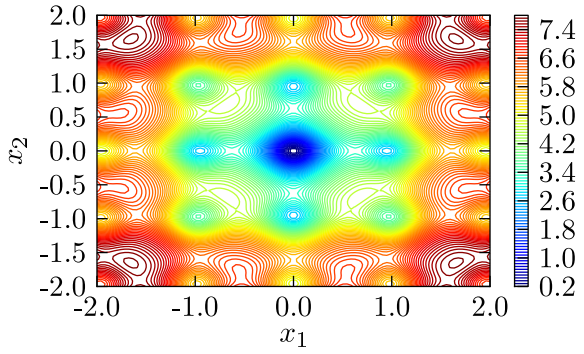


(c) MFP iterations with $n_d = 6$

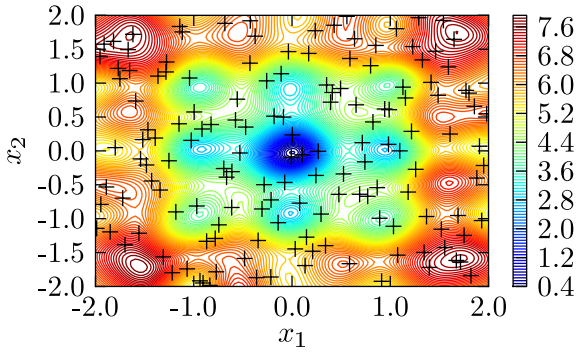


(d) NR fixed point iterations

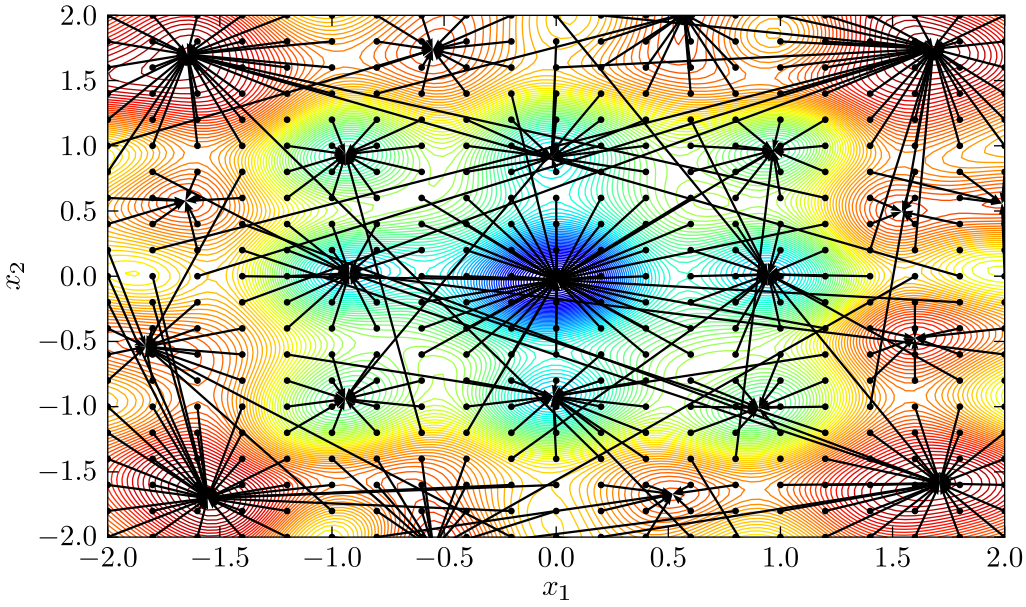
Figure 10.10: Iterative schemes for finding stationary points of the Bump function Krig; grid size $= 21 \times 21$



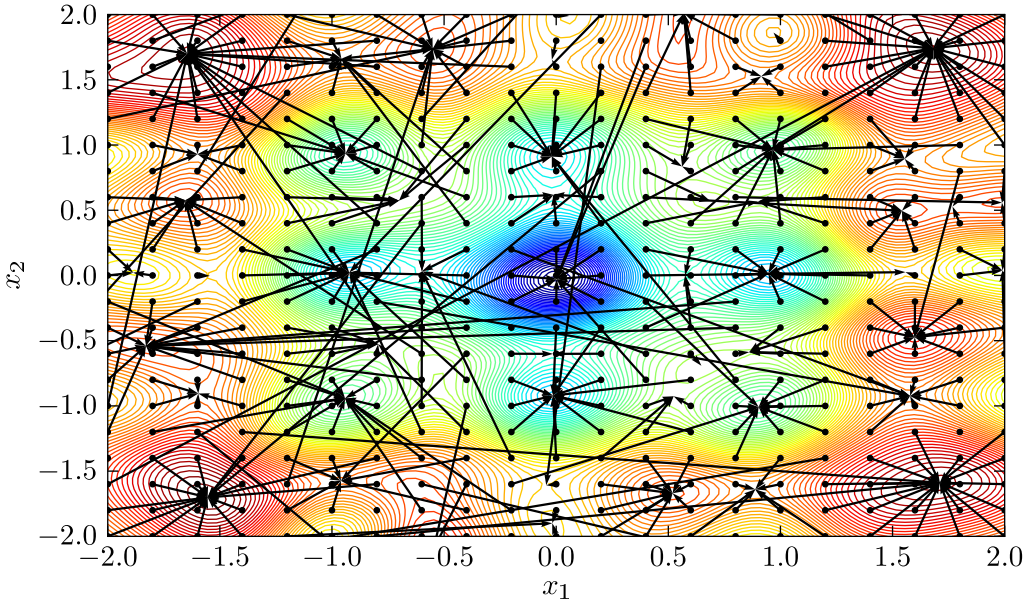
(a) Original Ackley's function



(b) Krig of Ackley's function (150 sample points shown in black 'plus')

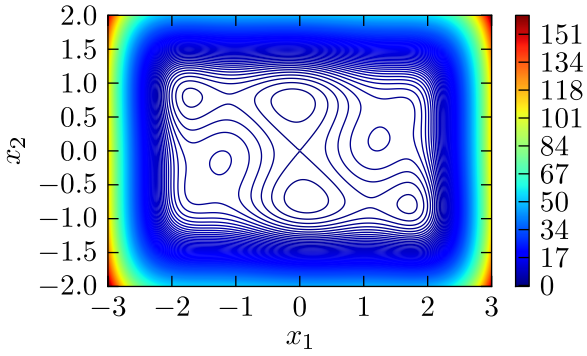


(c) MFP iterations with $n_d = 2$

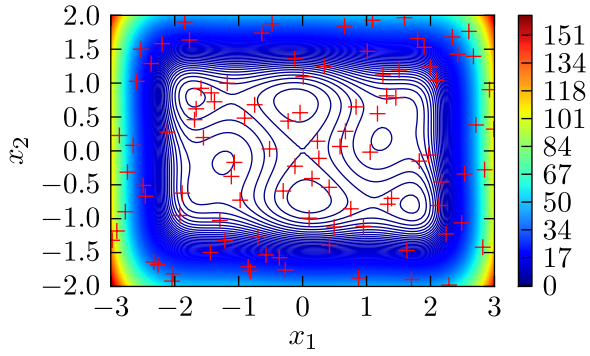


(d) NR fixed point iterations

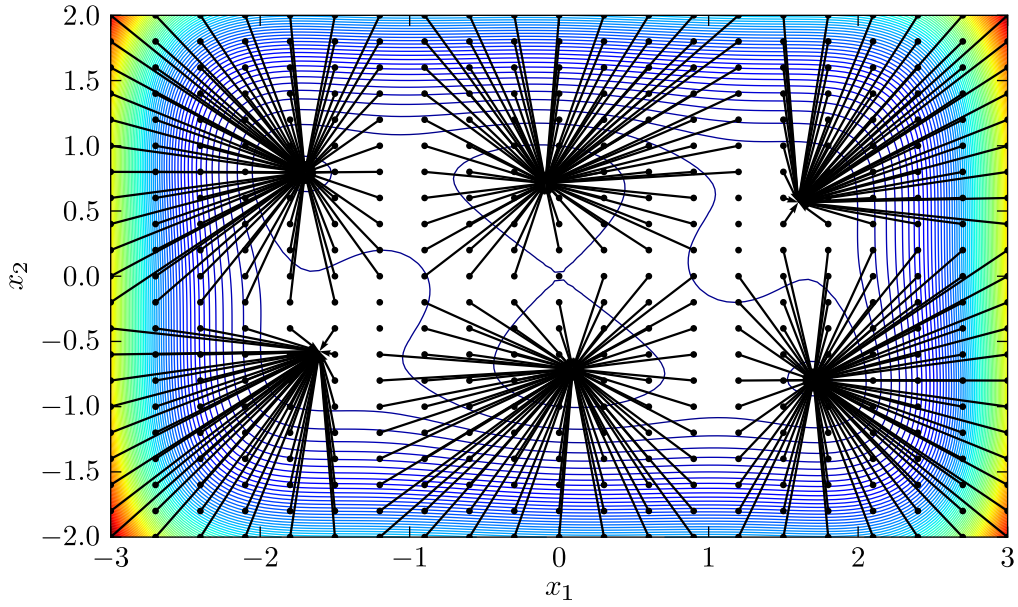
Figure 10.11: Iterative schemes for finding stationary points of Ackley's function Krig; grid size $= 21 \times 21$



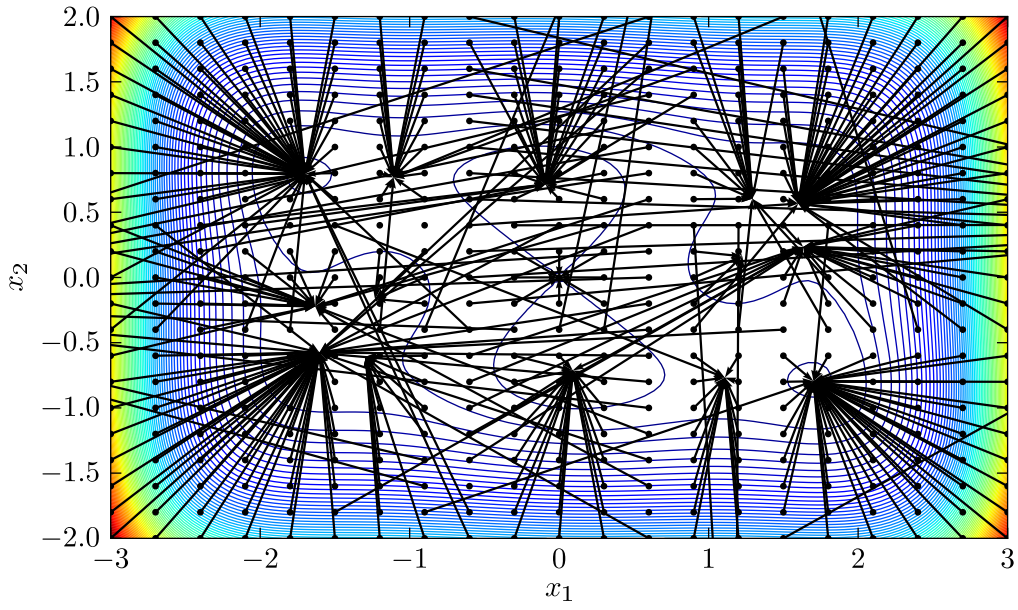
(a) Original Six-hump camel back function



(b) Krig of Six-hump camel back function (100 sample points shown in red 'plus')



(c) MFP iterations with $n_d = 1$



(d) NR fixed point iterations

Figure 10.12: Iterative schemes for finding stationary points of Six-hump camel back function Krig; grid size = 21×21

from all the points in the grid diverge. This is because, as discussed in section 10.3.5, the spectral radii of the six minima (found by $n_d = 1$) tend to one from the right as n_d increases. On the other hand, the NR sequence faces no such difficulties, and finds all the stationary points of the Krig.

10.5.8 Dropwave function

The Dropwave function is a highly multi-modal test function with only two variables and is defined as

$$f(x,y) = -\frac{1+\cos(12\sqrt{x^2+y^2})}{\frac{1}{2}(x^2+y^2)+2} \quad ; \; -1 \le x \le 1 \; ; \; -1 \le y \le 1. \tag{10.82}$$

Its global minimum, $f(x^*,y^*) \approx -1.0$, is located at $(x^*,y^*) = (0.0, 0.0)$. Figure 10.13 shows a surface plot of the Dropwave function. As can be seen this function has concentric circular shaped rings that represent alternate local minima and maxima. This is a challenging function for gradient based optimisers as the gradient at every point on each ring is zero and hence a local optimum, from a gradient point of view.

Figures 10.14a and 10.14b show, respectively, the original Dropwave function and its Krig created by sampling 150 points. Figures 10.14c and 10.14d show the behaviour of the MFP and NR sequences for a 21×21 grid of starting points. As can be seen from these figures, both the sequences seem apparently confused and converge to various points in the circular rings. Moreover, no MFP sequence converges to any of the rings that represent a local maximum. However, starting from a point inside the central basin of the function, both the sequences converge to the global minimum at $(x^*,y^*) = (0.0, 0.0)$.

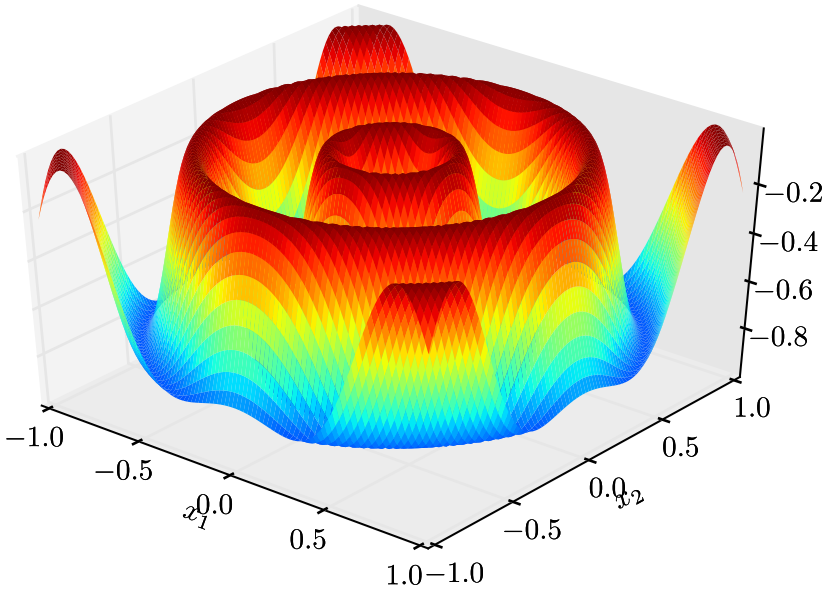
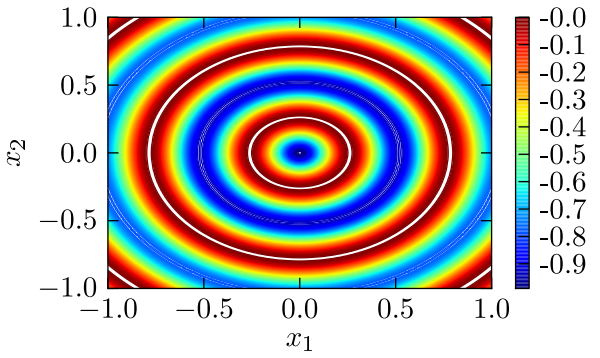
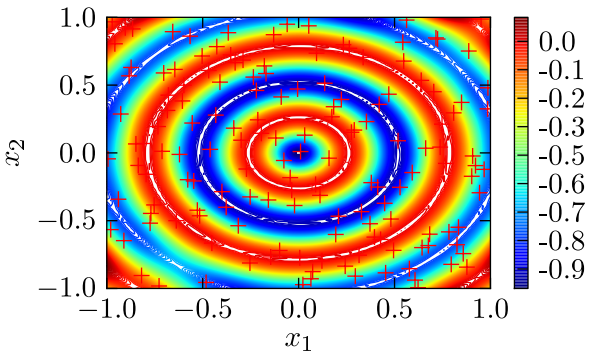


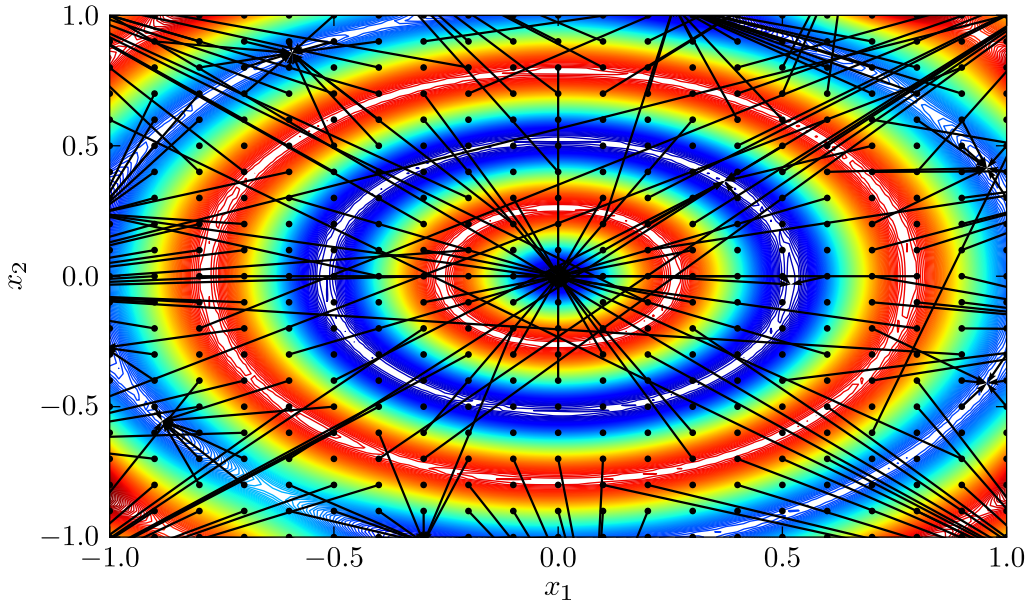
Figure 10.13: Surface plot of the Dropwave function



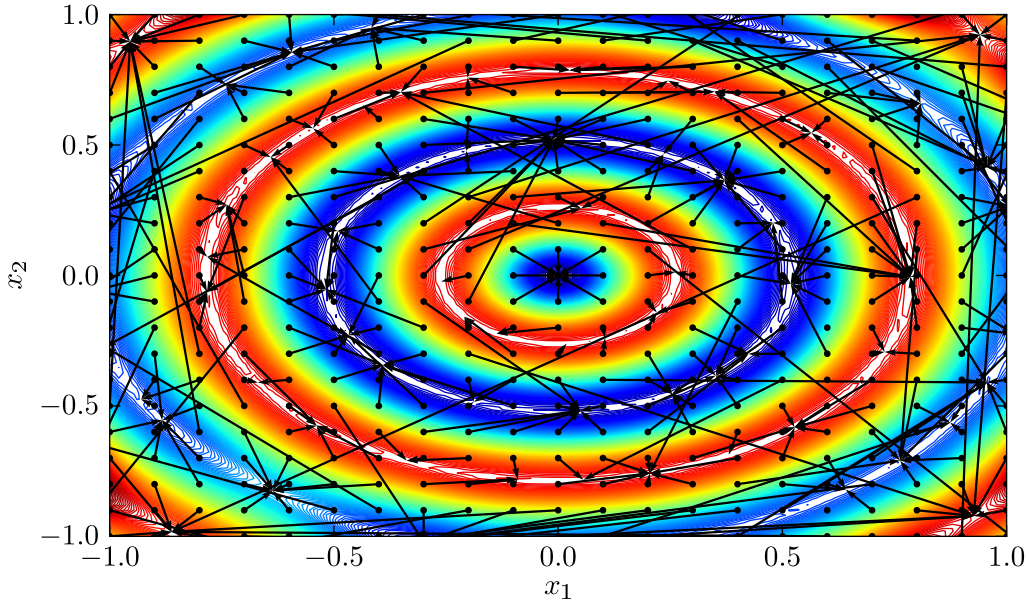
(a) Original Dropwave function



(b) Krig of Dropwave function (150 sample points shown in red 'plus')



(c) MFP iterations with $n_d = 6$



(d) NR fixed point iterations

Figure 10.14: Iterative schemes for finding stationary points of Dropwave function Krig; grid size $= 21 \times 21$

10.5.9 Easom's function

The Easom function is a unimodal test function with two variables. It is a difficult function to optimise as the global minimum has a very small basin of attraction and the function everywhere else is relatively flat, i.e. it yields no information about the location of the global minima. The Easom function is defined as

$$f(x,y) \;=\; -\cos(x)\cos(y)\exp\left[-(x-\pi)^2-(y-\pi)^2\right] \quad ;\; -20 \le x \le 20 \;;\; -20 \le y \le 20. \tag{10.83}$$

Its global minimum, $f(x^*,y^*) \approx -1.0$, is located at $(x^*,y^*) = (\pi,\pi)$.

A surface plot of Easom's function is shown in Figure 10.15. The relative flatness of the function and the relatively small size of the basin of the global minimum can readily be observed from this figure. Figures 10.16a and 10.16b show, respectively, the original Easom's function and its Krig created by sampling 200 points. A high sample size for this function is needed to capture the flatness of the function in an extremely large part of the domain. Figures 10.16c and 10.16d show the behaviour of the MFP and NR sequences for a 21×21 grid of starting points. Both the sequences converge to various local minima of the Krig distributed in the search domain. For the grid size chosen, only 4 points for the MFP sequence and 5 points for the NR sequence, that fall within the basin of the global minimum, converge to the global minimum.

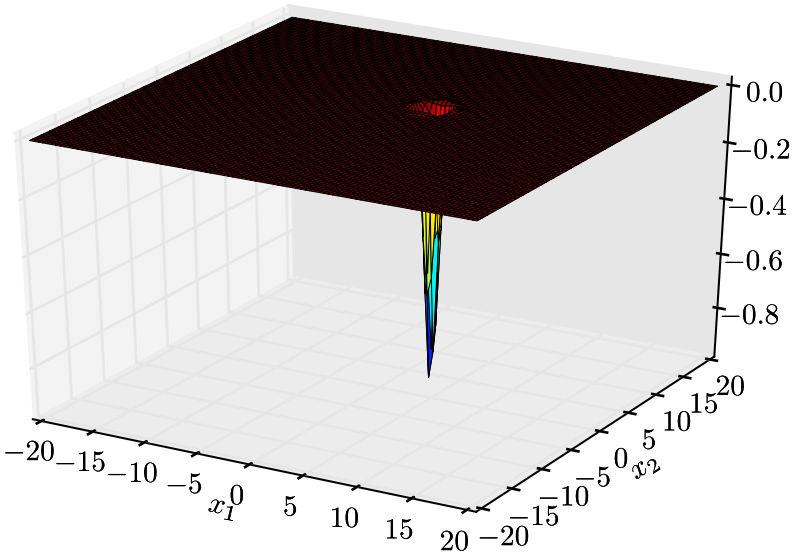
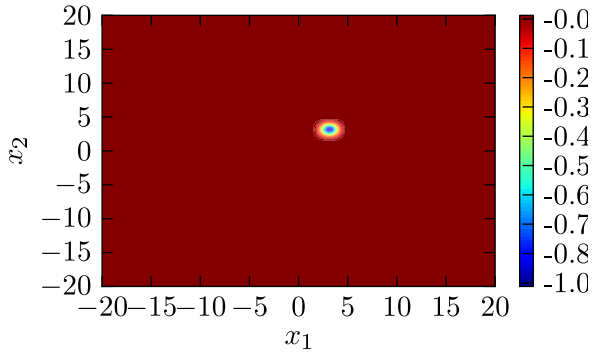
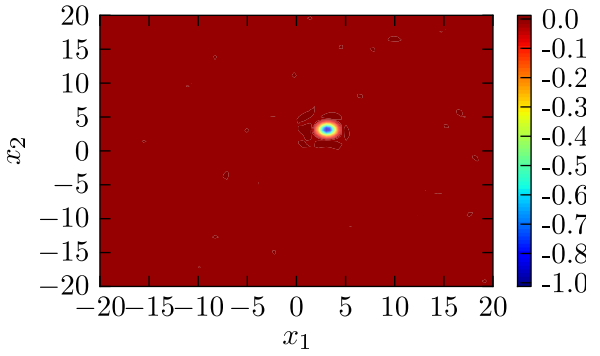


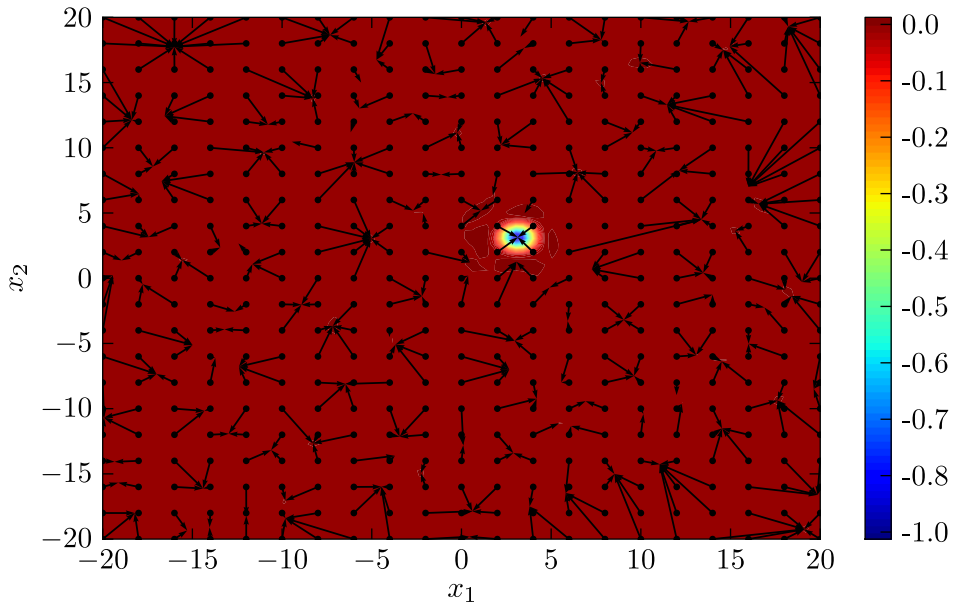
Figure 10.15: Surface plot of Easom's function



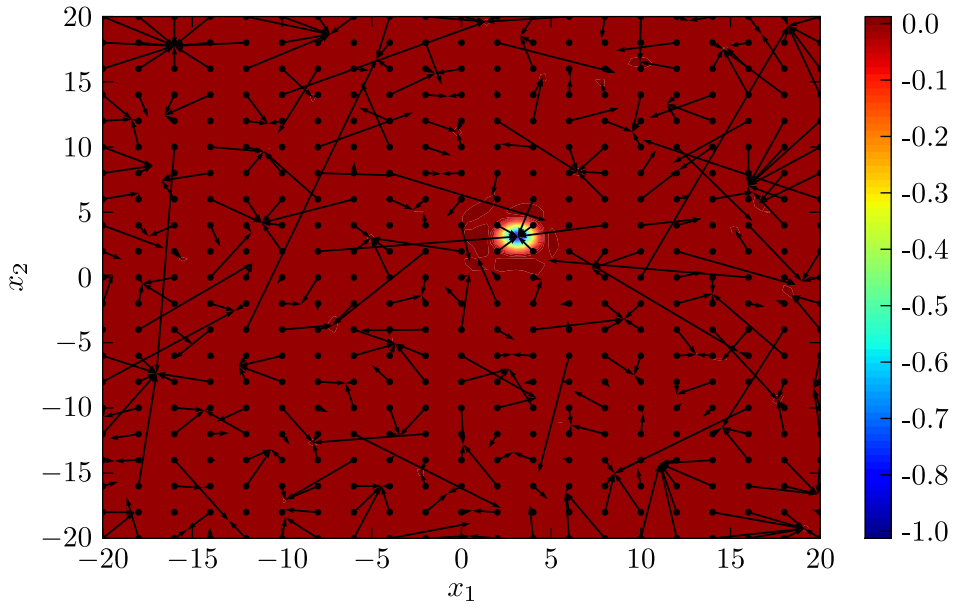
(a) Original Easom's function



(b) Krig of Easom's function (200 sample points not shown)



(c) MFP iterations with $n_d = 2$



(d) NR fixed point iterations

Figure 10.16: Iterative schemes for finding stationary points of Easom's function Krig; grid size $= 21 \times 21$

10.5.10 Goldstein-Price's function

The Goldstein-Price's function is a multi-modal test function with two variables. It is defined as

$$\begin{aligned} f(x,y) \;=\; & \Big[1+(x_1+x_2+1)^2(19-14x_1+3x_1^2-14x_2+6x_1x_2+3x_2^2)\Big]. \\ & \Big[30+(2x_1-3x_2)^2(18-32x_1+12x_1^2+48x_2-36x_1x_2+27x_2^2)\Big] \\ & -2\le x\le 2\;;\; -2\le y\le 2. \end{aligned} \tag{10.84}$$

Its global minimum, $f(x_1^*, x_2^*) \approx 3.0$, is located at $(x_1^*, x_2^*) = (0, -1)$.

This function is challenging for the optimisation algorithms as the local minima lie in relatively flat region. Figures 10.17a and 10.17b show, respectively, the original Goldstein-Price's function and its Krig created by sampling 150 points. Figures 10.17c and 10.17d show the behaviour of the MFP and NR sequences for a 21×21 grid of starting points. Interestingly, for no value of n_d in the MFP sequence, are any of the desired stationary points in the domain found. The NR sequence, however, finds all the stationary points.

Based on the aforementioned results, the following conclusions can be made regarding the FP and NR sequences:

1. The FP sequences finds stationary points, but is limited by the spectral radii of the stationary points. This limitation can be overcome up to a certain extent by using higher values of n_d. However, this still does not guarantee convergence of this sequence to all stationary points. Moreover, how to estimate the n_d factor *a priori* is not clear.

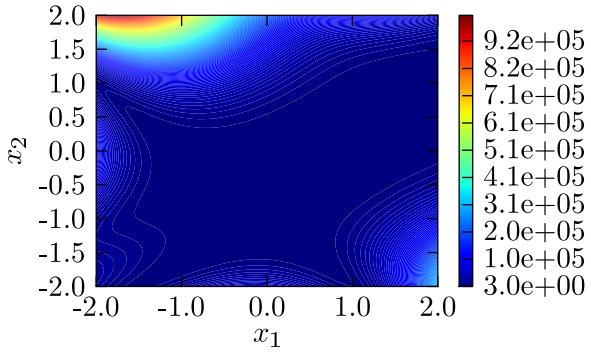
2. The NR sequence guarantees convergence to all stationary points, provided the starting point of the sequence lies within the basin of attraction of the stationary points.

10.6 Development of the Krige-Newton-Raphson-Sobol (KNRS) method

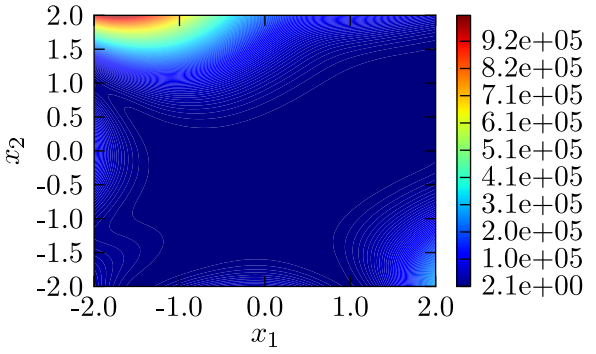
In this section an optimisation algorithm called the Krige-Newton-Raphson-Sobol (KNRS) method is proposed for global and multimodal optimisation.

From the results of the previous section, it is clear that the NR sequence is more robust in finding the stationary points when compared to MFP sequence. Moreover, based on the results and theorem 10.3.2, a strong proposition is that as long as an algorithm can get one starting point in the basin of attraction of the stationary point, the stationary point is guaranteed to be found. Finally, Newton-Raphson iterations, starting from a point sufficiently close to the stationary point, converge *q-quadratically* to the stationary point, when the Jacobian matrix is non-singular (see equation 10.73) [195]. This means that the error, difference between the solution ($\mathbf{u}^*$) and current point in the sequence ($\mathbf{u}^\kappa$), at $\kappa+1^{th}$ iterate is related to the error at κ^{th} iterate by $\|\mathbf{u}^{\kappa+1}-\mathbf{u}^*\| \le \tau \|\mathbf{u}^\kappa-\mathbf{u}^*\|^2$, for some $\tau > 0$ [196].

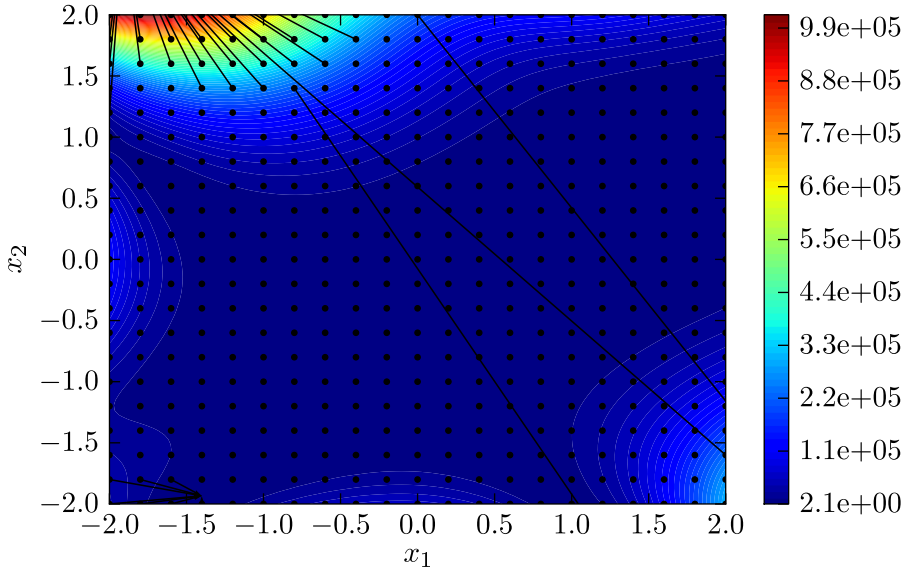
The above thoughts lead to a hypothesis that if an algorithm chooses successive starting points for the NR iterative sequence carefully, i.e. to explore regions of the search space uniformly, then all stationary points of the Krig should be found. For choosing successive starting points, a quasi-random (low discrepancy) sequence, called the SOBOL sequence (or the LP_τ sequence) [202, 203] is proposed. The SOBOL sequence has an attractive property in



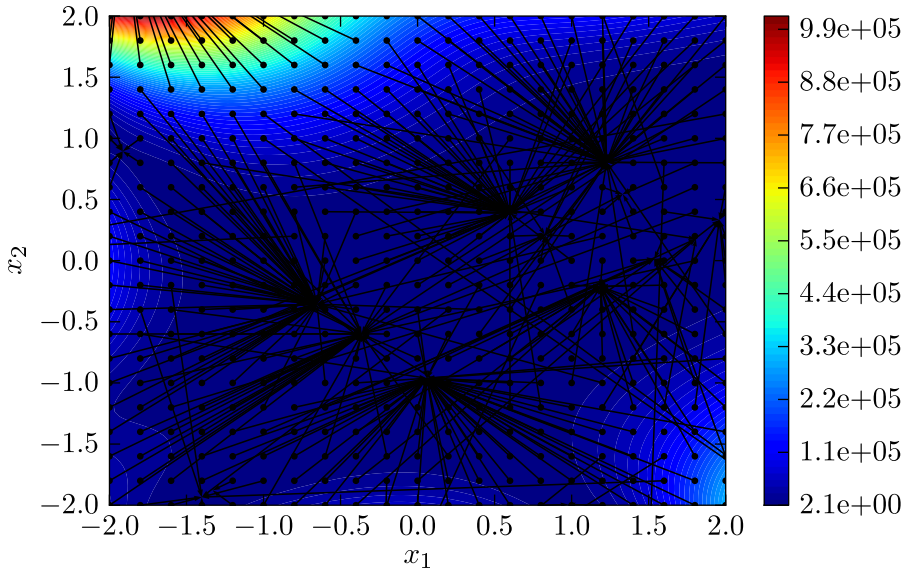
(a) Original Goldstein-Price's function



(b) Krig of Goldstein-Price's function (150 sample points not shown)

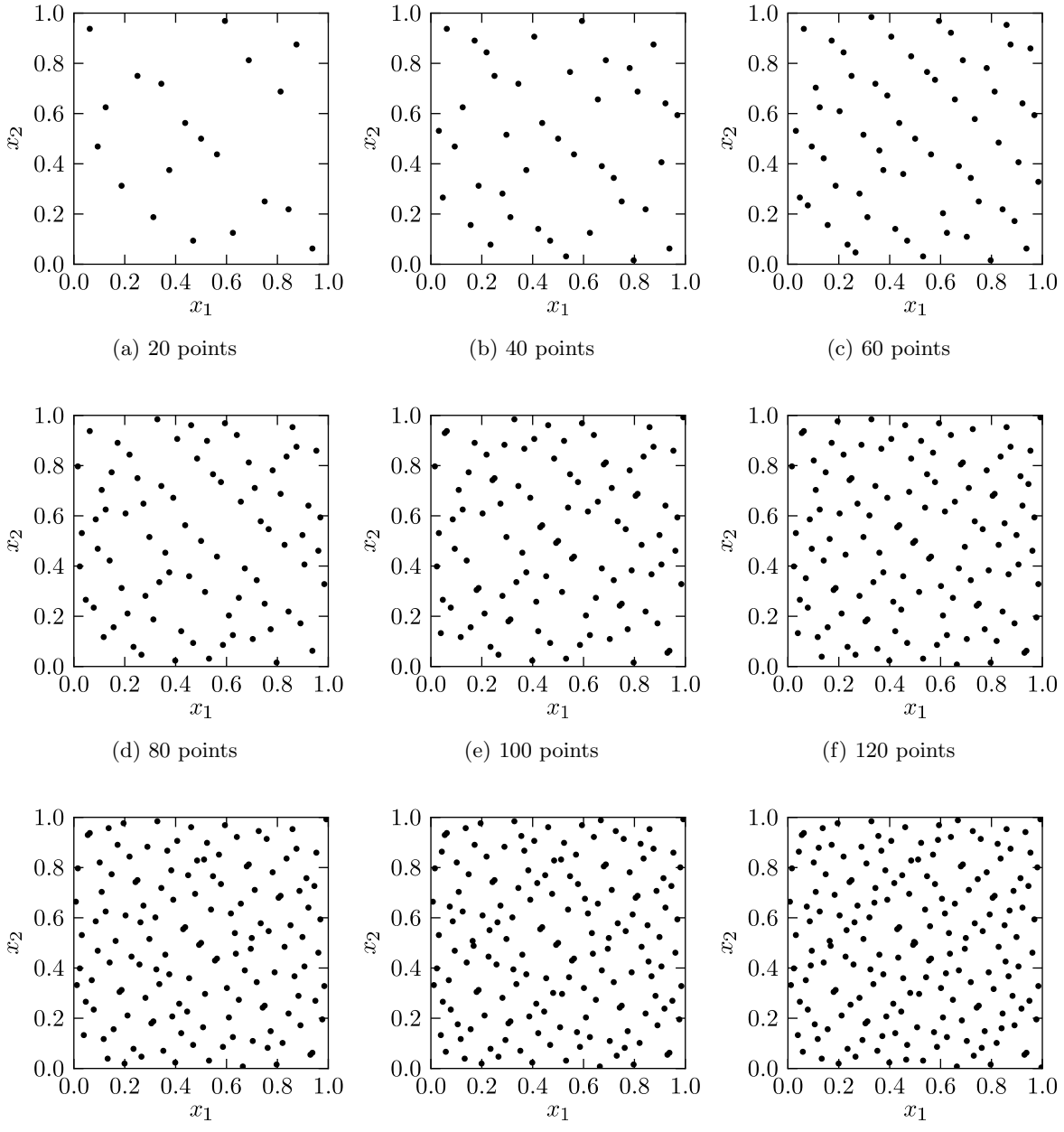


(c) MFP iterations with $n_d = 2$



(d) NR fixed point iterations

Figure 10.17: Iterative schemes for finding stationary points of Goldstein-Price's function Krig; grid size = 21×21



(a) 20 points (b) 40 points (c) 60 points

(d) 80 points (e) 100 points (f) 120 points

(g) 140 points (h) 160 points (i) 180 points

Figure 10.18: 20 points added successively from the LPτ sequence

that the points are uniformly space filling, i.e. as more points from the sequence are added the space is sampled more, uniformly. This property makes these sequences a reasonable choice for exploring a design space. Figure 10.18 shows an instance of this sequence in two dimensions where starting from a random point in the sequence, 20 points are added successively. Figure 10.19 shows 200 points from this sequence starting from a random point in the sequence. The space filling properties of such a sequence are apparent from these figures. For all the results presented in this chapter, the SOBOL sequences are generated by the code provided by Dr. John Burkardt at his homepage [204].

Based on the above reasoning, the KNRS algorithm is described as follows and depicted in figure 10.20. The algorithm begins by developing the necessary Kriging predictor for the required function or physical response. Then a random skip number (a positive integer, i) [205] is generated. This number means that if l points are needed from the SOBOL

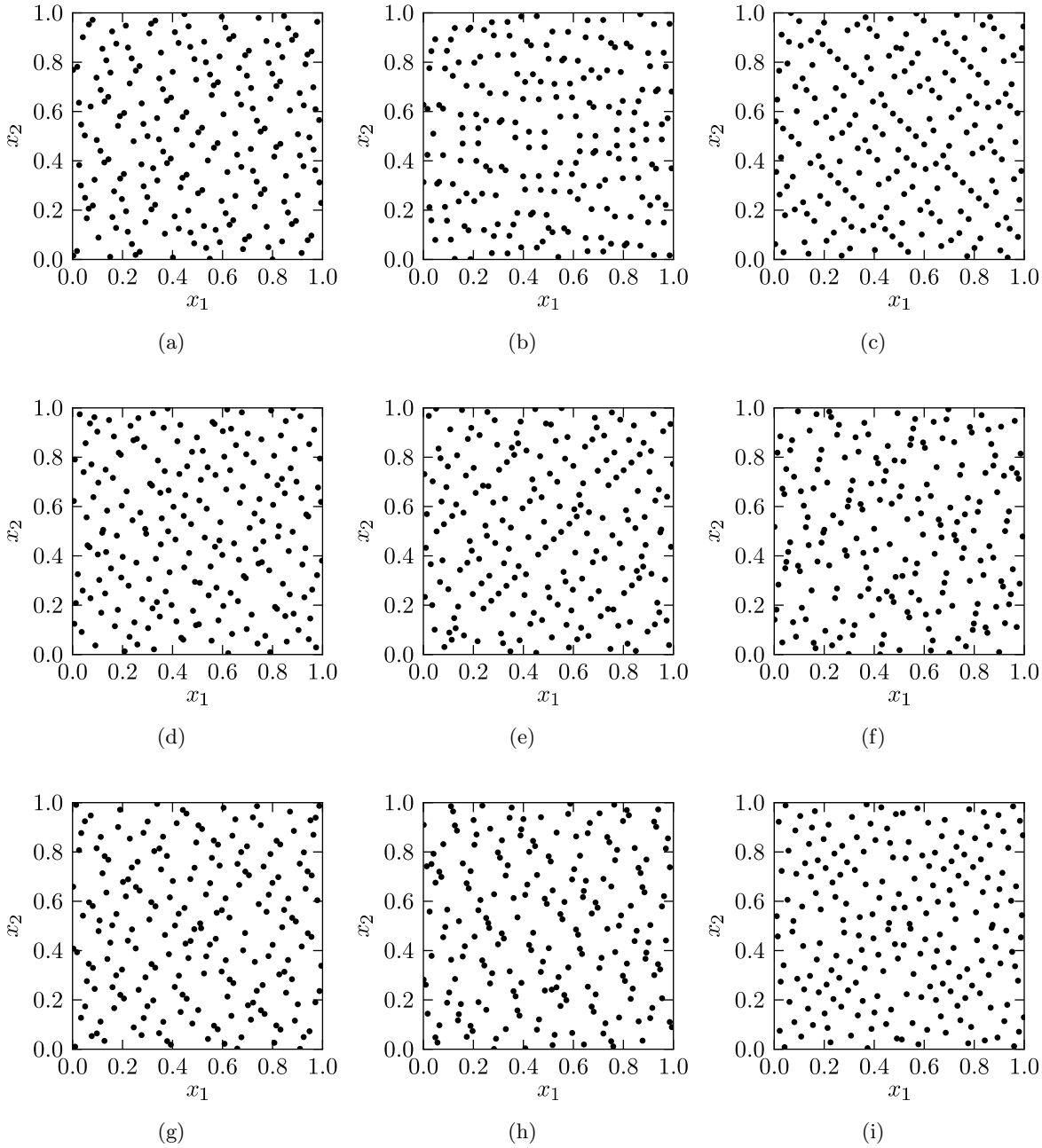


Figure 10.19: 200 points from the LPτ sequence with a random start between 0 and 10,000

sequence, $\mathbb{S} = (\mathbf{S}^1, \mathbf{S}^2, \mathbf{S}^3, \ldots, \mathbf{S}^n)$, then instead of taking the first l points from the sequence, $(\mathbf{S}^1, \mathbf{S}^2, \mathbf{S}^3, \ldots, \mathbf{S}^l)$, the points used are $(\mathbf{S}^{i+1}, \mathbf{S}^{i+2}, \mathbf{S}^{i+3}, \ldots, \mathbf{S}^{i+l})$. A NR iteration (equation 10.73) is then performed starting from the first point $\mathbf{S}^{i+1}$. This iteration either diverges, oscillates, or converges to a stationary point. If the iteration converges, then the results are appended to an offline list of found stationary points. Next, the second point from the SOBOL sequence, $\mathbf{S}^{i+2}$ is chosen as the starting point and a NR iteration carried out from this point. This procedure of taking the next point from the SOBOL sequence and running an NR iteration starting from this point is repeated until any of the convergence criteria are met: (a) either the computational budget is reached, or (b) if the user is satisfied by the hitherto found minimum value from the stationary points, or (c) the number of desired stationary points are found. As an additional step on the list of unique stationary points the following two operations can be performed: first, if the user is only looking for the global minimum, then

the list can be sorted and the minimum value point can be to the user; and second if the user is interested in differentiating between minima, maxima, and the saddle points, then each of the stationary points found can be classified into these categories by performing $2m$ function evaluations, where m is the dimension of the function, around each stationary point and evaluating the function value at these points.

To conclude the formulation of the KNRS algorithm, the following three important points underlying the motivation to develop the KNRS algorithm are highlighted.

1. NR iteration guarantees quadratic convergence to the stationary point, if starting sufficiently close (i.e. in the basin of attraction) of a stationary point. Since an exact expression, i.e. not evaluated by finite differencing, is used for the Kriging predictor, the NR sequence proposed is very robust, i.e. free from discretisation errors.

2. SOBOL sequences sample the space uniformly. Starting anywhere from a SOBOL sequence, successive points are space filling, and hence explore all the regions of the search domain equally.

3. A combination of the local search provided by the NR sequence coupled by the exploratory power of SOBOL sequence offers a powerful paradigm to perform optimisation, and forms the basis for the KNRS algorithm proposed.

10.7 Comparison of the KNRS algorithm with other algorithms for Global optimisation

In this section, the above proposed KNRS algorithm is compared to two algorithms: a real coded Genetic algorithm [36] and dynamic hill climbing [206, 13], for testing global optimisation capabilities. In the following sub-sections, these algorithms are first briefly described and then the results of the comparison are presented.

10.7.1 Genetic algorithms

The working of a Genetic algorithm is described in section 3.3.4.3. In all the comparisons made in this section, a real coded GA with a binary tournament operator is used for selection, the SBX operator is used for crossover, and polynomial mutation operator is used for mutation. The GA used is coded in the Python programming environment [207] by the author. The parameter settings used for comparison are tabulated in Table 10.3.

10.7.2 Dynamic Hill Climbing

Dynamic Hill Climbing (DHC) is a search algorithm proposed by Yuret and Maza [206, 13] which is based on the ideas of genetic algorithms, hill climbing, and conjugate gradients. Yuret and Maza, in their publications, showed this algorithm to be very effective on a test-suite of De Jong's five functions. Keane [188] has used DHC for the optimisation of Kriging hyperparameters.

The central idea of Dynamic Hill Climbing lies in having an inner and outer loop: the inner loop being a hill climber which starts from a seed point and performs a hill climb;

BEGIN KNRS

Step 0: Start with a sampling plan (such as a Latin hypercube sampling plan) to generate points where the function is sampled.

Step 1: Sample the function (Analysis: CFD, FEA etc.) at the sample points.

Step 2: Construct the Kriging predictor for the function

Step 3: If errors for the predictor are less than the required accuracy of the Krig then go to **Step 5**

Step 4: Estimate additional number of points to be added to improve the Kriging predictor based on an in-fill plan (such as the expected improvement criterion). Return to **Step 1**

Kriging

Step 5: Generate a random *skip* number (an integer i) for the SOBOL sequence. From the SOBOL sequence, pick the $i+1^{th}$ point, $\mathbf{S}_{i+1}$. Store this point in variable $\mathbf{x}_{start}$, i.e. $\mathbf{x}_{start} = \mathbf{S}_{i+1}$. Set counter, $\kappa = 1$

Step 6: Starting from the point $\mathbf{x}_{start}$, run the NR sequence (equation 10.73)

Step 7: If the iteration in **Step 6** converges go to **Step 8**, otherwise $\mathbf{x}_{start} = \mathbf{u}_{i+2}$, go to **Step 6**

Newton-Raphson

Step 8: Append the converged location of the iteration in **Step 6** to the list of stationary points found.

Step 9: Find the unique points from the list of stationary minima found; update the list of unique stationary points found.

Database

Step 10: Check stopping criterion: (a) if the computational budget is reached, or (b) if the minimum value found so far is satisfactory, or (c) if the number of desired unique stationary points are found. If not, then $\mathbf{x}_{start} = \mathbf{S}_{i+2}$, $\kappa = \kappa + 1$, go to **Step 6**.

SOBOL

Step 11: If needed, perform a check on the unique stationary points found to classify them as maxima, minima, or saddle points.

Classification

END KNRS

Figure 10.20: KNRS algorithm

Table 10.3: GA settings for comparion of KNRS for global optimisation

Parameter	Value
Coding	Real-coded
population size	50
number of generations	50
Selection operator	Binary tournament
Crossover operator	Simulated Binary Crossover (SBX)
SBX parameter (η_c)	10.0
Crossover probability	0.5
Mutation operator	Polynomial mutation
Mutation probability	0.05
Polynomial mutation parameter (η_m)	10.0

and the outer loop seeding the next start point far from the already found local minima. The inner loop has a dynamic coordinate frame: initially the directions of movement are the coordinate axis directions and a step length in these directions is decided. Hence, for a problem of n dimensions, $2n$ evaluations are made and a move is made along the direction of the best point found. If none of the $2n$ points are better than the current point, the step length is halved and another $2n$ evaluations are carried out. Additionally, the previous successful directions of movement are stored in a buffer. After every successful move, a Gram-Schmidt orthogonalization procedure is used to construct a new coordinate frame based on the difference of previous moves. This provides additional flexibility as the set of orthogonal search directions are not fixed to the coordinate axes. Convergence is determined when the step length becomes smaller than a predetermined limit. The outer loop's goal is to seed the next starting point far from the previously found local minima, which are stored in a buffer, in order to promote exploration. Yuret and Maza proposed several methods for this. For binary coded algorithms they proposed maximising the Hamming distance (difference in the bits of two vectors) of the new point from already found local minima. They [13] also proposed two other methods for the outer loop: first, to choose a set of points randomly in the search space and initialize the inner loop with the best among these points; second, to divide the search space into hyper-rectangles based on the local minima already found, and choosing a random point in the hyper-rectangle of largest area (hyper-volume). The latter procedure is similar to the binary search algorithm proposed by Hughes [208].

DHC has three parameters to set: a) the first starting point; b) the starting step length; and c) the minimum step length. For all the DHC comparisons in this chapter, the first starting point is chosen at random within the search domain, the starting step length is set $D/2$, where, D is the difference between the upper and lower bound on decision variables (as suggested by the authors), and the minimum step length is set to 10^{-3}. The DHC used in this study is coded by the author in the Python programming environment [207]. For the outer loop, the method of evaluating 50 random points in the search domain and choosing the best among these as the seed for the inner loop is used. The results of this code are validated against the results reported by the Yuret and Maza [13] for a test suite of De Jong's five functions, as shown in Figures 10.21 and 10.22. Comparing these figures it can be seen that for all the test functions, except the Quartic function, the mean of the best fitness value reaches the minimum at approximately similar number of function evaluations for the two implementations: $\approx$ 100 for Sphere function; $\approx$ 1,000 for Rosenbrock's function; $\approx$ 10,000 for

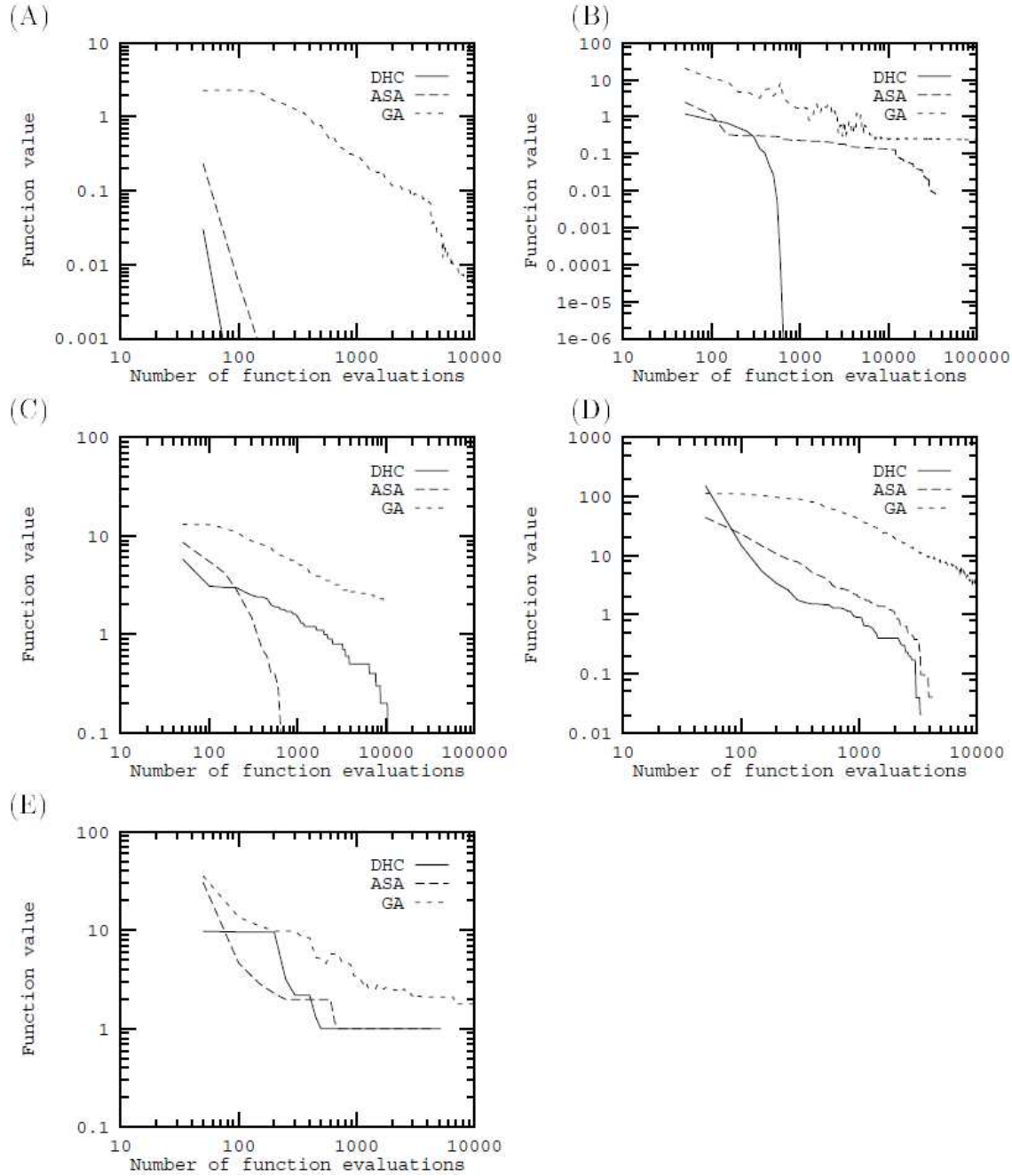
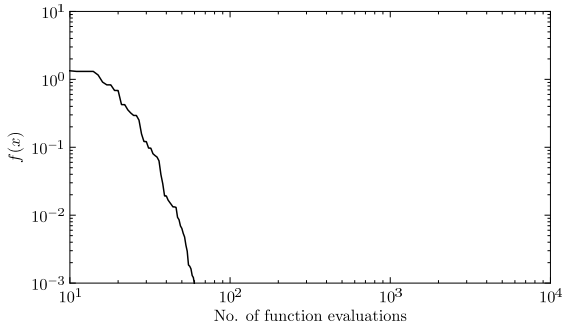
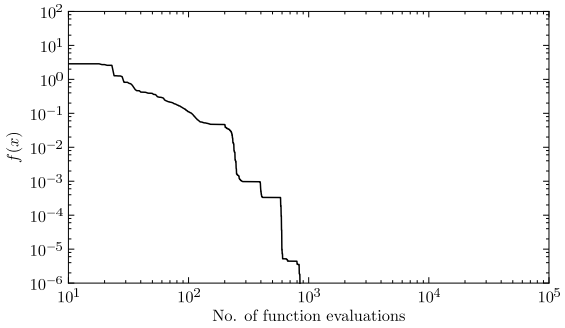


Figure 10.21: Best point average over 10 runs for five functions by the developers of DHC [13]: (A) Sphere function, (B) Rosenbrock's function, (C) Step function, (D) Quartic function, and (E) Shekel's function [13]

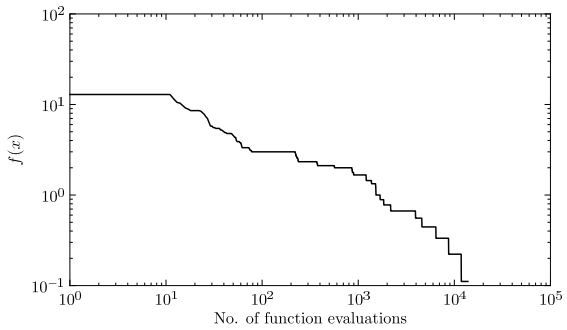
Step function; and ≈ 1,000 for the Shekel's function.



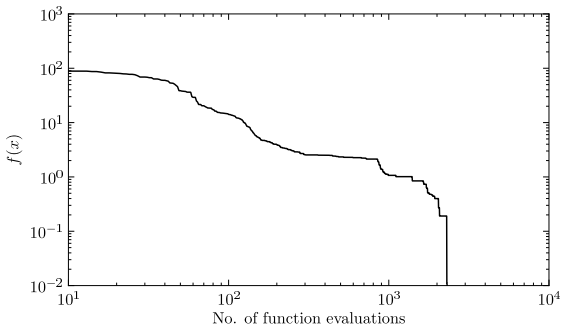
(a) Sphere function



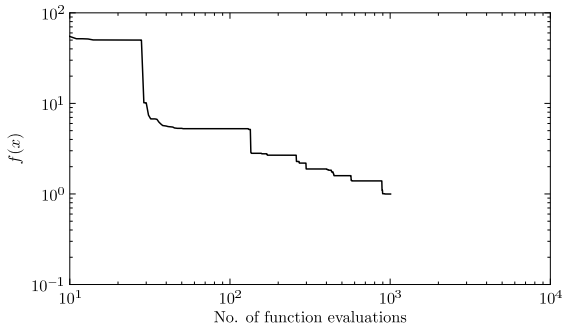
(b) Rosenbrock's function



(c) Step function



(d) Quartic function



(e) Shekel's function

Figure 10.22: Best point average over 10 runs for five functions to validate the DHC code (Author's code)

10.7.3 Comparison results and discussion

To compare the performance of KNRS against the performance of GA and DHC, 50 runs of all the three algorithms are performed for the 10 test functions in table 10.2.

Figures 10.23 – 10.32 show the optimisation histories for the best point and the mean of the best point for 50 runs of the KNRS, GA, and DHC algorithms on the 10 test functions. Table 10.5 shows the corresponding times taken for the 50 runs, and time per evaluation. Before any comments on these numbers, the working of the KNRS algorithm is explored. Let us consider Figure 10.23a. One observation is that since each NR iteration converges to one of the stationary points of a function, the best point values in any run are likely to be the stationary point function values for a long time, i.e. until any better point is found. This leads to a band formation around the function values of the stationary points. In other words, depending on the convergence accuracy of the inner NR iterations, the converged values of the NR iterations lie in a small interval around the stationary points function values. For example for the Branin function, the function value around the stationary points is approximately 20 and hence a a dark line (i.e. a set of lines corresponding to this band) is observed in Figure 10.23a. To further understand the working of this algorithm two independent runs on the Branin function and key points along the run are shown in Table 10.4. Comparing the first run in Table 10.4 to Figure 10.23a, it can be seen that this is the only run on Figure 10.23a that has not converged to any of the minima of Branin function Krig. The most likely reason for the non-convergence of this run is that a NR sequence is actually diverging. This is an important issue, especially when a fixed number of evaluations are allowed in the algorithm. In the case of the Branin function, this limit is 50 evaluations. Now, if the number of iterations in an NR sequence (without convergence) before which it is declared that the sequence is either oscillating or diverging, is greater than 50, then if the first point of the SOBOL sequence starts to diverge, then there will be no function improvement in the entire run. This issue can be resolved by either checking inside each iteration as to whether the series is diverging or oscillating, or by keeping a sensible limit on the maximum iterations for the NR sequence. In all the comparisons presented in this section, the latter approach is adopted and the limit on maximum iterations in an NR sequence is set to 50. The effect of the diverging solution in the case of Branin function is also reflected in the plot for the mean of the best point (Figure 10.23b): the dashed line represents the mean of all the 50 runs and the solid line shows the mean when the run with the diverging solution is removed. This is a direct result of choosing 50 as the inner iteration limit, and the fact that only 50 evaluations are considered in these runs, thereby disallowing any further point to be evaluated before the maximum evaluation limit is reached.

Returning to the comparison of the three algorithms on the 10 test functions (Figures 10.23 – 10.32), it is important to compare the time taken by each algorithm per evaluation. Each algorithm has a different mechanism to converge to the optima, based on what the algorithm does after evaluating a point. It should be noted that the time per evaluation here implies, the time taken by one complete inner iteration of an algorithm, i.e. time taken to evaluate the function value once plus other inner workings of the algorithm. For example, time per evaluation for KNRS includes time taken for generation of the next point in the SOBOL sequence plus one iteration of the NR sequence and convergence checks within the iteration. For the GA, time per evaluation is the time taken for selection, crossover, and mutation, per generation divided by the number of members in a generation. Similarly, for

Table 10.4: Two independent runs of the KNRS algorithm on the Brainin function

Eval. count	$f(x)$	Comment		Comment	$f(x)$	Eval. count
1	37.52934	1^{st} point in SOBOL sequence		1^{st} point in SOBOL sequence	39.48397	1
2	19.60173				39.48397	2
3	19.60173				0.40519	3
4	19.60173				0.39996	4
5	19.60173				0.39996	5
6	19.60173				0.39996	6
7	19.60173				0.39996	7
8	19.59964				0.39996	8
9	19.59964				0.39996	9
10	19.59964				0.39996	10
11	19.59964				0.39996	11
12	19.59964	1^{st} NR converges to a saddle point			0.39996	12
13	16.76393	2^{nd} point in SOBOL sequence			0.39996	13
14	16.76393				0.39996	14
15	16.76393				0.39996	15
16	16.76393				0.39996	16
17	16.76393				0.39996	17
18	16.76393				0.39996	18
19	16.76393				0.39996	19
20	16.76393			NR sequences up till here converged to a minima	0.39996	20
21	16.76393			Next NR sequence converges to a different minima	0.25473	21
22	16.76393				0.24775	22
23	16.76393				0.24775	23
24	16.76393				0.24775	24
25	16.76393	NR sequences up till here, starting from the second point in the SOBOL sequence, converge to either a maxima outside bounds, or to a saddle point			0.24775	25
26	16.24129	Next point in SOBOL sequence			0.24775	26
27	16.24129				0.24775	27
28	16.24129				0.24775	28
29	16.24129				0.24775	29
30	16.24129				0.24775	30
31	16.24129				0.24775	31
32	16.24129				0.24775	32
33	16.24129				0.24775	33
34	16.24129				0.24775	34
35	16.24129				0.24775	35
36	16.24129				0.24775	36
37	16.24129				0.24775	37
38	16.24129				0.24775	38
39	16.24129				0.24775	39
40	16.24129				0.24775	40
41	16.24129				0.24775	41
42	16.24129				0.24775	42
43	16.24129				0.24775	43
44	16.24129				0.24775	44
45	16.24129				0.24775	45
46	16.24129				0.24775	46
47	16.24129				0.24775	47
48	16.24129				0.24775	48
49	16.24129				0.24775	49
50	16.24129	The NR sequences since the previous NR sequence are either converging to saddle points again or the last sequence is diverging; max. eval. reached		No subsequent NR sequence could converge to a better minima; max. eval reached	0.24775	50

the DHC, time per evaluation implies time taken for generating the search direction from a point and deciding whether to move in that direction or to decrease the step length, plus convergence checks. These times per evaluation are calculated on an average for all 50 runs and are reported in Table 10.5.

Table 10.5: Global optimisation: Comparison of time per evaluation for KNRS, GA, and DHC (50 runs for each algorithm in each test function)

Function	KNRS		GA (50×50)		DHC		Time per eval. (10^{-3})s		
	n_{eval}	time (s)	n_{eval}	time (s)	n_{eval}	time (s)	KNRS	GA	DHC
Branin	50	14.07	2500	210.56	500	81.76	5.63	1.68	3.27
De Jong's	50	6.23	2500	95.45	500	37.68	2.49	0.76	1.51
Rosenbrock's	50	15.42	2500	209.25	500	75.00	6.17	1.67	3.00
Rastrigin's	500	412.17	2500	608.96	2500	1316.33	16.49	4.87	10.53
Bump	500	594.79	2500	819.54	2500	1206.55	23.79	6.56	9.65
Ackley's	500	358.96	2500	610.72	2500	945.84	14.36	4.89	7.57
Six-hump camel back	500	277.91	2500	409.06	2500	619.23	11.12	3.27	4.95
Dropwave	500	451.73	2500	631.82	2500	912.31	18.07	5.05	7.3
Easom's	500	636.72	2500	850.36	2500	1216.11	25.47	6.8	9.73
Goldstein-Price's	500	433.26	2500	613.82	2500	936.68	17.33	4.91	7.49
Average							**14.09**	**4.05**	**6.50**

Averaged over all the runs and functions the time taken for one evaluation of the KNRS algorithm is longer than a GA and DHC by a factor or 3.5 and 2, respectively. However, for all the test functions, except Easom's and Six-hump camel back function, the implementation of KNRS, finds the global optimum in considerably fewer evaluations than a GA (see Table 10.6). Comparison with DHC yields mixed results. On average KNRS finds the global optima earlier than DHC (see Table 10.6) for the Branin function (by a factor of ≈ 4), De Jong's function (by a factor of ≈ 9), Rosenbrock's function (by a factor of ≈ 3), Bump function (by a factor of ≈ 2), and Goldstein-Price's function (by a factor of ≈ 1.5). On the other hand, DHC finds the global optima earlier than KNRS for the Six-hump camel back function (by a factor of ≈ 5). The comparison times for the Dropwave function are comparable and for the highly multimodal Rastrigin's and Ackley's functions are inconclusive as the mean for the KNRS algorithm for these functions has not converged to the global minimum value for the 500 evaluations used. The issues and possible causes for this observed behaviour, and the special Easom's function, are discussed next.

10.7.3.1 Branin, De Jong's, Rosenbrock's, Bump, and Goldstein-Price's functions

The behaviour for the De Jong's function is the simplest to explain. Being a quadratic function, the NR iterations for the original function converge in one NR evaluation. Since we are working with the Krig instead of the original function, it takes 3-4 iterations for each NR sequence, irrespective of where the iteration starts, to converge to the global minimum. This is apparent in Figure 10.24a where all runs converge to the minimum in less than five evaluations. DHC, on the other hand, takes about 30 evaluations on average to converge to the minimum, as it does not use gradient information directly.

Table 10.6: Average number of evaluations before the error between the mean of the best point (for 50 runs) for each algorithm and the global minimum reaches 2%; '-' represents a case when the algorithm, in the number of maximum evaluations specified (Table 10.5), has not converged close to the minimum

Function	Absolute no. of evaluations			No. Evaluations relative to KNRS		
	KNRS	DHC	GA	KNRS	DHC	GA
Branin	32	129	2050	1	4	64.1
De Jong's	4	35	1150	1	8.8	287.5
Rosenbrock's	44	142	-	1	3.2	-
Rastrigin's	-	112	-	-	-	-
Bump	282	476	2100	1	1.7	7.4
Ackley's	-	56	1700	-	-	-
Six-hump camel back	165	35	700	1	0.2	4.2
Dropwave	483	384	-	1	0.8	-
Easom's	-	722	-	-	-	-
Goldstein-Price's	180	255	-	1	1.4	-

The Branin function has seven stationary points, two outside the decision variable bounds and five within these bounds. The basins of attraction for the five stationary points within the decision variable bounds can be seen in Figure 10.6b. On average one NR sequence for the Krig of the Branin function used here converges in seven evaluations. Hence, according to Figure 10.23b on average, about 5 NR sequences ($\approx 35/7$) have to be evaluated, starting from anywhere in the SOBOL sequence, to get a starting point that converges to any of the three global minima. DHC, on the other hand, owing to its strategy of decreasing the step lengths sequentially to assure convergence, takes many more evaluations (≈ 150) to converge to the global optima. Similar arguments hold for the Rosenbrock's, Bump, and Goldstein-Price's function. An important observation is made for the Bump function. In the same way that there are two maxima outside the search domain of the Krig of the Branin function Krig, for the Krig of the Bump function the global minima, if the bounds are not considered, lies below the bound-constrained minima (see Figure 10.10d: arrows within the box $[x_1 \in (1.5, 2.0), x_2 \in (0, 0.4)]$ point towards this bound-unconstrained minima). The KNRS, following the NR sequence, converges to this minima. On the other hand, since DHC is constrained to search only within the search domain, it converges to the bound-constrained global minima. This explains why in the mean plots (Figure 10.27b and 10.27f) the mean for KNRS converges to a value lower than that for DHC. Nonetheless, this behaviour of the KNRS algorithm is undesirable a from constrained optimisation point of view, where the bounds must be honoured.

10.7.3.2 Six-hump camel back and Dropwave functions

The relatively poor performance of the KNRS algorithm with respect to DHC on the Six-hump camel back function is primarily because the basins of attraction of the two global minima, at (0.0898, -0.7126) and (-0.0898, 0.7126), are relatively small when compared to the surrounding stationary points (see Figure 10.12d). Consequently, more points in the SOBOL sequence are needed to locate a point in the basin of attraction of the global minima. Nonetheless, on average, the mean of the best point in 50 runs converges to the global minima in approximately 140 evaluations.

For the Dropwave function, the performance of KNRS and DHC are comparable. Similar to the Six-hump camel back function, the basin of attraction of the global minima for the Dropwave is small relative to the search domain (see Figure 10.14d). DHC struggles equally for this function due to the presence of concentric ridges which can cause DHC to oscillate between these ridges.

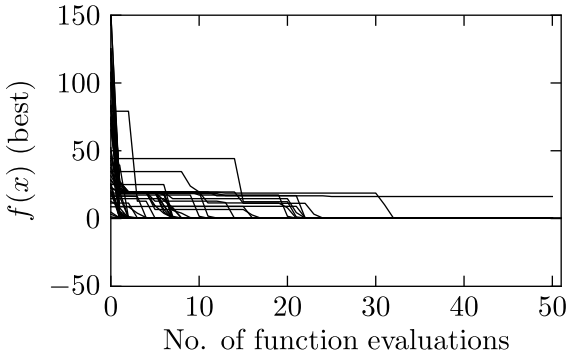
10.7.3.3 Rastrigin's & Ackley's function

The Rastrigin's and Ackley's functions are the most multimodal among the test functions used here. They both have regularly alternating minima and maxima, and saddle points between them (Figures 10.9b and 10.11b). At first glance, comparing Figure 10.26b with Figure 10.26f, and more importantly Figure 10.28b with Figure 10.28f, it may be concluded that the KNRS algorithm has performed extremely poorly when compared to DHC on these functions. This is indeed true, when the goal is to find the global optimum. However, further insight is obtained by considering that facts that (i) a limited number of evaluations of the algorithm have been performed for both the algorithms and (ii) the non-convergence of the mean to the global minimum for the KNRS algorithm on these two functions is neither because any of the starting points in the NR sequence are diverging, nor because the SOBOL sequence is performing insufficient exploration of the design space. Every NR iteration is actually converging to the closest stationary point. Indeed, it is the sheer number of such stationary points for these functions that 500 evaluations are not enough to gather entire information about the landscape, i.e. the location of all stationary points including the global minimum. This aspect, although magnified for the case of Rastrigin's and Ackley's functions, is relevant to all the functions. However, in other functions, the number of evaluations used was sufficient to explore the search region sufficient enough to find the global minima. This aspect is demonstrated in Figure 10.34 where the results of one random run of the 50 runs for all the test functions are plotted. The red dots show the locations where the various NR sequences converged during this single run. Notice how many stationary points are found and some are missed during each of these runs. Particularly notice the plots for Rastrigin's and Ackley's functions. Here, inevitably, a few stationary points are missed if only 500 evaluations are used. For some runs, these missed stationary points could as well be the global minima. However, if the algorithm is allowed to run for a higher number of evaluations, the global minima is to be found, inevitably, due to the space-filling nature of the SOBOL sequence.

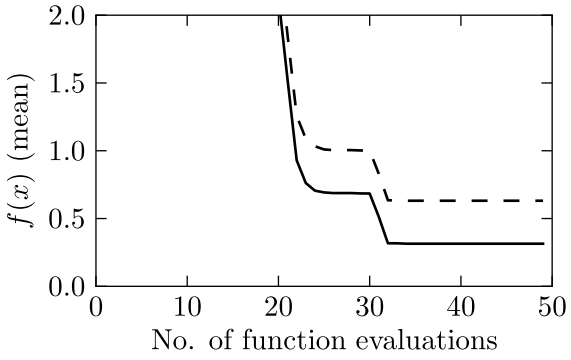
Encouragingly, every NR sequence of the KNRS algorithm, unless diverging (which is not the case for the Rastrigin's and Ackley's function), results in yielding some information about the landscape of the function. Hence, it is the author's hypothesis that the merit of the KNRS algorithm, although effective in a global optimisation problem, lies in multimodal optimisation, where the goal is not just to find the global minima but other optima as well irrespective of their relative function values. This aspect is explored further in section 10.8.

10.7.4 Easom's function

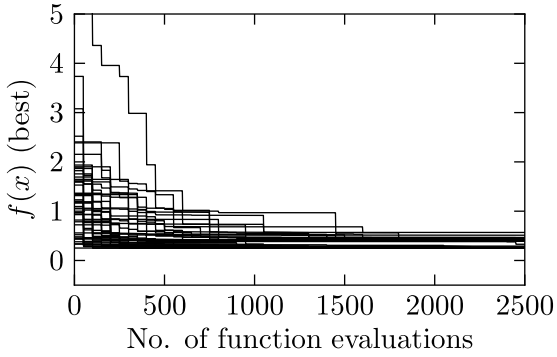
As mentioned earlier Easom's function is an extremely challenging function for optimisation algorithms because: a) the size of the basin of attraction of the global minimum is very small when compared to the search domain; and b) the function elsewhere is relatively flat, with a function value close to zero. In Figures 10.31a and 10.31b, 23 out of 50 runs have not found



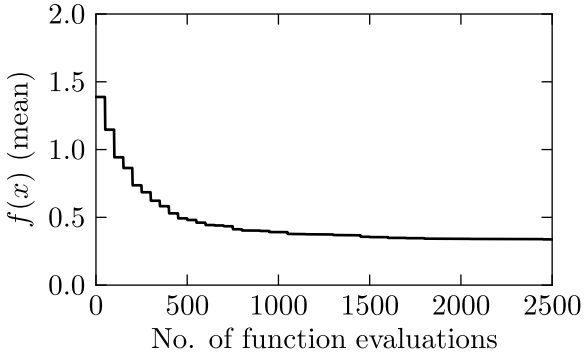
(a) Optimisation history KNRS : best point for 50 runs



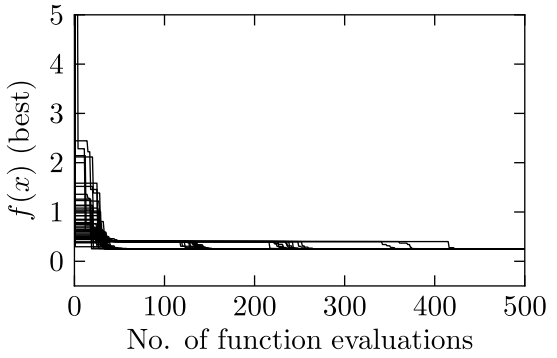
(b) Optimisation history KNRS: mean of the best point for 50 runs



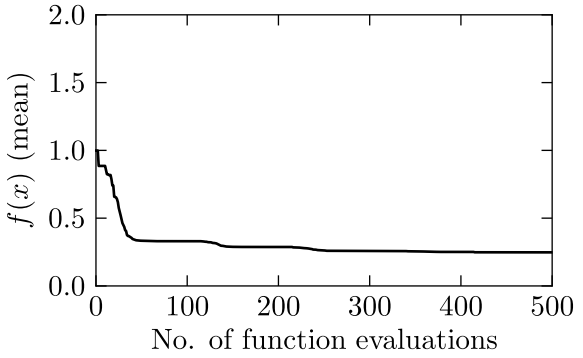
(c) Optimisation history GA : best point for 50 runs



(d) Optimisation history GA: mean of the best point for 50 runs



(e) Optimisation history DHC : best point for 50 runs

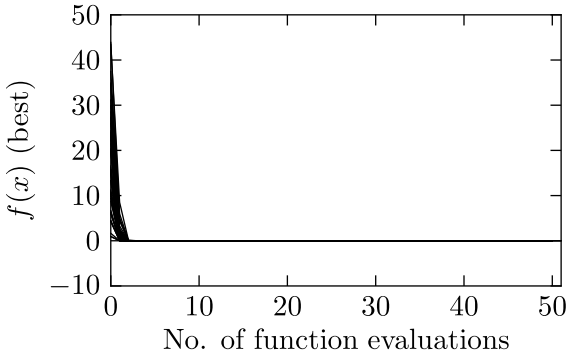


(f) Optimisation history DHC: mean of the best point for 50 runs

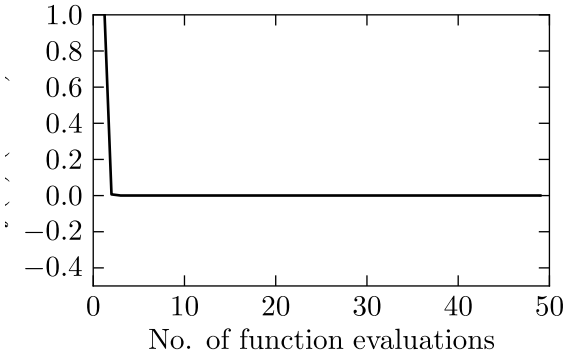
Figure 10.23: KNRS, GA, and DHC, algorithms on the Branin function: optimisation history for best point and mean of the best point for sample of 50 runs

the global minima. One of the reasons for the relatively better performance of DHC for this function could be the outer loop of DHC. In order to seed the inner loop, 50 random points are evaluated and the inner loop is seeded with the best point among these 50 points. Clearly, if any of these points lies in the basin of the global minima, DHC's inner loop will quickly converge to it. Nonetheless, if KNRS is allowed higher maximum evaluations, then the evenly exploratory nature of the SOBOL sequence should find the minima. This indeed happens, as shown in Figure 10.33.

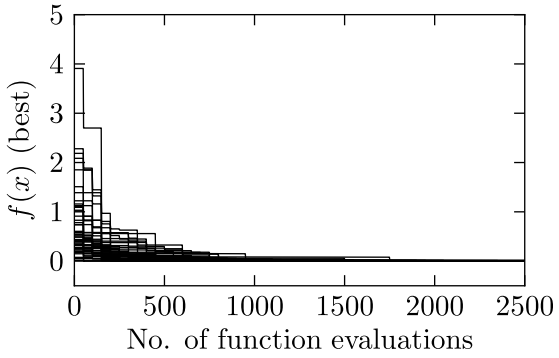
Lastly, as can be observed from Figures 10.23–10.32, a standard GA for the chosen test functions, performs relatively poorly when compared to KNRS and DHC. In the next section the multimodal optimisation capabilities of KNRS are explored in detail.



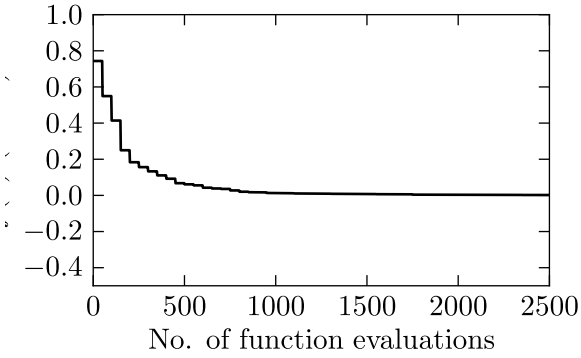
(a) Optimisation history KNRS: best point for 50 runs



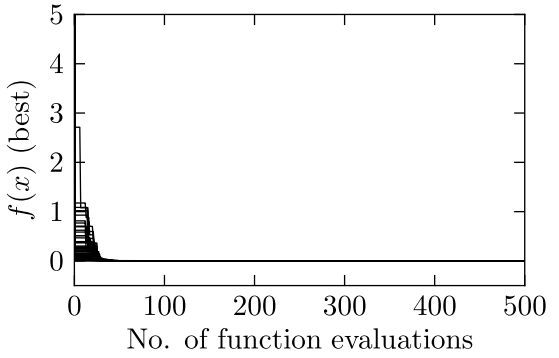
(b) Optimisation history KNRS: mean of the best point for 50 runs



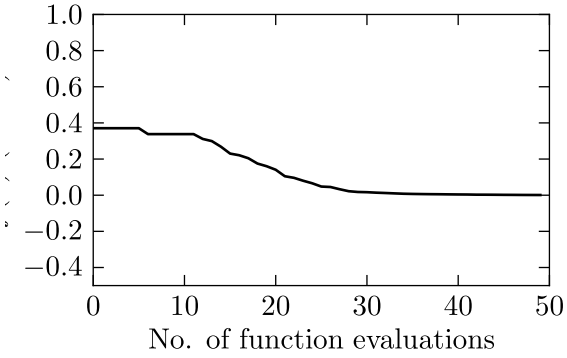
(c) Optimisation history GA : best point for 50 runs



(d) Optimisation history GA: mean of the best point for 50 runs

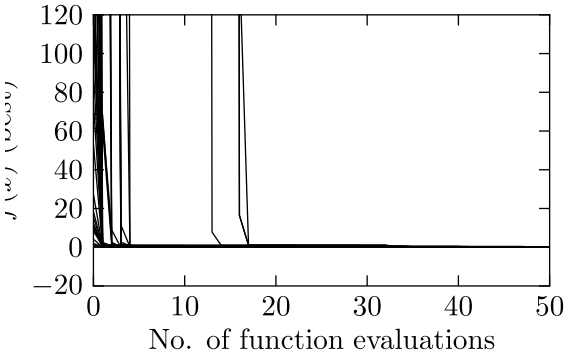


(e) Optimisation history DHC : best point for 50 runs

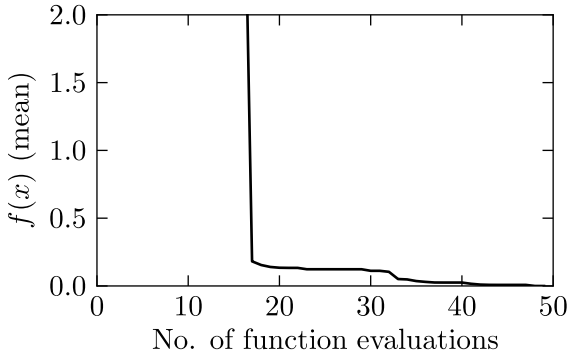


(f) Optimisation history DHC: mean of the best point for 50 runs

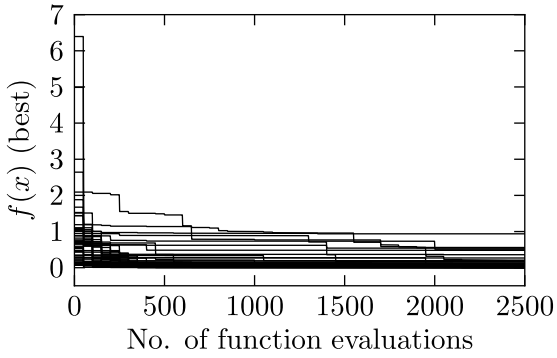
Figure 10.24: KNRS, GA, and DHC, algorithms on the De Jong's function: optimisation history for best point and mean of the best point for sample of 50 runs



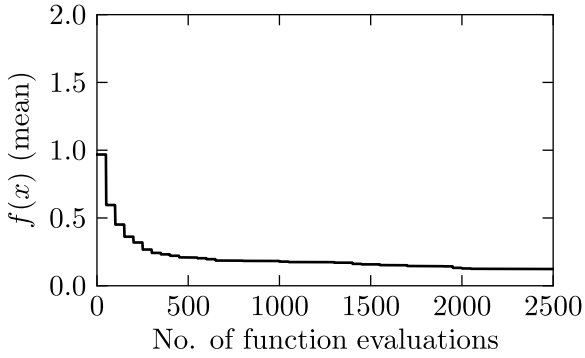
(a) Optimisation history KNRS: best point for 50 runs



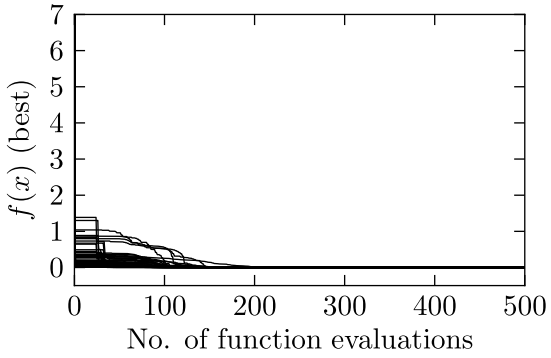
(b) Optimisation history KNRS : mean of the best point for 50 runs



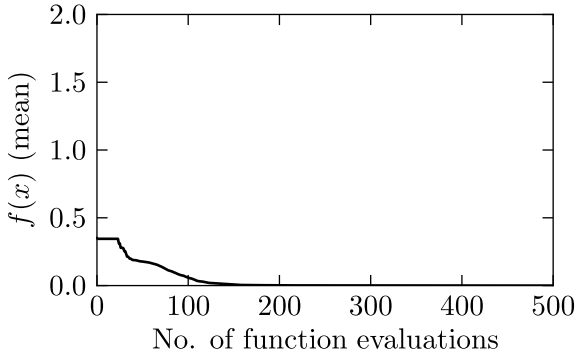
(c) Optimisation history GA : best point for 50 runs



(d) Optimisation history GA: mean of the best point for 50 runs

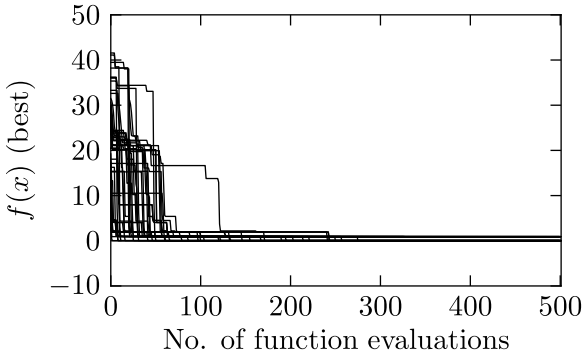


(e) Optimisation history DHC : best point for 50 runs

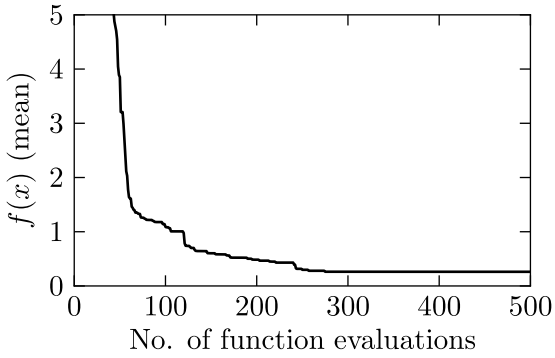


(f) Optimisation history DHC: mean of the best point for 50 runs

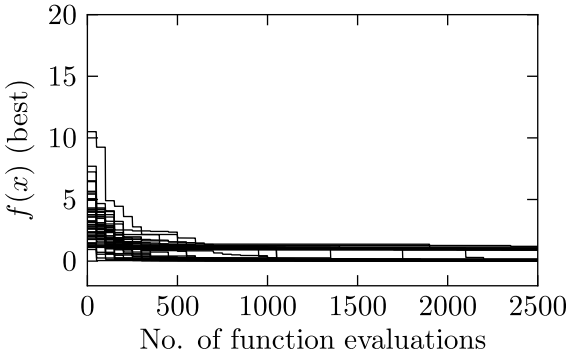
Figure 10.25: KNRS, GA, and DHC, algorithms on the Rosenbrock's function: optimisation history for best point and mean of the best point for sample of 50 runs



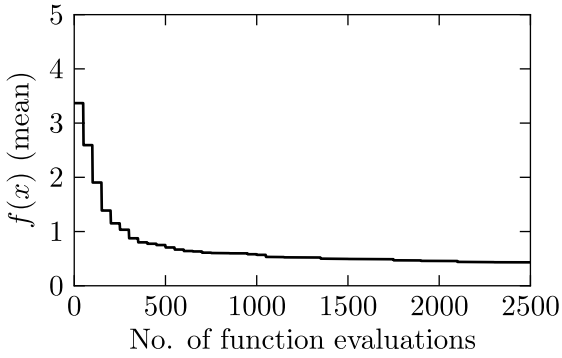
(a) Optimisation history KNRS: best point for 50 runs



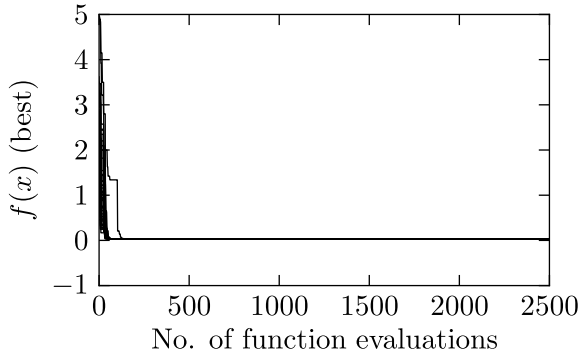
(b) Optimisation history KNRS: mean of the best point for 50 runs



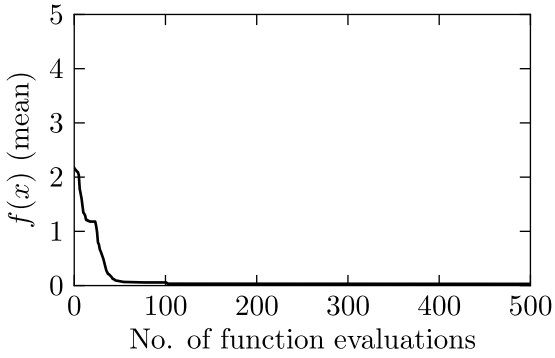
(c) Optimisation history KNRS : best point for 50 runs



(d) Optimisation history GA: mean of the best point for 50 runs

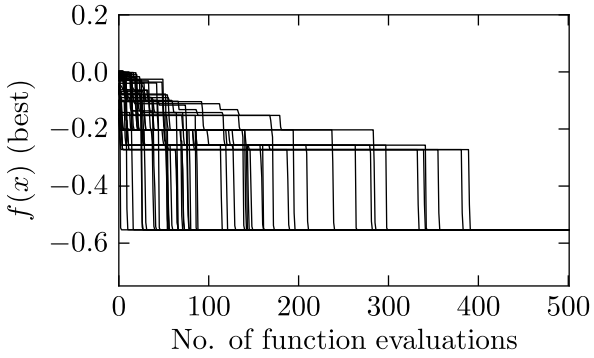


(e) Optimisation history DHC : best point for 50 runs

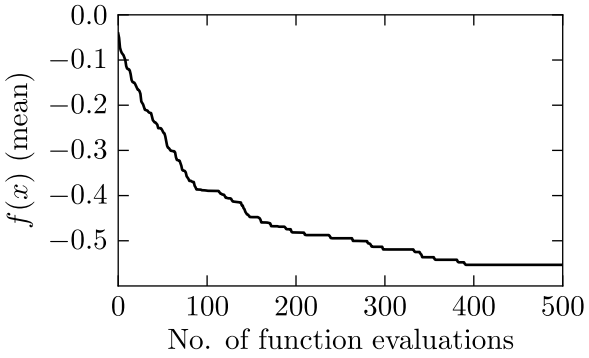


(f) Optimisation history DHC: mean of the best point for 50 runs

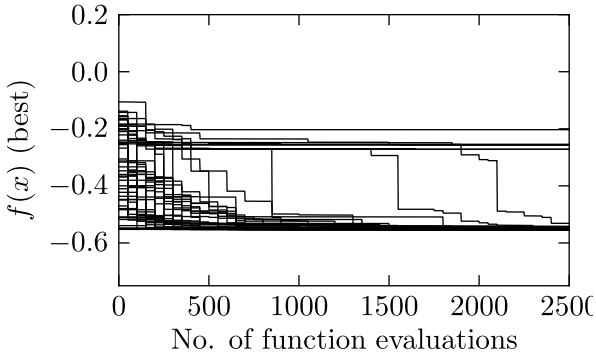
Figure 10.26: KNRS, GA, and DHC, algorithms on the Rastrigin's function: optimisation history for best point and mean of the best point for sample of 50 runs



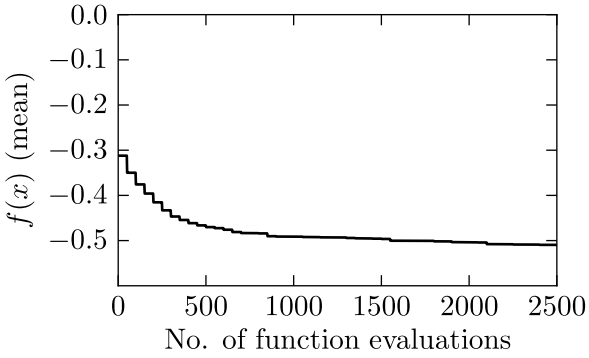
(a) Optimisation history KNRS: best point for 50 runs



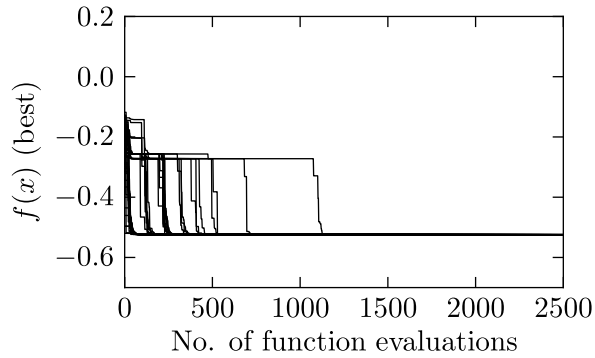
(b) Optimisation history KNRS: mean of the best point for 50 runs



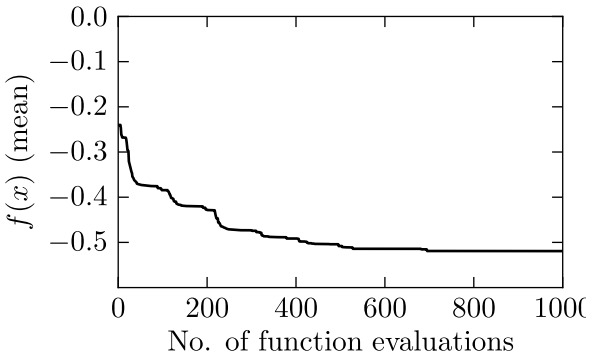
(c) Optimisation history KNRS : best point for 50 runs



(d) Optimisation history GA: mean of the best point for 50 runs

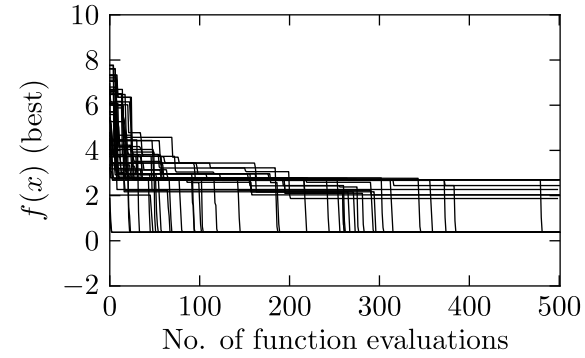


(e) Optimisation history DHC : best point for 50 runs

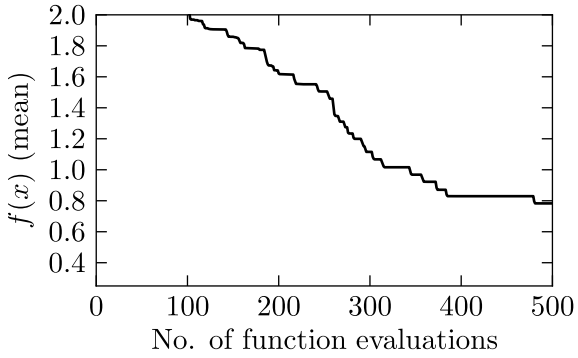


(f) Optimisation history DHC: mean of the best point for 50 runs

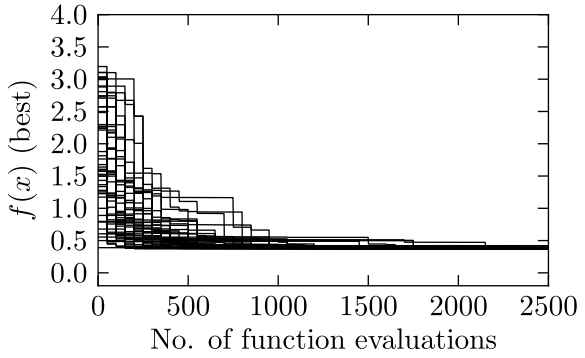
Figure 10.27: KNRS, GA, and DHC, algorithms on the Bump function: optimisation history for best point and mean of the best point for sample of 50 runs



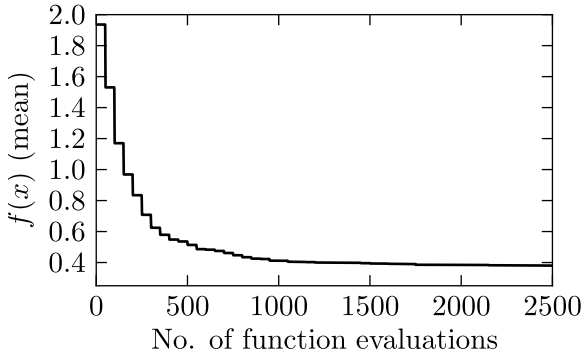
(a) Optimisation history KNRS: best point for 50 runs



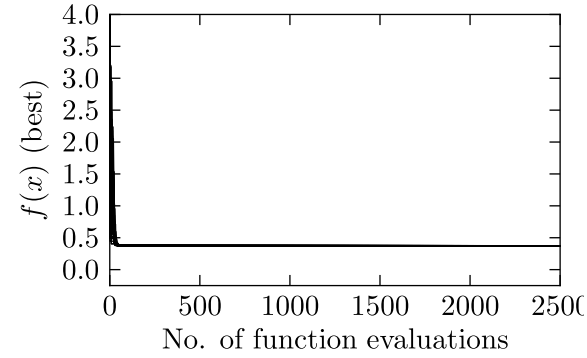
(b) Optimisation history KNRS: mean of the best point for 50 runs



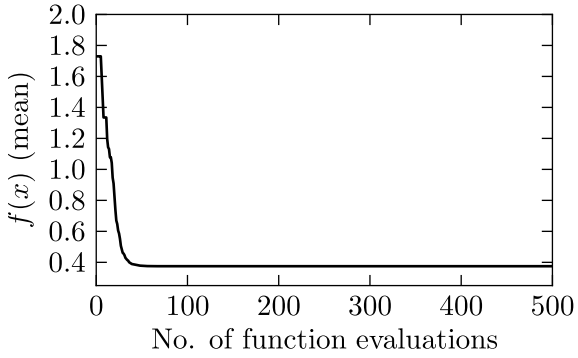
(c) Optimisation history KNRS : best point for 50 runs



(d) Optimisation history GA: mean of the best point for 50 runs

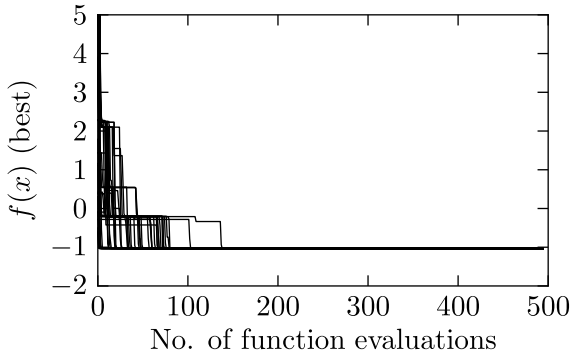


(e) Optimisation history DHC : best point for 50 runs

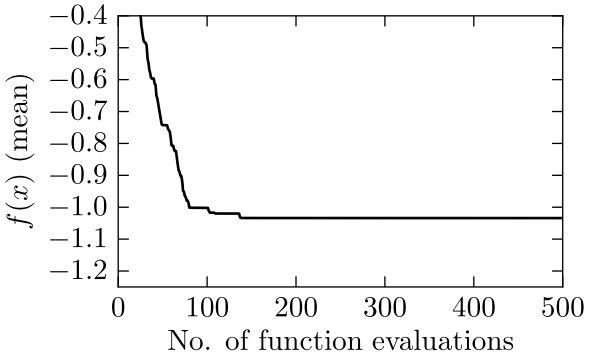


(f) Optimisation history DHC: mean of the best point for 50 runs

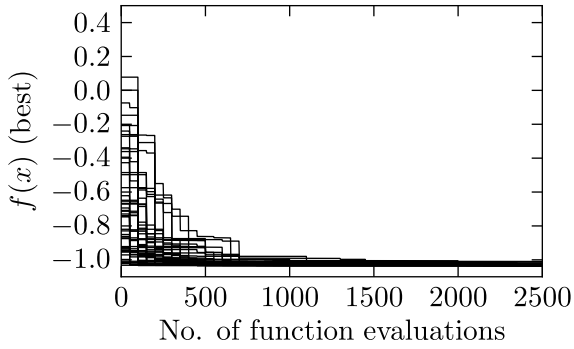
Figure 10.28: KNRS, GA, and DHC, algorithms on the Ackley's function: optimisation history for best point and mean of the best point for sample of 50 runs



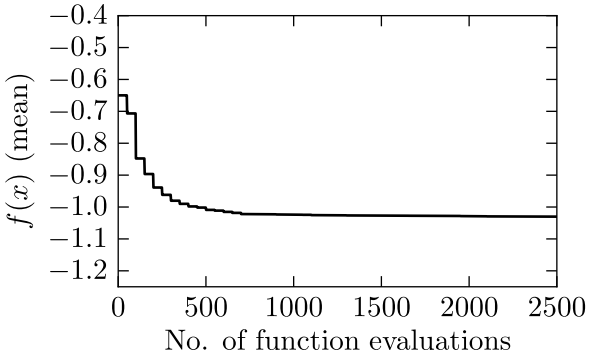
(a) Optimisation history KNRS: best point for 50 runs



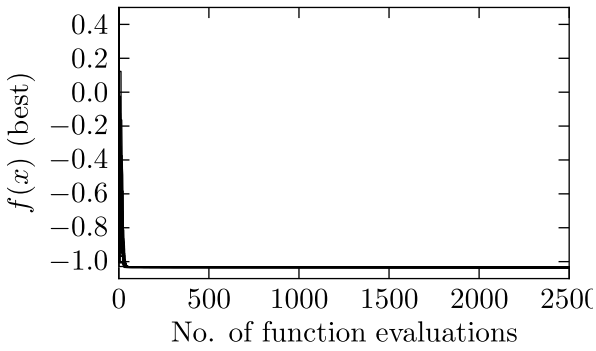
(b) Optimisation history KNRS: mean of the best point for 50 runs



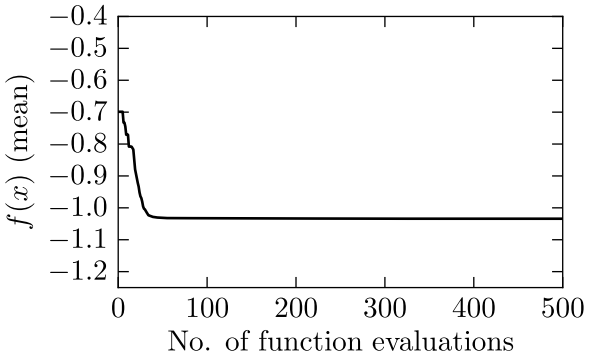
(c) Optimisation history KNRS : best point for 50 runs



(d) Optimisation history GA: mean of the best point for 50 runs

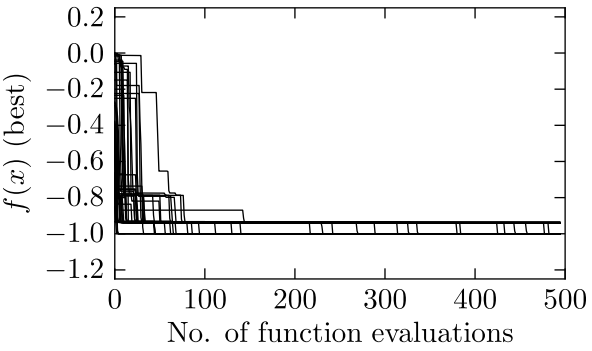


(e) Optimisation history DHC : best point for 50 runs

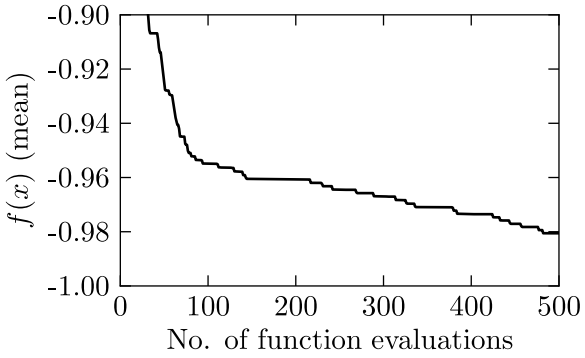


(f) Optimisation history DHC: mean of the best point for 50 runs

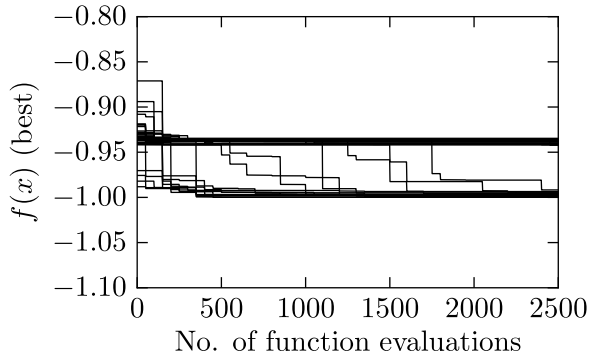
Figure 10.29: KNRS, GA, and DHC, algorithms on the Six-hump camel back function: optimisation history for best point and mean of the best point for sample of 50 runs



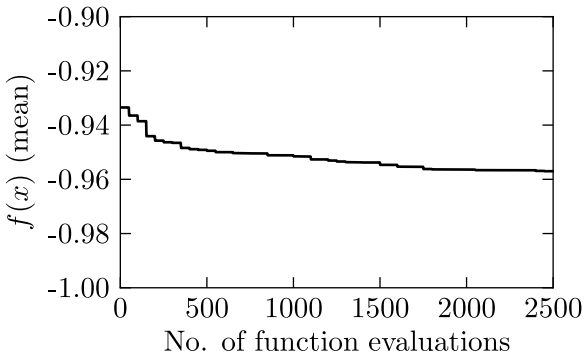
(a) Optimisation history KNRS: best point for 50 runs



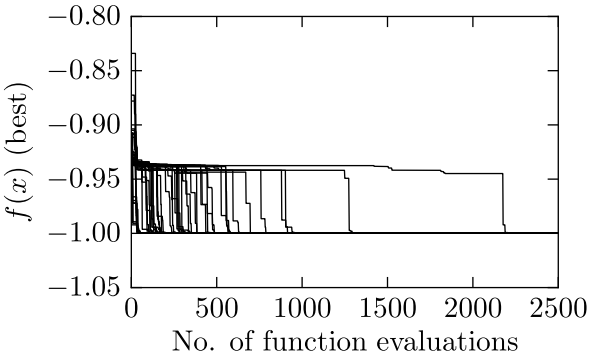
(b) Optimisation history KNRS: mean of the best point for 50 runs



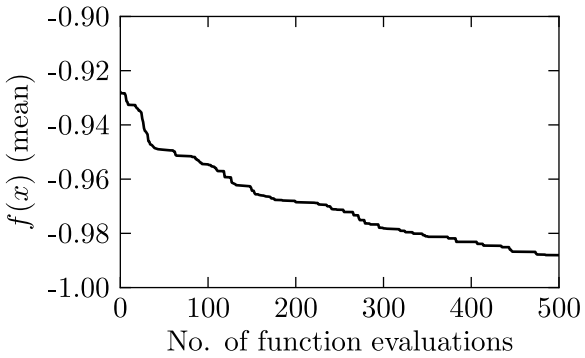
(c) Optimisation history KNRS : best point for 50 runs



(d) Optimisation history GA: mean of the best point for 50 runs

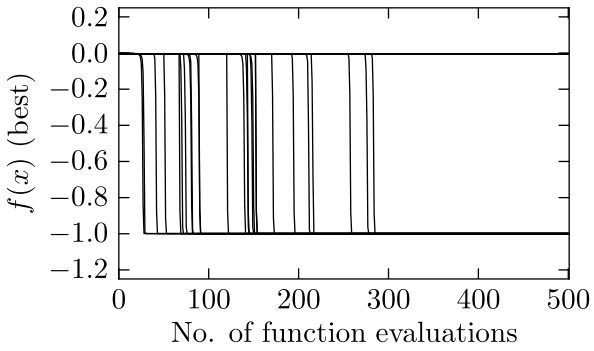


(e) Optimisation history DHC : best point for 50 runs

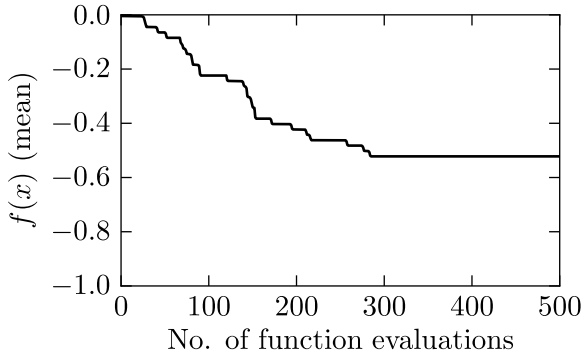


(f) Optimisation history DHC: mean of the best point for 50 runs

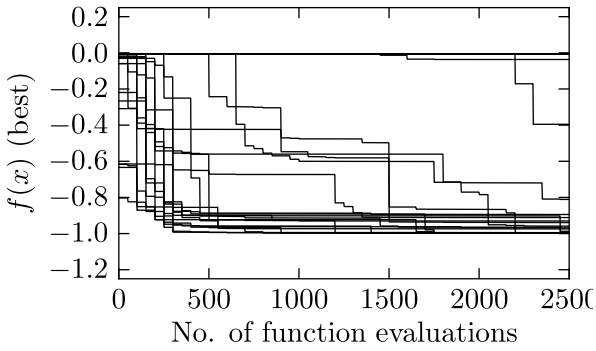
Figure 10.30: KNRS, GA, and DHC, algorithms on the Dropwave function: optimisation history for best point and mean of the best point for sample of 50 runs



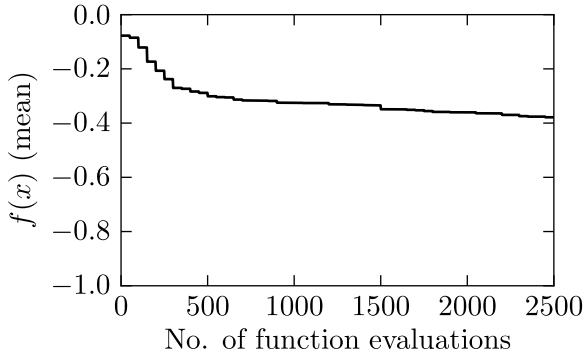
(a) Optimisation history K: best point for 50 runs



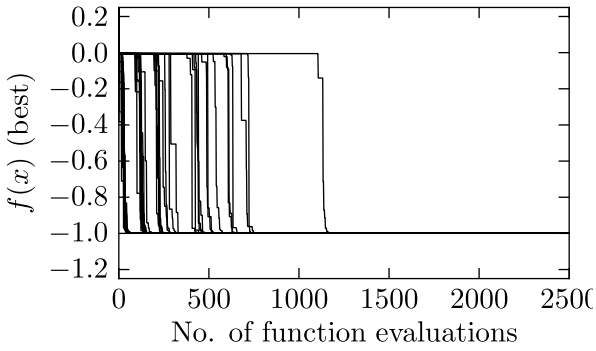
(b) Optimisation history KNRS: mean of the best point for 50 runs



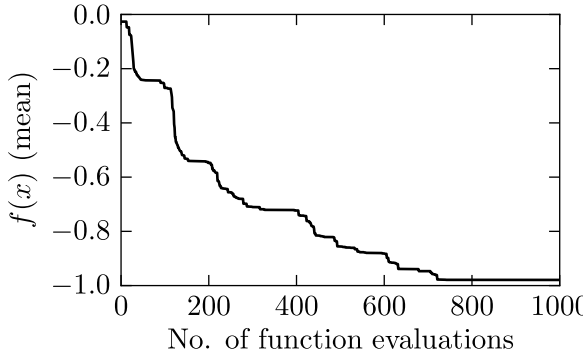
(c) Optimisation history GA : best point for 50 runs



(d) Optimisation history GA: mean of the best point for 50 runs

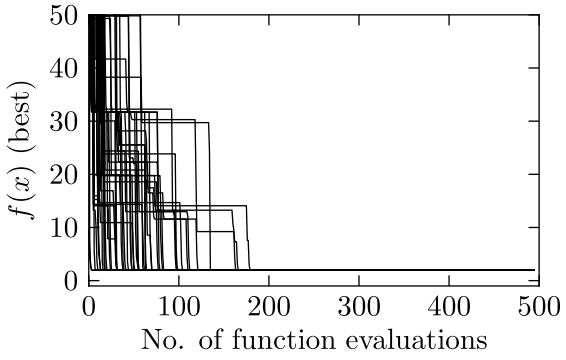


(e) Optimisation history DHC : best point for 50 runs

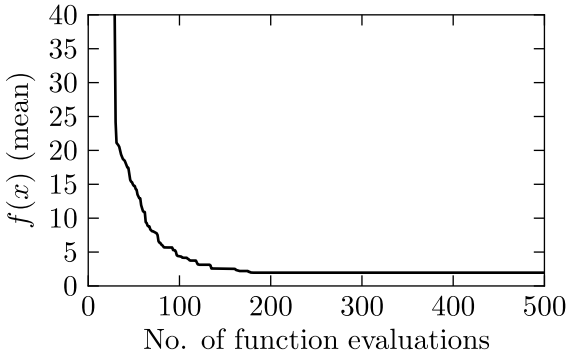


(f) Optimisation history DHC: mean of the best point for 50 runs

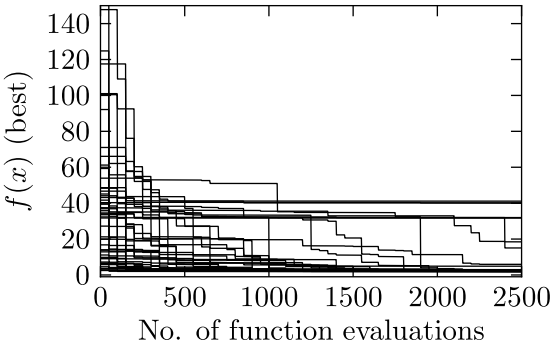
Figure 10.31: KNRS, GA, and DHC, algorithms on the Easom's function: optimisation history for best point and mean of the best point for sample of 50 runs



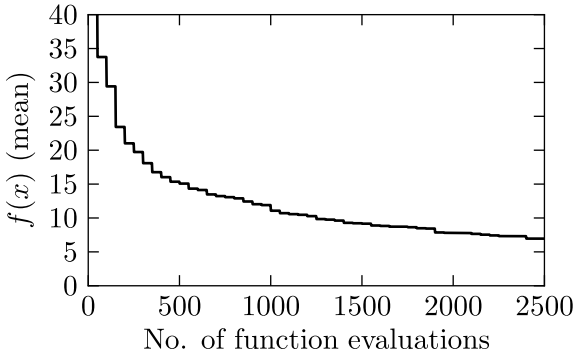
(a) Optimisation history KNRS: best point for 50 runs



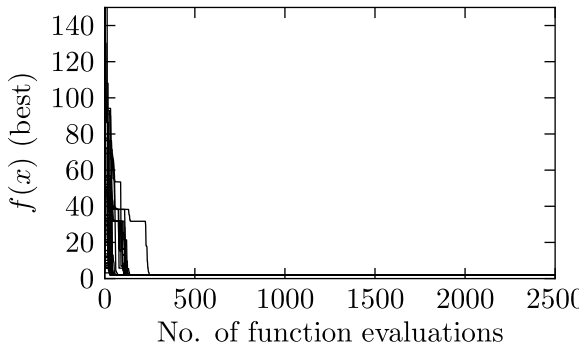
(b) Optimisation history KNRS: mean of the best point for 50 runs



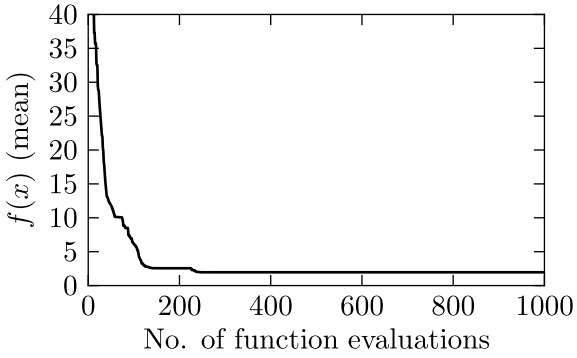
(c) Optimisation history KNRS : best point for 50 runs



(d) Optimisation history GA: mean of the best point for 50 runs

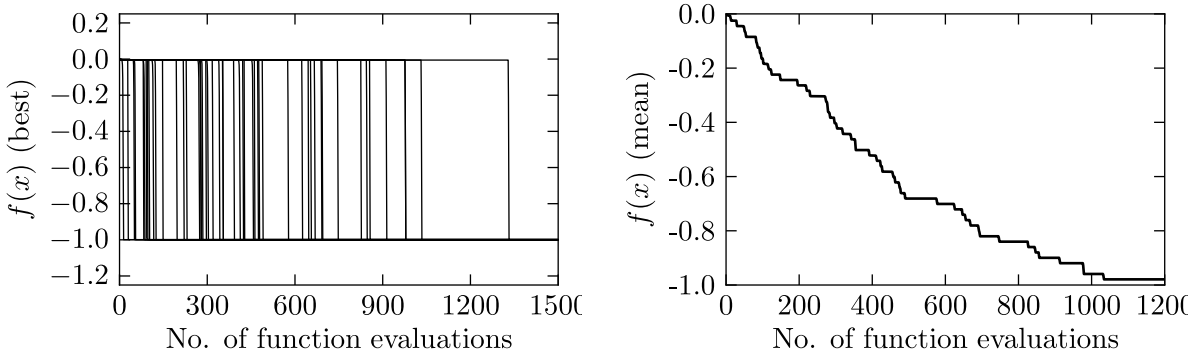


(e) Optimisation history DHC : best point for 50 runs



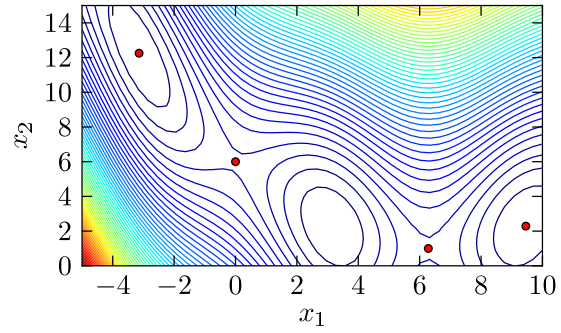
(f) Optimisation history DHC: mean of the best point for 50 runs

Figure 10.32: KNRS, GA, and DHC, algorithms on the Goldstein-Price's function: optimisation history for best point and mean of the best point for sample of 50 runs

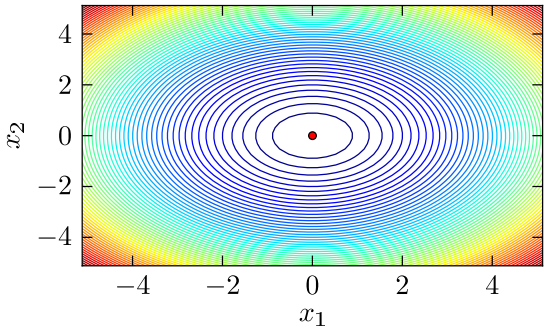


(a) Optimisation history KNRS: best point for 50 runs (b) Optimisation history KNRS: mean of the best point for 50 runs

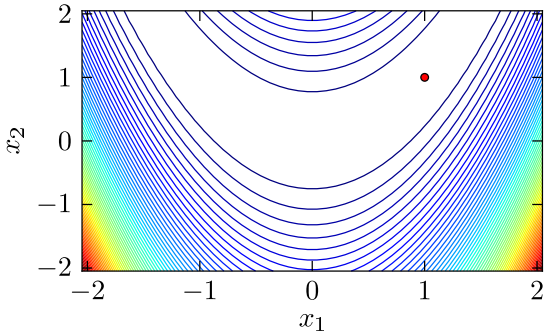
Figure 10.33: KNRS on the Easom's function: higher maximum evaluations allowed



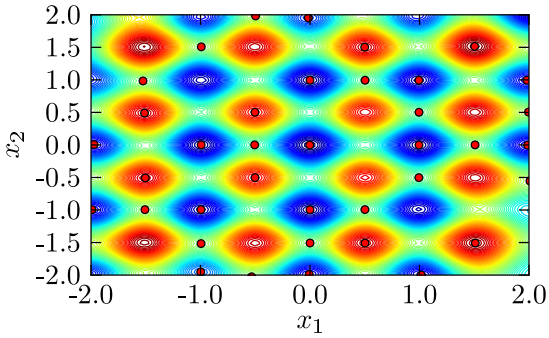
(a) Branin function



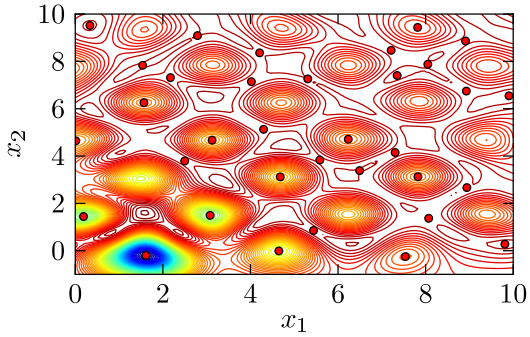
(b) De Jong's function



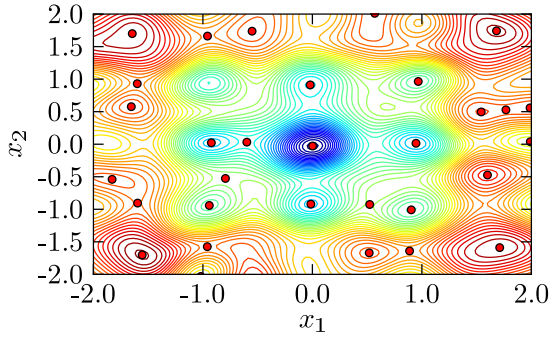
(c) Rosenbrock's function



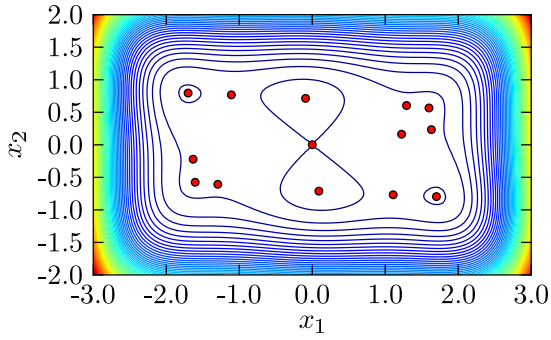
(d) Rastrigin's function



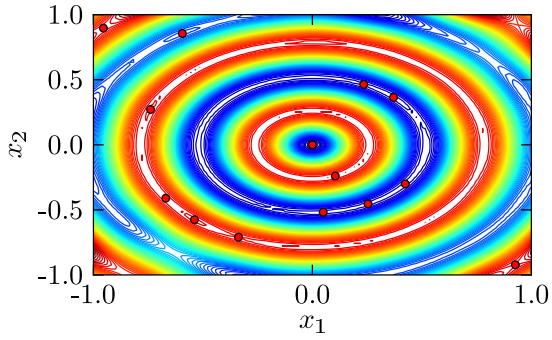
(e) Bump function



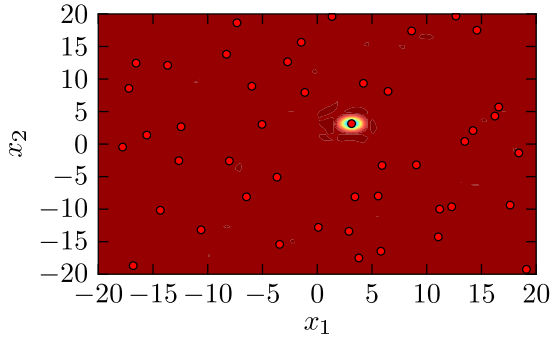
(f) Ackley's function



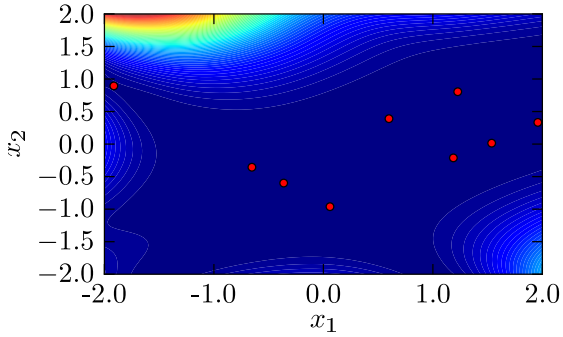
(g) Six-hump camel back function



(h) Dropwave function



(i) Easom's function



(j) Goldstein-Price's function

Figure 10.34: Random KNRS run on the test functions: optimal solutions found in red circles

10.8 Comparison of KNRS with other algorithms for multimodal optimisation

Multimodal optimisation is applicable to functions that are multimodal, i.e. such functions have multiple optimal solutions: either more than one global minima, or a global minimum with several local minima, or a combination of both. The objective of a multimodal optimisation algorithm is to find all the minima (global and local) [15]. The following section outlines the basic concepts of evolutionary algorithms proposed for multimodal optimisation.

10.8.1 Evolutionary algorithms for multimodal optimisation

Since evolutionary algorithms work with a population rather than a point-by-point search, they are thought to be at an inherent advantage at finding more than one optimal solution. However, even in the presence of multiple optima with the same fitness value, a standard GA eventually converges to a single optimum solution. De Jong [209] describes this as *generic drift*. Hence, in order to avoid convergence of a GA to a single peak and hence find multiple solutions, sufficient diversity in the population must be maintained [210, 15]. In GA literature each peak is considered to be a *niche*, that can only support a certain number of population members. The methods used to maintain population diversity so that members can be sustained at different peaks are called *niche formation* methods [210]. Many methods for niche formation have been proposed in the past. The most prominent of such approaches are the *crowding* approach [209, 211, 212] and *sharing* approach [213, 214, 215, 216, 217]. The central idea behind niche formation strategies is to modify the genetic operators, usually the selection operator, in a standard GA. Crowding schemes discourage crowding of population members by employing a strategy that when offsprings are created they replace an individual from the population based on a similarity metric. Thus a similar offspring replaces a like parent, thereby preserving diversity. On the other hand sharing approaches work by degrading the fitness of similar solutions, thereby penalising the fitness of members based on the number of similar individuals in the population. Such a GA is known as a fitness sharing GA and is described next. Lastly, it should be mentioned that recently in 2012, Deb and Saha [218] have proposed a bi-objective algorithm for multimodal optimisation. This algorithm converts a multimodal optimisation into a bi-objective problem and proposes to solve this problem using multiobjective evolutionary algorithms. The discussion of this algorithm and its performance comparison with the KNRS algorithm is out of scope of this thesis.

10.8.1.1 Fitness sharing Genetic algorithms

As mentioned earlier a fitness sharing GA works by degrading the fitness of members based on a similarity metric. If there are q peaks and the objective function value at each peak is f_i, then the basic idea is that if the fitness of each member belonging to a peak is assigned to be its objective function value divided by the number of members on the peak, then the optimal distribution of points, m_i, on a peak should follow the following

$$\frac{f_1}{m_1} = \frac{f_2}{m_2} = ... = \frac{f_q}{m_q} \quad \text{and} \quad \sum_{i=1}^{q} = N, \tag{10.85}$$

where N is the population size. The above is optimal in the sense that if peak i recieves any more members than m_i, as calculated above, then the fitness of all the individuals at that

peak will become less than the optimal ratio stated in the above equation, thereby decreasing the population count at that peak in subsequent generations. Similarly if peak i has less members than m_i, as calculate above, then the fitness of the members at that peak will be higher than others, thereby encouraging selection of more copies of the members at that peak in subsequent generations. In order to clasify members belonging to each peak (niche), Goldberg and Richardson [213] proposed the following sharing function

$$\mathrm{Sh}(d) = \begin{cases} 1 - \left(\frac{d}{\sigma_{share}}\right)^{\alpha} & d \le \sigma_{share} \\ 0 & \text{otherwise} \end{cases}, \tag{10.86}$$

where d is the distance (phenotype distance for real-coded GAs) between two members, and α controls the shape of the sharing function. Thus, two members share fitness only if the distance between them is less than a pre-specified sharing parameter σ_{share}, implying that solutions are close (similar) to each other. The fitness of a member is then defined by its objective function value divided by its niche count, nc_i, defined as

$$\mathrm{nc}_i = \sum_{i=1}^{N} \mathrm{Sh}(d_{ij}), \tag{10.87}$$

where the summation is carried out over the entire population size, N. Thus the fitness value becomes:

$$\mathrm{fitness}_i = \frac{f_i}{\mathrm{nc}_i}, \tag{10.88}$$

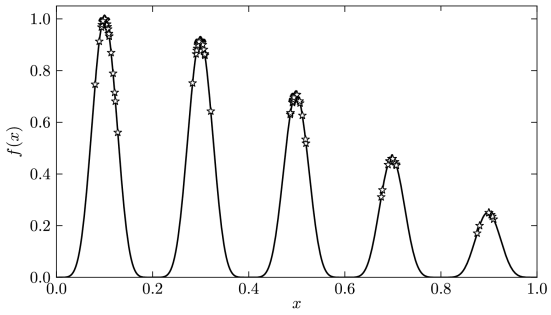
where f_i is the objective function value. In essence, the parameter σ_{share} denotes the radius of sharing effect: a member's fitness is degraded by all the members within the radius σ_{share} depending on the distance (higher distance means less sharing; this is dictated by the sharing function $\mathrm{Sh}(d)$).

The two parameters in a fitness sharing GA are α and σ_{share}. Deb [219] showed that the parameter α does not have too much effect on the performance of fitness sharing GAs. On the other hand, the parameter σ_{share} is important as it determines the niche size of each peak. Based on a guess for the number of peaks in a function, say q, Deb and Goldberg [214] suggested the following to calculate the value of σ_{share} assuming that the q minima are divided equally in the n-dimensional search space

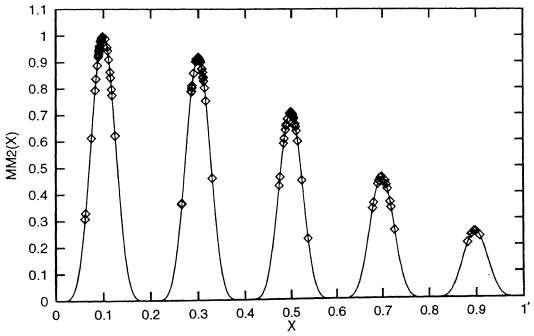
$$\sigma_{share} = \frac{\sqrt{\sum_{i=1}^{n} \left(x_i^U - x_i^L\right)^2}}{2\sqrt[n]{q}}, \tag{10.89}$$

where x_i^U and x_i^L represent the upper and lower bounds on the i_{th} decision variable, respectively. Using the above method, σ_{share} can be calculated if the number of optima is known or estimated. Thus, in a fitness sharing GA, only the selection operator needs modification, without any changes to the crossover and mutation operators. Oei et. al. [220] showed that the standard binary tournament resulted in chaotic fluctuations in a fitness sharing GA as shared fitness values for a population are calculated based on the previous generation. They suggested that a continuously updated tournament strategy, where the shared fitness values are calculated from the population of a new generation, as it is created, reduced the chaotic fluctuations in the number of members at each peak.

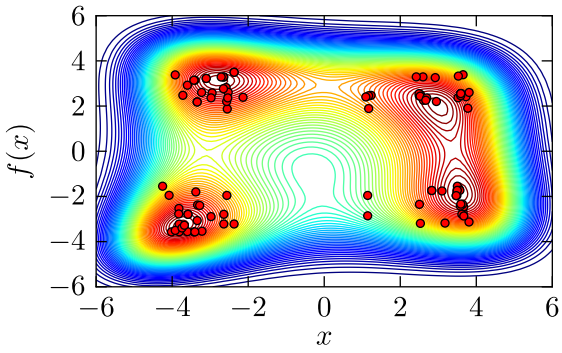
The above stated fitness sharing GA, using the continuously updated binary tournament strategy, is coded in the Python programming environment [207] by the author. The results



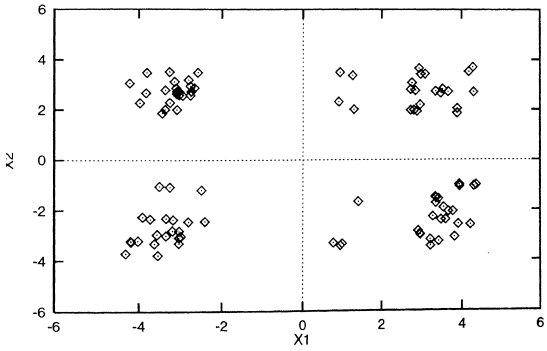
(a) Author's fitness sharing GA on MM2 function



(b) Result of fitness sharing GA on MM2 function by Kumar [14] and Deb & Kumar [15]



(c) Author's fitness sharing GA on Himmelblau's function



(d) Result of fitness sharing GA on the Himmelblau's function by Kumar [14] and Deb & Kumar [15]

Figure 10.35: Comparison of Author's fitness sharing GA implementation with that of [14, 15]: results of fitness sharing GA with 100 members after 200 generations ($\eta_c = 35$)

of this are validated against the results of the implementation by Deb and Kumar [15] and Kumar [14] for the MM2 and MM5 functions. The reader is referred to these studies for the details of MM2 function. The MM5 function, the modified Himmelblau's function, is described as follows. This function is described as this is one of the two dimensional test problem used for comparison in this section.

$$f(x_1, x_2) = \left[\frac{1 - \left((x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2\right)}{2186}\right] \quad -6 \leq x_1, x_2 \leq 6. \tag{10.90}$$

Figure 10.35 shows the results of the author's fitness sharing GA and Deb and Kumar's implementation [15, 14] of such a GA for the functions MM2 and the Himmelblau's function. The results are presented at the end of 200 generations with 100 members per generation, using the SBX crossover parameter, $\eta_c = 35$, and a good correspondence between the results of Kumar [14] and the implementation used in this chapter is found.

The parameter settings used for comparison of a fitness sharing GA with KNRS and DHC are tabulated in Table 10.7.

10.8.2 Dynamic Hill Climbing for multimodal optimisation

As mentioned earlier in section 10.7.2, Yuret and Maza proposed different ways to perform exploration through the outer loop of DHC. In the previous section, for comparing the global optimisation capabilities of DHC, the outer loop implementation where a set of random points are chosen, evaluated for fitness, and the best point chosen as the starting seed for the inner

Table 10.7: Sharing GA settings for comparion of KNRS for multimodal optimisation

Parameter	Value
Coding	Real-coded
population size	100 (except for Rastrigin's function, where population size is 200)
number of generations	200
Selection operator	Continuously updated binary tournament [220]
Crossover operator	Simulated Binary Crossover (SBX)
SBX parameter (η_c)	35.0 [15]
Crossover probability	0.9
Mutation probability	0
Sharing function	Equation 10.86; $\alpha = 1$
Sharing parameter σ_{share}	Equation 10.89

loop was used. However, for multimodal optimisation, where the goal is to find all optima, the author's opinion is that the hyper-rectangle method (see [13] for details), similar to the binary search algorithm by Hughes [208] is preferable as it promotes exploration. Such an implementation is used for all the comparisons in this section. It is worth mentioning at this stage that although DHC is not designed for multimodal optimisation. It is designed to store the local minima locations in memory to guide the search towards the global minimum.

10.8.3 Comparison of KNRS with fitness sharing GA and DHC

A set of five functions are chosen for the comparison of KNRS with a fitness sharing GA and DHC, in the context of multimodal optimisation. The functions are:

1. Branin function (Equation 10.59)
2. Rosenbrock's function (Equation 10.77)
3. Himmelblau's function (Equation 10.90)
4. Six-hump camel back function (Equation 10.81)
5. Rastrigin's function (Equation 10.78)

For all these functions 10 runs are performed and the times per evaluation are calculated for these. These times are reported in Table 10.8. The relative times between the KNRS and DHC remain similar as in the previous section. However, the time per evaluation for a fitness sharing GA is significantly larger than the standard GA. This is not surprising, as a fitness sharing GA takes a significant amount of time in calculating the shared fitness for each member of the population as it has to loop over all the members of the population being generated and calculate the distances between these points.

Since the goal here is to find multiple optima, it is not clear if the best point history and mean of the best point over evaluations are of any significance. Instead, for KNRS and DHC, the number of stationary points found and the number of minima (local or global) found, respectively, are plotted against evaluation count. These are shown in Figure 10.36. Moreover, for a GA, it is not clear how such a count can be made as a function of generations. Hence, for multimodal optimisation, the comparisons with GA are more based on the abilities of the algorithms to find the minima as opposed to how fast this task is accomplished.

Table 10.8: Multimodal optimisation: comparison of time per evaluation for KNRS, GA, and DHC (10 runs for each algorithm for each test function)

Function	KNRS		GA (50×50)		DHC		Time per eval.(10^{-3}s)		
	n_{eval}	time (s)	n_{eval}	time (s)	n_{eval}	time (s)	KNRS	GA	DHC
Branin	500	22.5	10,000	2450.8	2500	56.9	4.5	24.51	2.28
Rosenbrock's	500	22.6	10,000	2442.7	10,000	217.7	4.52	24.43	2.18
Himmelblau	2500	228.5	10,000	3331.4	10,000	116.0	9.14	33.31	1.16
Six-hump camel back	2500	228.5	10,000	3354.4	10,000	480.0	9.14	33.54	4.80
Rastrigin's	2500	357.0	10,000	4229.1	10,000	734.9	14.28	42.29	7.35
Average							**8.32**	**31.62**	**3.55**

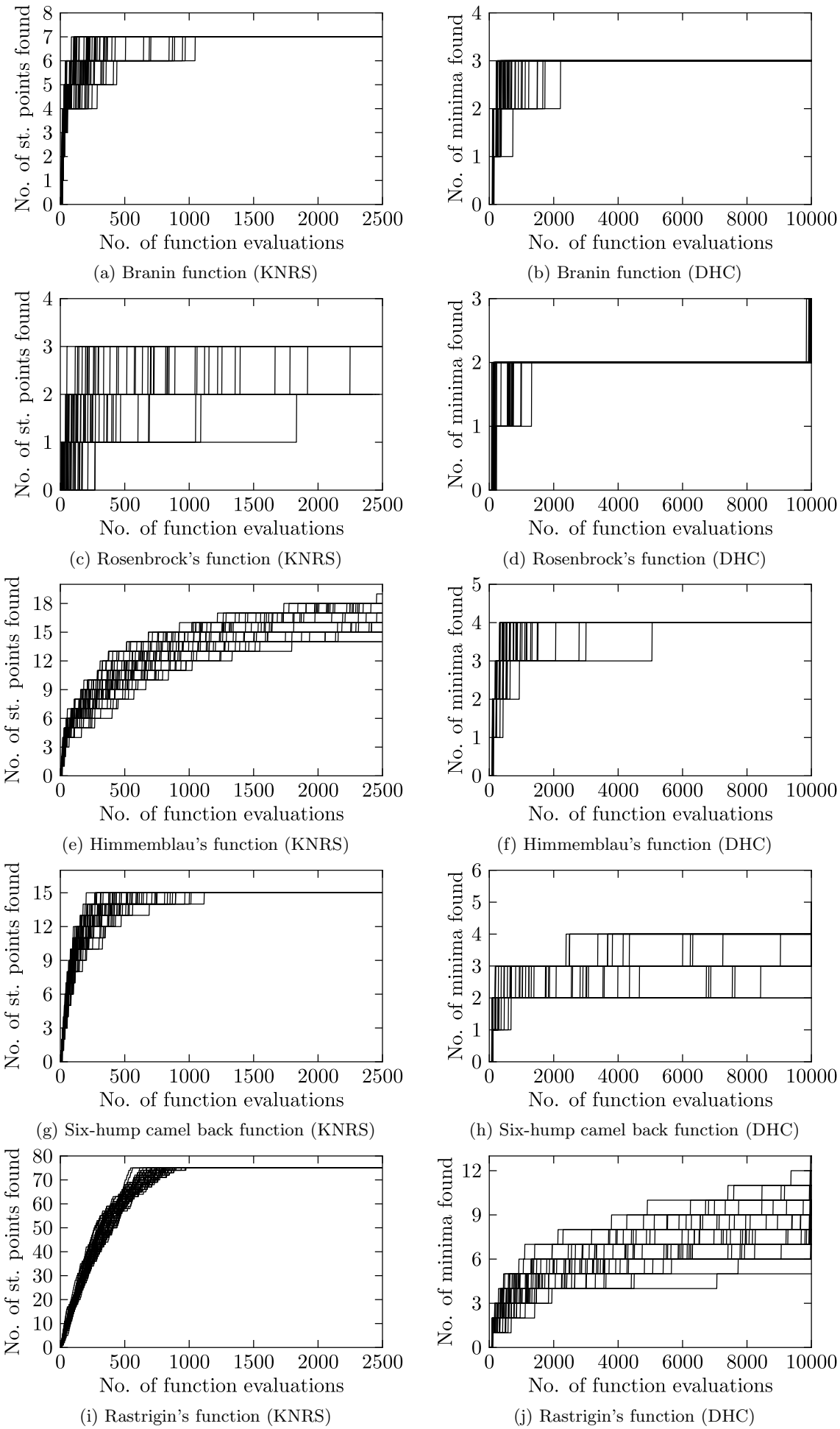
As can be seen in Figure 10.36 the KNRS algorithm finds all the stationary points, and hence all the minima, for all the five test functions in 2500 evaluations. In the last section the issue with Rastrigin's and Ackley's functions while performing global optimisation was discussed and it was said that given a higher evaluation count the algorithms would find the global minima. It can be seen in Figure 10.36i, that this has indeed happened for a maximum evaluation count of 2500 as oppposed to that of 500 in the previous section. For this function, it is clear that for all the 50 runs, 81 stationary points (which includes all the minima) have been found in just over 1000 evaluations. Moreover, DHC has only found a maximum of 12 minima (from a total of 25 minima) in 10,000 evaluations. Similar performance is true for all the test functions used when comparing KNRS with DHC (Figure 10.36). However, this should not be held against DHC as it was mentioned earlier that DHC has not been designed to find local minima, but to use local minima information to guide its search to the global minimum.

To compare the robustness of a fitness sharing GA with KNRS, in Figure 10.37 all the minima found at the end of 2500 evaluations by KNRS, and the final population at the end of 200 generations for the fitness sharing GA are plotted, respectively. The fitness sharing GA performs well for the Rastrigin's, Himmelblau's, and the Six-hump camel back functions. However, it performs poorly on the Branin and Rosenbrock's functions.

Unlike a fitness sharing GA, one advantage of KNRS is that it does not require a guess regarding how many minima there could be in the function. The setting of appropriate σ_{share}, and a corresponding population size that can maintain sub-populations at all peaks is very hard to determine *a priori*. On the other hand, KNRS explores the entire space sequentially and uniformly. The only parameter that KNRS, in its current implementation, requires is to set the maximum number of inner iterations for the NR sequence. This parameter can be completely eliminated by performing a check for oscillation or divergence inside the NR iteration loop.

10.9 Limitations

Even though this chapter has compared the performance of KNRS, genetic algorithms (both standard and fitness sharing), and DHC, for a suite of 10 test functions, it is important to note that all the test functions are two dimensional. The performance of KNRS relative to any other algorithm on functions in higher dimensions remains to be evaluated. In higher dimensions, KNRS , owwing to the space filling features of the SOBOL sequence will even-



(a) Branin function (KNRS)

(b) Branin function (DHC)

(c) Rosenbrock's function (KNRS)

(d) Rosenbrock's function (DHC)

(e) Himmemblau's function (KNRS)

(f) Himmemblau's function (DHC)

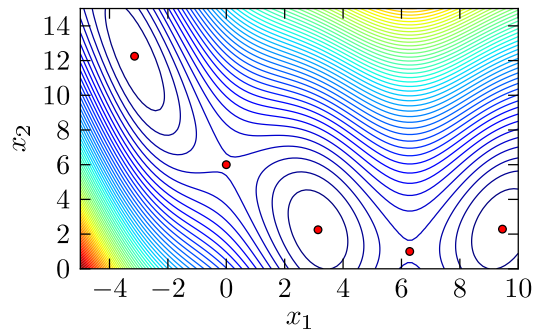
(g) Six-hump camel back function (KNRS)

(h) Six-hump camel back function (DHC)

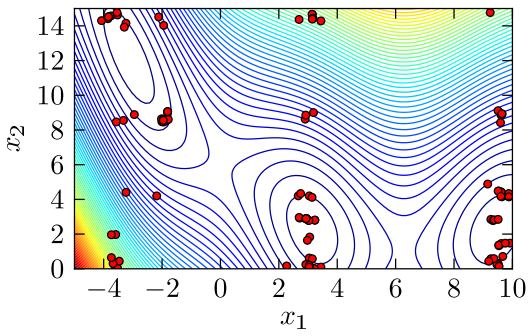
(i) Rastrigin's function (KNRS)

(j) Rastrigin's function (DHC)

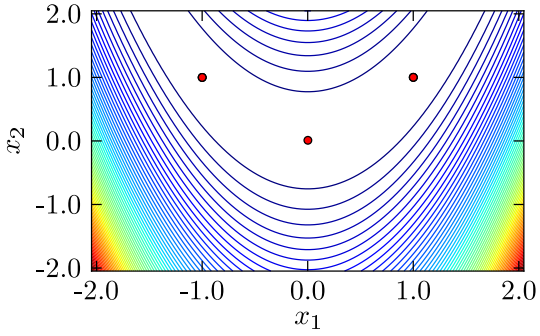
Figure 10.36: Performance of KNRS and DHC for five test functions



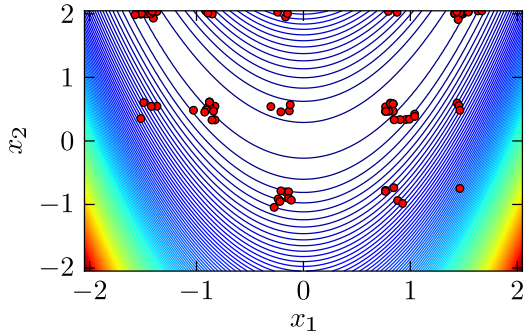
(a) Branin function (KNRS)



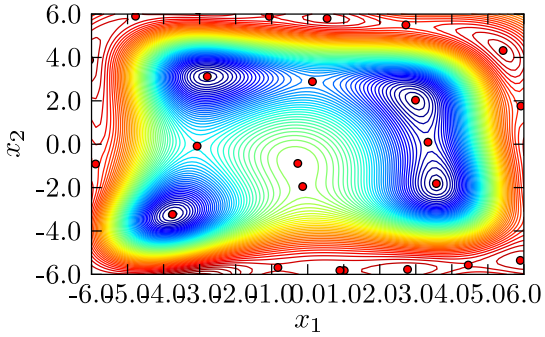
(b) Branin function (GA)



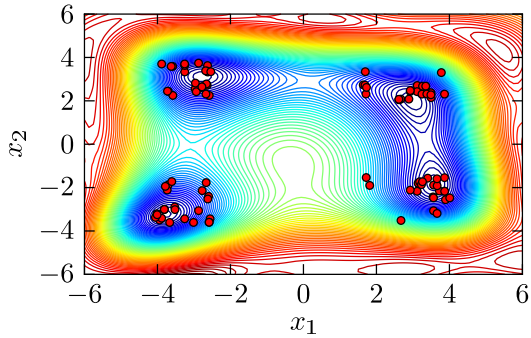
(c) Rosenbrock's function (KNRS)



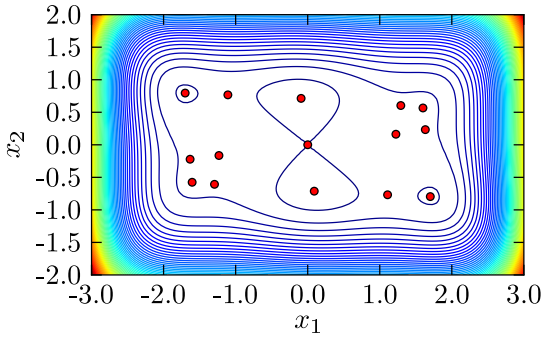
(d) Rosenbrock's function (GA)



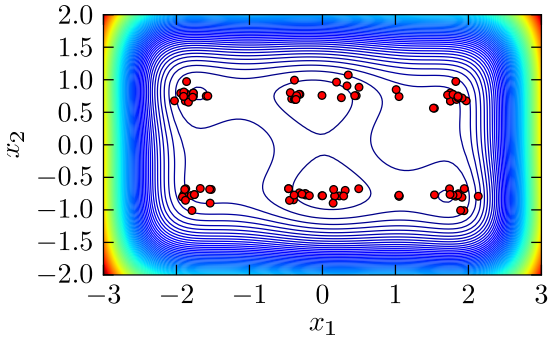
(e) Himmemblau's function (KNRS)



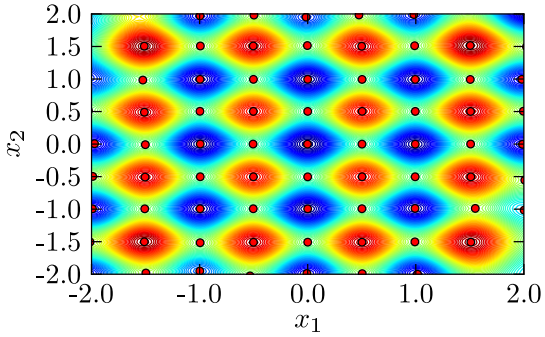
(f) Himmemblau's function (GA)



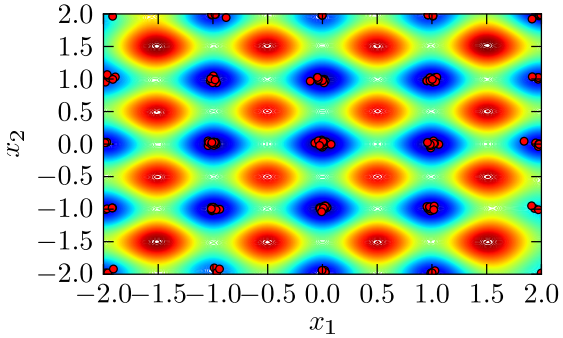
(g) Six-hump camel back function (KNRS)



(h) Six-hump camel back function (GA)



(i) Rastrigin's function (KNRS)



(j) Rastrigin's function (GA)

Figure 10.37: Performance of KNRS and fitness sharing GA for five test functions: results at 200 generations of GA with a population of 100 members (except for Rastrigin where population size is 200); and KNRS for 2500 evaluations of each function

tually find all the optima; however this chapter has not addressed how effective this process might be as opposed to standard optimisation algorithms. Lastly, as observed in the case for Bump function, in section 10.7.3.1, KNRS does not, in its current form, honour bounds. Consequently, for constrained optimisation a modification to KNRS is required.

10.10 Note on SOBOL sampling in KNRS

Even though the idea of global exploration through the use of a low-discrepancy sequence, such as the SOBOL sequence, is developed independently by the author, there have been similar use of low-discrepancy sequences for global optimisation in the past. These methods go by the name of *Single Linkage* (SL) and *Multi-Level Single Linkage* (MLSL) algorithms. The reader is referred to [221, 222, 223, 224, 225] for details of such algorithms. An important aspect of these methods, which is also applicable to the KNRS algorithm, is that based on the number of starting points for the local search, such as steepest descent (or NR sequence in KNRS), and the number of local minima found, Bayesian estimates on the number of local minima in the underlying function can be ascertained, see for example [221, 222, 223]. The evaluation of such estimates, which can be used as a stopping criterion for the KNRS algorithm is out of scope of this thesis, but forms an important part for future work. Similarly, theoretical analyses such as those presented in [226, 227], can be used to obtain statistical evidence that all the local minima of a function have been found. Lastly, it should be noted that low-discrepancy sequences have also been used in evolutionary algorithms and evolutionary strategies. The reader is referred to [228, 229] for details of these.

10.11 Conclusions

The following conclusions can be made from this chapter:

1. Starting with a realisation that Kriging predictors can be differentiated and that the derivative information has hitherto not been explored, an expression for the derivative of the Kriging predictor is derived. This expression is then manipulated to form a fixed-point iterative sequence to find the stationary points of a Kriging predictor.

2. The fixed point iterative sequence is evaluated for its convergence and its convergence is explored on the Branin function. In order to overcome the limitations of the fixed point iteration, a modification of the sequence is proposed. Moreover, a Newton-Raphson equivalent of the same fixed-point sequence is derived, which guarantees convergence to stationary points provided the starting point is close enough to the stationary point (within the basin of attraction).

3. The convergence of the MFP and NR sequences are further explored and tested on a test suite of 10 two-dimensional functions.

4. Based on the performance of the Newton-Raphson sequence on the above test functions, a new algorithm, called KNRS, is proposed.

5. The performance of KNRS is tested for global optimisation against a genetic algorithm and a dynamic hill climbing algorithm for a test suite of 10 two-dimensional functions.

6. For global optimisation KNRS performs favourably relative to GA for all the test functions, and favourably relative to DHC on most test functions. From the results it is realised that an important merit of the KNRS algorithm lies in multimodal optimisation.

7. The performance of KNRS for multimodal optimisation is tested against a fitness sharing genetic algorithm and a dynamic hill climber.

8. On the test functions chosen, KNRS performs favourably relative to both a fitness sharing GA and DHC.

9. The capabilities of KNRS, both for global optimisation and multimodal optimisation, have not been evaluated in higher (greater than two) dimensional spaces.

Chapter 11

Conclusions & recommendations for further work

11.1 Conclusions

The following subsections present the conclusions made from this thesis.

11.1.1 Most Significant contributions

Even though the area of interventional cardiology has witnessed a rapid evolution in coronary stent designs – from bare-metal stents to drug-eluting stents to biodegradable stents – there exists a substantial gap between the engineering analysis of stents, their comparison based on such analysis, and the use of such comparisons to design better stents by the use of optimisation methods. To fill this gap, the most significant contributions made by this thesis, to the field of coronary stent design, analysis, and optimisation are:

1. the development of engineering models to evaluate a stent's performance based on various features relevant to in-stent restenosis, and formulation of the corresponding stent evaluation metrics,

2. demonstration of how design improvement can be obtained in both constrained and multiobjective optimisation studies for the problem of stent design,

3. demonstration of how the parameters defining stent geometry influence various physical features that are directly related to in-stent restenosis, and

4. demonstration of various paradigms which determine the choice of an ideal stent, from a set of stent designs, based on various needs.

11.1.2 Multiobjective study

A NURBS based three parameter parameterisation for the widely studied CYPHER stent, Cordis corporation, Johnson & Johnson company, is proposed. Based on this parameterisation, a multiobjective optimisation study, using surrogate modelling and NSGA-II, is

performed to minimise acute recoil, minimise volume average stresses, minimise haemodynamic alteration, maximise volume average drug delivered, maximise uniformity of drug-distribution, and maximise flexibility. It is found that the following pairs of objectives are in conflict with each other

- volume average stresses (arterial injury) and acute recoil;
- volume average drug delivered and volume average stresses (arterial injury);
- flexibility and volume average stresses (arterial injury);
- flexibility and flow disturbance;
- volume average drug delivered and flow disturbance;
- uniformity of drug distribution and volume average stresses.

The complex interplay between stent design (distribution of struts, link design, strut thickness, and circumferential ring design) and stent performance, from the perspective of the various conflicting/desirable properties, is demonstrated. It is found that while strut width and the axial length of the circumferential rings most affect volume average stresses and recoil, the length of the links in the cross-flow direction significantly affects volume average drug, flexibility, and the flow index. The non-dominated solutions, which represent a potentially optimum family of CYPHER like stents, for the proposed parameterisation are obtained and discussed. The position of a representative CYPHER stent in various slices of Pareto front is also shown. Based on the results, it is also hypothesised that the proposed metric for measuring the uniformity of drug-distribution, D_{dev}, might be misleading for comparing stents, especially when using Gaussian process based surrogate modelling.

Three paradigms, viz. *conservative, constrained, and experimental*, are proposed for the selection of optimal designs from the set of non-dominated solutions. Designs 20 and 30 are identified as the optimal designs under a conservative approach. Similarly, designs 15, 21, and 26, are identified as optimal from one of the views of constrained paradigm. The relation of these designs obtained by the multiobjective study with the results of constrained optimisation study is also shown.

11.1.3 Constrained optimisation study

A parameterisation technique for creating general circumferential-rings and wavy-links based stent geometries is proposed. A polynomial form for the parameterisation of links (flex connectors) is formulated to control the waviness of the links. Based on this parameterisation, a constrained optimisation study to obtain design improvement from the baseline geometry one objective at a time, without compromising any other objective, is performed. Significant design improvement is obtained for the three cases of individually minimising volume average stresses, maximising volume average drug delivered, and maximising flexibility, without deteriorating any other objective. It is found that the optimal designs have a higher strut width compared to the baseline geometry but a lower length of circumferential rings. It is concluded that strut width is a key determinant of acute recoil, volume average stresses, and drug delivered; and the axial length of the circumferential rings affects acute recoil most. It is also shown that the optimal designs minimise the number of curved regions in the flex connectors.

11.1.4 Findings

The following contributions are made from the engineering analysis of stents performed in this thesis

1. It is found that the length of flex connectors in contemporary stent designs, which are introduced to improve flexibility, significantly influences flow features relevant to in-stent restenosis. In particular, it is shown that a higher length of the flex connectors leads to deterioration of the relevant flow-features.

2. It is shown that while the plastic deformations in the curved regions of the stent are key in determining acute recoil, the contact area between the stent and the artery significantly affect volume average stresses.

3. It is shown that the ability of the flex connectors to delay self contact with increasing curvature index is key in determining the flexibility of a stent. In particular, it is shown that a higher delay in self contact and minimising the number of potential contact locations while bending, leads to improved flexibility.

4. It is shown that the balloon expansion of a stent in a stenosed coronary artery leads to significant tissue prolapse. This tissue prolapse affects the extent of flow recirculation zones formed in the stented artery. It is also shown that the formation of recirculation zones in and around the struts promotes delivery of drug from the lumen to the tissue.

11.1.5 Analysis models and evaluation metrics

The following contributions are made in terms of engineering analysis of stents in relation to in-stent restenosis

1. The analysis of coronary stents to assess haemodynamic alteration relevant to the adverse response of in-stent restenosis is demonstrated. To this effect, a computational fluid dynamics model is developed to simulate blood flow in a stented coronary artery.

2. The haemodynamic low and reverse flow index, $HLRFI$, that quantifies the flow features relevant to in-stent restenosis in a stented coronary artery is formulated.

3. The analysis of coronary stents to assess arterial injury during balloon expansion and minimum post-procedural stent area is demonstrated. To this effect, a finite element analysis model is adopted to model the balloon-expansion of stent in a representative stenosed coronary artery.

4. The metrics of $Recoil$, that measures acute recoil post stent implantation, and VAD, that measures the volume average stresses in the artery post stenting, are formulated. These are shown to measure structural strength of a stent and arterial injury due to stenting procedure, respectively.

5. The analysis of coronary stents to assess their flexibility is demonstrated. To this effect, a finite element analysis model is adopted to study the response of stent deformation on application of bending loads.

6. The flexibility metric, FM, which quantifies the flexibility of a stent, in both elastic and plastic deformation phases, in a single numeric quantity is formulated.

7. The analysis of coronary stents to assess their drug distribution capabilities is demonstrated. To this effect, a computational fluid dynamics model is developed to simulate drug-delivery in a stented coronary artery.

8. Two metrics of volume average drug, VAD, and standard deviation of drug delivered, D_{dev}, are formulated to quantify the properties of drug delivery in a stented coronary artery. The metrics quantify the amount of drug delivered and the uniformity of the drug delivered, respectively.

11.1.6 Conclusions for the KNRS algorithm

In the final chapter of this thesis, a methodology to search the Krig (Gaussian process model) of a function, based on the hitherto unexploited and analytically available derivative information is explored. To this end, the following contributions to the field of Gaussian process assisted optimisation are made:

1. An expression for the derivative of the Gaussian process predictor is derived.

2. The expression for the above derivative when equated to zero (in order to find stationary points) is manipulated to yield a standard fixed point iterative sequence (SFP).

3. To improve convergence of the SFP sequence to more stationary points, a modification of the sequence, the modified fixed point iterative (MFP) sequence, is proposed.

4. A Newton-Raphson (NR) equivalent of the SFP is proposed. It is shown that the NR sequence guarantees convergence to the stationary points of the Krig provided the starting point is close enough to the stationary point (within the basin of attraction).

5. The convergence of the MFP and NR sequences are further explored and tested on a suite of 10 two-dimensional functions.

6. Based on the performance of the Newton-Raphson sequence on the above test functions, a new algorithm, called Krige-Newton-Raphson-Sobol (KNRS) algorithm, is proposed.

7. The performance of KNRS is tested for global optimisation against a genetic algorithm and a dynamic hill climbing algorithm for a test suite of 10 two-dimensional functions.

8. It is found that for global optimisation KNRS performs favourably to GA for all the test functions, and favourably to DHC on most test functions. From the results it is realised that an important merit of the KNRS algorithm lies in multimodal optimisation.

9. The performance of KNRS for multimodal optimisation is tested against a fitness sharing genetic algorithm and a dynamic hill climber.

10. It is found that on the test functions chosen, KNRS performs favourably to both a fitness sharing GA and DHC.

11. The capabilities of KNRS on higher dimensional spaces remains to be explored.

11.2 Recommendations for Further Work

Although a number of conclusions have been made regarding the influence of stent design parameters on factors affecting in-stent restenosis, a number of improvements can be made in both stent analysis and stent optimisation. In the area of searching the GP models, KNRS offers a range of extensions for further investigation. The areas of further work can be broadly classified into

1. Improvements in stent analysis
    a) Realistic geometry of stenosed arteries;
    b) Realistic morphology of stenosed arteries (layers, anisotropy);
    c) Transient release of drug-distribution (including factors like reverse binding of drug).
2. Improvements in stent optimisation methodology
    a) Parameterisation techniques for stent design;
    b) Refinement of objective functions;
    c) Refinement in terms of surrogate update methods;
    d) Treatment of the design problem as a many-objective problem.
3. Investigation and extension of KNRS
    a) Investigation of performance in higher dimensional spaces;
    b) Possibility of a surrogate update method by KNRS;
    c) Extension to constrained optimisation;
    d) Extension to multiobjective optimisation.

The following sections outline each of the above themes.

11.3 Improvements in stent analysis

Further work in this area is primarily needed to make the stent analyses more realistic. In particular, efforts in the following areas are required

11.3.1 Realistic geometry of stenosed arteries

Throughout this thesis, a representative model for the geometry of stenosed artery is used. The artery is assumed to be cylindrical and the plaque is defined by the Hicks Henne bump function. In reality, the geometries of stenosed arteries can be very complex with high curvature and asymmetric distribution of plaque. An inclusion of such geometries, which can be obtained by intravascular ultrasound and magnetic resonance imaging techniques, in the analyses models – for both balloon-expansion and haemodynamic alteration – will further enhance the understanding of the interaction between the stent and the artery.

11.3.2 Realistic morphology of stenosed arteries

Apart from the complex geometries of stenosed arteries as discussed above, the morphology of the disease can be quite complex as well. In this thesis, an isotropic single layer for both the plaque and the artery is assumed. In reality, as discussed in chapter 2.2.3, the artery is composed of three layers: the intima, the media, and the adventitia. Furthermore, the variation in plaque composition can be significant, ranging from soft fibrous plaque to hard calcified plaque in the same lesion. Lastly, all human tissue is usually anisotropic. An inclusion of such morphological properties in the stent analyses will bridge the gap between the results of computational analyses and clinically observed outcomes.

11.3.3 Transient release of drug-distribution

In this thesis, a steady-state drug-distribution model is used. Even though this model captures the result of geometrical variation in the amount of drug-delivered, the model does not give much insight on the transient process of drug-delivery. To this end, further investigation in terms of development of a transient advection-diffusion drug-distribution model, which includes effects of reverse-binding of the drug (endocytosis), is required.

11.4 Improvements in stent optimisation methodology

Further work in the area of improving the stent design optimisation method includes the following

11.4.1 Parameterisation techniques for stent design

One of the most important features of any optimisation study is the parameterisation chosen to represent geometries. The optimisation algorithms can only find the optimum solution(s) in the design space represented by the chosen parameterisation. Hence, better techniques to parametrically represent stent geometries can be implemented, so that a wider design space represented by the parameterisation. A few possible parameterisation techniques are listed below

- Using an n control point approach to define a network of NURBS curves representing the stent struts on a flat plane. Specification of appropriate constraints on such curves can lead to a vast and novel design space, which can then be used for design optimisation studies.
- The Class-Shape-Transformation (CST) method suggested by Kulfan [230] can be used to define the 2-D patterns, which can then be repeated periodically to create stent designs using approach 2 (see Section 2.6.1.1).
- A custom shape function for a baseline stent geometry can be defined, which when combined with polynomials such as the Bernstein polynomials will lead to creation of new designs.
- Shape grammar can be employed to use the information from the currently existing stents to create new designs which combine the features of multiple, currently existing, stent designs.

- A mesh morphing technique can be implemented to directly manipulate the mesh, rather than the curves/surfaces defining the stent geometry. This is particularly attractive as automated meshing of complex geometries is currently quite hard.

- Intellectually most challenging, inspiration from nature such as the bifurcation patterns of the arteries, patterns of vein bifurcation in leaves, can be used to create stent parameterisation techniques. These, by virtue of natural evolution are optimised at least for one objective (for example, the vein bifurcation patterns of leaves is optimised for nutrient transport which can be seen as analogous to drug-delivery).

11.4.2 Refinement of objective functions

In this thesis several objective functions which relate measurable quantities from engineering analyses to the factors that contribute towards in-stent restenosis are formulated. As a first approximation, average measures for stress and drug delivered are introduced. However, it is not yet clinically known whether it is the peak stresses in the artery, irrespective of the location where such peak stresses occur, or it is the volume average measures that the human body responds to in terms of restenosis. Similarly, for drug-delivery, perhaps only the volumes of tissue which receive less drug than a certain therapeutic threshold, and not the average amount of drug delivered, are relevant in terms of restenosis. Further investigation, more from the clinical community, is needed in such regards.

11.4.3 Refinement in terms of surrogate update methods

It was discussed in chapter 9 that many methods to decide update points for multi-objective kriging assisted optimisation have been proposed [187, 188, 189, 190, 191, 192, 193]. The choice of these update methods, as opposed to the simple pure exploitative approach adopted in this chapter, could lead to an improved set of non-dominated solutions. Further investigation, regarding the performance of such algorithms for the problem of stent design will be very useful.

11.4.4 Treatment of the design problem as a many-objective problem

This results of this thesis have shown that many objective pairs in the optimisation problem defined in chapter 9 are correlated to each other, rather than being in conflict. Methods from the field of many-objective optimisation methods, can be employed to reduce the number of objectives to fewer than six to decrease the complexity of the problem. This will both minimise the time required for high-fidelity simulations and with fewer objectives will lead to better estimation of the Pareto front.

11.5 Investigation and extension of KNRS

In the development of KNRS, an intial algorithm is proposed in this thesis. However, further investigation of the performance of this algorithm is needed in the following areas

11.5.1 Bayesian analysis for an estimate of the number of local minima

Using methods such as those presented in [221, 222, 223], an analysis which determines a Bayesian estimate on the number of local minima of the underlying function, based on the number of NR sequence starting points and the number of local minima found by such starting points, needs to be performed in future. Such an analysis can lead to an efficient stopping criterion for the KNRS algorithm.

11.5.2 Investigation of performance in higher dimensional spaces

The 10 test functions that are chosen in this thesis are all complex but 2-dimensional. Consequently, the relative performance of KNRS with other algorithms like GA, which are known to have a better search power in higher dimensional spaces, for both global and multimodal optimisation remains to be explored.

11.5.3 Possibility of a surrogate update method by KNRS

With the search performed by KNRS, one advantage is that all the stationary points of the Krig are located. Perhaps, using this information about the landscape of the current form of the Krig, an update methodology to balance exploration and exploitation can be developed. The ideas for this are still at a nascent stage, for which further work is required.

11.5.4 Extension to constrained optimisation

The results of the performance of KNRS on the 10 test functions, in particular the Bump function, showed that KNRS, in its current form, does not honour any constraints. Hence, a modification of KNRS to include both equality and inequality constraints is needed. Encouragingly, since the form of equations used in KNRS is universal, i.e. it is related to the Krig rather than the function itself and hence has the same mathematical form for every function, methods from classical optimisaion (such as gradient projection methods) can be used to formulate the constrained version of the algorithm.

11.5.5 Extension to multiobjective optimisation

Lastly, an extension of KNRS for multiobjective optimisation is needed. The author has made significant progress in this regard. This formulation and some preliminary results are presented in appendix D. This extension to multiobjective optimisation, however, needs more testing in higher dimensional objective spaces.

Appendices

Appendix A

Conditional distribution for a Gaussian Process

In chapter 3, section 3.2.3, it was said that the joint probability distribution of the function value at a new point, y^*, and the function values at all the sample points, $\mathbf{y}$, is assumed to be jointly Normal. Here the conditional distribution of y^* given the values of $\mathbf{y}$ is derived. From equation 3.33 we have

$$\begin{bmatrix} \mathbf{y} \\ y^* \end{bmatrix} \backsim \mathcal{N}\left(\begin{bmatrix} \mathbf{1}_{p+1}\beta \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Gamma} & \boldsymbol{\gamma}(\boldsymbol{x}^*) \\ \boldsymbol{\gamma}(\boldsymbol{x}^*)^T & \Gamma(\mathbf{x}^*, \mathbf{x}^*) \end{bmatrix} \right) \tag{A.1}$$

The conditional distribution of y^* given $\mathbf{y}$ is

$$p(y^*|\mathbf{y}) = \frac{p(\mathbf{y}, y^*)}{p(\mathbf{y})} \tag{A.2}$$

where p represents the probability. From equation A.1 $p(\mathbf{y}, y^*)$ can be written as

$$p(\mathbf{y}, y^*) = (2\pi)^{-(p+1)/2}|\Sigma|^{-1/2}\exp\left(-\frac{1}{2}\begin{bmatrix}(\mathbf{y} - \mathbf{1}_p\,\beta) & (y^* - \beta)\end{bmatrix}\Sigma^{-1}\begin{bmatrix}\mathbf{y} - \mathbf{1}_p\,\beta \\ y^* - \beta\end{bmatrix}\right) \tag{A.3}$$

where

$$\Sigma = \begin{bmatrix} \boldsymbol{\Gamma} & \boldsymbol{\gamma}(\boldsymbol{x}^*) \\ \boldsymbol{\gamma}(\boldsymbol{x}^*)^T & \Gamma(\mathbf{x}^*, \mathbf{x}^*) \end{bmatrix} \tag{A.4}$$

hence

$$p(\mathbf{y}, y^*) = (2\pi)^{-(p+1)/2}|\Sigma|^{-1/2}\exp\left(-\frac{1}{2}\begin{bmatrix}(\mathbf{y} - \mathbf{1}_p\,\beta)^T & (y^* - \beta)\end{bmatrix}\begin{bmatrix} \boldsymbol{\Gamma} & \boldsymbol{\gamma}(\boldsymbol{x}^*) \\ \boldsymbol{\gamma}(\boldsymbol{x}^*)^T & \Gamma(\mathbf{x}^*, \mathbf{x}^*) \end{bmatrix}^{-1}\begin{bmatrix}\mathbf{y} - \mathbf{1}_p\,\beta \\ y^* - \beta\end{bmatrix}\right) \tag{A.5}$$

The partitioned inverse using Schur complement of Σ can be written as

$$\Sigma^{-1} = \begin{bmatrix} \mathbf{I}_p & -\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}(\boldsymbol{x}^*) \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \boldsymbol{\Gamma}^{-1} & 0 \\ 0 & \mathbf{M} \end{bmatrix}\begin{bmatrix} \mathbf{I}_p & 0 \\ -\boldsymbol{\gamma}(\boldsymbol{x}^*)^T\boldsymbol{\Gamma}^{-1} & 1 \end{bmatrix} \tag{A.6}$$

where $\mathbf{I}_p \in \mathbb{R}^{p\times p}$ is an identity matrix, and $\mathbf{M}$ is the Schur complement

$$\mathbf{M} = \left(\Gamma(\mathbf{x}^*, \mathbf{x}^*) - \boldsymbol{\gamma}(\boldsymbol{x}^*)^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}(\boldsymbol{x}^*)\right)^{-1} \tag{A.7}$$

Using this, the terms inside the exponential in equation A.5 can be written as

$$-\frac{1}{2}\left[(\mathbf{y}-\mathbf{1}_p\beta)^T \quad (y^*-\beta)\right]\begin{bmatrix}\mathbf{I}_p & -\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}(\boldsymbol{x}^*)\\ 0 & 1\end{bmatrix}\begin{bmatrix}\boldsymbol{\Gamma}^{-1} & 0\\ 0 & \mathbf{M}\end{bmatrix}\begin{bmatrix}\mathbf{I}_p & 0\\ -\boldsymbol{\gamma}(\boldsymbol{x}^*)^T\boldsymbol{\Gamma}^{-1} & 1\end{bmatrix}\begin{bmatrix}\mathbf{y}-\mathbf{1}_p\beta\\ y^*-\beta\end{bmatrix} \tag{A.8}$$

or

$$-\frac{1}{2}\left[(\mathbf{y}-\mathbf{1}_p\beta)^T \quad (y^*-\beta)-(\mathbf{y}-\mathbf{1}_p\beta)^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}(\boldsymbol{x}^*)^T\right]\begin{bmatrix}\boldsymbol{\Gamma}^{-1} & 0\\ 0 & \mathbf{M}\end{bmatrix}\begin{bmatrix}\mathbf{y}-\mathbf{1}_p\beta\\ (y^*-\beta)-\boldsymbol{\gamma}(\boldsymbol{x}^*)^T\boldsymbol{\Gamma}^{-1}(\mathbf{y}-\mathbf{1}_p)\beta\end{bmatrix} \tag{A.9}$$

or

$$-\frac{1}{2}\left[(\mathbf{y}-\mathbf{1}_p\beta)^T \quad \left((y^*-\beta)-\boldsymbol{\gamma}(\boldsymbol{x}^*)\boldsymbol{\Gamma}^{-1}(\mathbf{y}-\mathbf{1}_p\beta)\right)^T\right]\begin{bmatrix}\boldsymbol{\Gamma}^{-1}(\mathbf{y}-\mathbf{1}_p\beta)\\ \mathbf{M}\left((y^*-\beta)-\boldsymbol{\gamma}(\boldsymbol{x}^*)^T\boldsymbol{\Gamma}^{-1}(\mathbf{y}-\mathbf{1}_p)\beta\right)\end{bmatrix} \tag{A.10}$$

or

$$-\frac{1}{2}\left[(\mathbf{y}-\mathbf{1}_p\beta)^T\boldsymbol{\Gamma}^{-1}(\mathbf{y}-\mathbf{1}_p\beta)\right]\ -\frac{1}{2}\left[\left((y^*-\beta)-\boldsymbol{\gamma}(\boldsymbol{x}^*)\boldsymbol{\Gamma}^{-1}(\mathbf{y}-\mathbf{1}_p\beta)\right)^T\mathbf{M}\left((y^*-\beta)-\boldsymbol{\gamma}(\boldsymbol{x}^*)^T\boldsymbol{\Gamma}^{-1}(\mathbf{y}-\mathbf{1}_p)\beta\right)\right] \tag{A.11}$$

$$-\frac{1}{2}\left[(\mathbf{y}-\mathbf{1}_p\beta)^T\boldsymbol{\Gamma}^{-1}(\mathbf{y}-\mathbf{1}_p\beta)\right]\ -\frac{1}{2}\left[\left(y^*-\left[\beta+\boldsymbol{\gamma}(\boldsymbol{x}^*)\boldsymbol{\Gamma}^{-1}(\mathbf{y}-\mathbf{1}_p\beta)\right]\right)^T\mathbf{M}\left(y^*-\left[\beta+\boldsymbol{\gamma}(\boldsymbol{x}^*)^T\boldsymbol{\Gamma}^{-1}(\mathbf{y}-\mathbf{1}_p)\beta\right)\right]\right] \tag{A.12}$$

or

$$-\frac{1}{2}\left[(\mathbf{y}-\mathbf{1}_p\beta)^T\boldsymbol{\Gamma}^{-1}(\mathbf{y}-\mathbf{1}_p\beta)\right]\ -\frac{1}{2}\left[(y^*-\mu^*)^T\mathbf{M}(y^*-\mu^*)\right] \tag{A.13}$$

where

$$\mu^* = \left[\beta+\boldsymbol{\gamma}(\boldsymbol{x}^*)\boldsymbol{\Gamma}^{-1}(\mathbf{y}-\mathbf{1}_p\beta)\right] \tag{A.14}$$

substituting this in equation A.5

$$p(\mathbf{y},y^*) = (2\pi)^{-(p+1)/2}|\Sigma|^{-1/2}\exp\left(-\frac{1}{2}\left[(\mathbf{y}-\mathbf{1}_p\beta)^T\boldsymbol{\Gamma}^{-1}(\mathbf{y}-\mathbf{1}_p\beta)\right]\ -\frac{1}{2}\left[(y^*-\mu^*)^T\mathbf{M}(y^*-\mu^*)\right]\right) \tag{A.15}$$

Now, in equation A.2, $p(\mathbf{y})$, is

$$p(\mathbf{y}) = (2\pi)^{-(p)/2}|\boldsymbol{\Gamma}|^{-1/2}\exp\left(-\frac{1}{2}\left[(\mathbf{y}-\mathbf{1}_p\beta)^T\boldsymbol{\Gamma}^{-1}(\mathbf{y}-\mathbf{1}_p\beta)\right]\right) \tag{A.16}$$

Hence, using equations A.15 and A.16, the condition distribution given by equation A.2 is

$$p(y^*|\mathbf{y}) = \frac{p(\mathbf{y},y^*)}{p(\mathbf{y})} = (2\pi)^{-(1)/2}\frac{|\Sigma|^{-1/2}}{|\boldsymbol{\Gamma}|^{-1/2}}\exp\left(-\frac{1}{2}\left[(y^*-\mu^*)^T\mathbf{M}(y^*-\mu^*)\right]\right) \tag{A.17}$$

From equation A.6, it can be seen that $(|\Sigma|^{-1/2}/|\boldsymbol{\Gamma}|^{-1/2}) = |\mathbf{M}|^{-1/2}$. Hence the conditional distribution is

$$p(y^*|\mathbf{y}) = (2\pi)^{-(1)/2}|\mathbf{M}|^{-1/2}\exp\left(-\frac{1}{2}\left[(y^*-\mu^*)^T\mathbf{M}(y^*-\mu^*)\right]\right) \tag{A.18}$$

which is a Gauusian with mean

$$\mu^* = \left[\beta + \boldsymbol{\gamma}(\boldsymbol{x}^*)\boldsymbol{\Gamma}^{-1}(\mathbf{y} - \mathbf{1}_p\,\beta)\right] \tag{A.19}$$

and variance

$$\mathbf{M} = \left(\Gamma(\mathbf{x}^*, \mathbf{x}^*) - \boldsymbol{\gamma}(\boldsymbol{x}^*)^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}(\boldsymbol{x}^*)\right)^{-1} \tag{A.20}$$

that is,

$$y^*|\mathbf{y} \backsim \mathcal{N}\left(\beta + \boldsymbol{\gamma}(\boldsymbol{x}^*)^T\boldsymbol{\Gamma}^{-1}(\mathbf{y} - \mathbf{1}\beta)\ ,\ \Gamma(\mathbf{x}^*, \mathbf{x}^*) - \boldsymbol{\gamma}(\boldsymbol{x}^*)^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}(\boldsymbol{x}^*)\right) \tag{A.21}$$

Appendix B

Effect of shear-thinning on results of chapter 4

This appendix shows the results of comparison between the WSS patterns when blood is considered to be a Newtonian and non-Newtonian fluid. Yilmaz and Gundogdu [231] present an excellent review of the various Non-Newtonian models that have been proposed to model blood flow. In this appendix a comparison of the model used in this chapter with the widely used Carreau model [231, 16] is presented. The Carreau model relates the shear rate, $\dot{\gamma}$ to the fluid viscosity, μ, as

$$\mu = \mu_\infty + (\mu_0 - \mu_\infty)[1 + (\lambda\dot{\gamma})^2]^{(n-1)/2} \tag{B.1}$$

where μ is the effective viscosity, μ_∞ and μ_0 are fluid viscosities at infinite and zero shear rates, $\dot{\gamma}$ is the shear rate, λ is a time constant, and n is the power law index.

A steady state CFD simulation for Stent-C (see chapter 4) is used for a comparison between the Newtonian model and the Carreau model. The values of Carreau model parameters are taken from [16]. Figure B.2 shows a plot of WSS magnitude, for the two models. Similarly, figure B.2 shows a contour plot of WSS magnitude for the two cases. It is found from these figures that the Newtonian model underpredicts WSS marginally when compared to the non-Newtonian model. This marginal underprediction minimally effects the comparison of stents based on the *HLRFI* metric as the difference in WSS is related to the fluid viscosity properties (not the stent geometry) implying that such underprediction is applicable to all stents irrespective of their geometry.

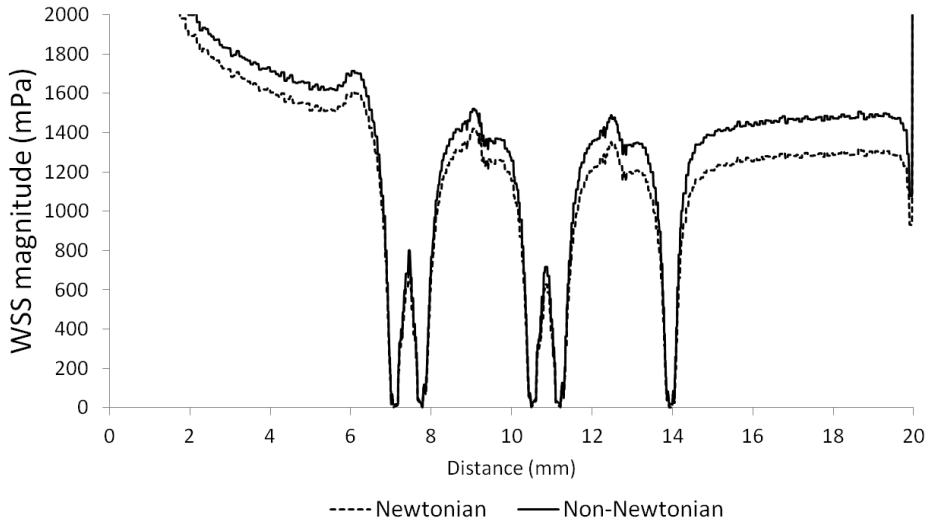


Figure B.1: Wall shear stress magnitude on the central line of a representative NIR stent (Stent-C in chapter 4): results for steady state flow comparison with Newtonian and non-Newtonian blood properties. The non-newtonian model adopted is the Carreu model [16]; steady state flow velocity = 0.1382 m/s

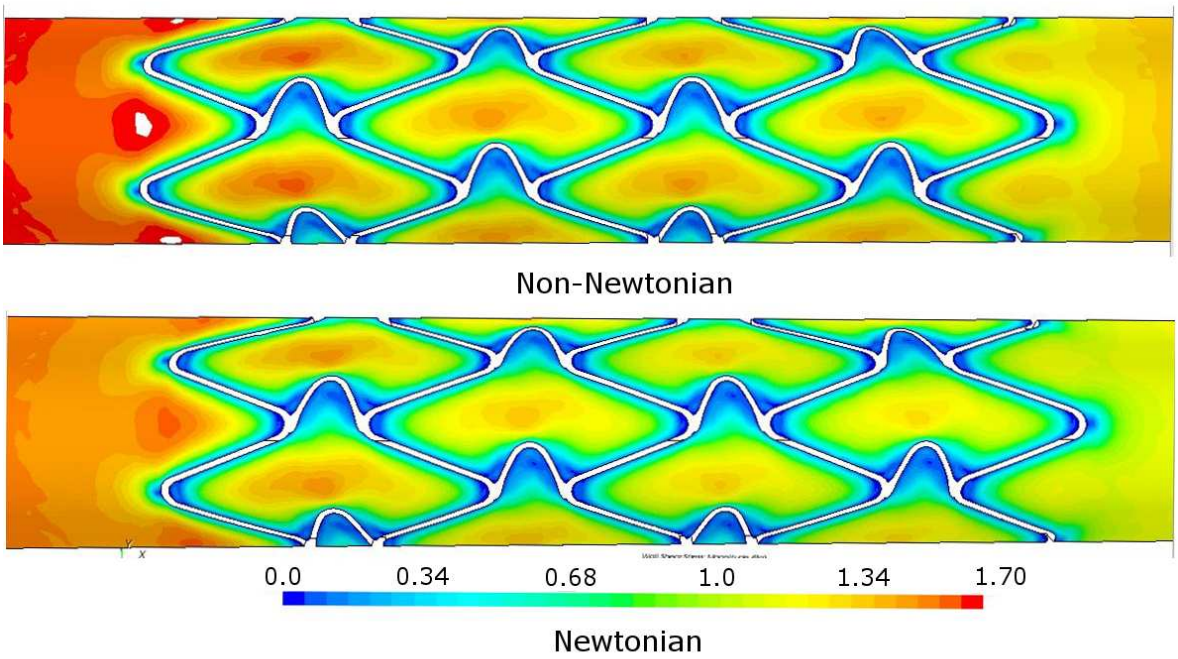


Figure B.2: Wall shear stress magnitude color plot on a representative NIR stent (Stent-C in chapter 4): results for steady state flow comparison with Newtonian and non-Newtonian blood properties. The non-newtonian model adopted is the Carreu model [16]; steady state flow velocity = 0.1382 m/s

Appendix C

Results for baseline geometry for constrained optimisation study

In this appendix the results for balloon expansion analysis and flexibility analysis for the baseline geometry used in chapter 7 (figure 7.3) are presented.

Figure C.1 shows the different stages of the transient balloon expansion of the stent. In the initial phase of expansion (cf Fig. C.1c) the stent expands unevenly along its length. This phenomenon, commonly referred to as dogboning, where the expansion is relatively larger in the longitudinal ends than the centre, can be observed. A recoil of 0.17 mm, calculated using Eq. 5.10, can be observed visually in Figs. C.1e and C.1f, which show the states of the assembly at the peak load and post-unloading conditions, respectively. This acute recoil can be seen as the 'spring back' effect owing to the elasto-plastic properties of the stent and the loading imparted on to the stent by the stretched artery. The average radius vs. time plot shows this effect clearly in Fig. C.2d. The average radius maintains a peak value during the hold phase of the loading cycle, and as the load is removed, the stent contracts as a result of reduced loading stresses. One of the goals while designing a good stent is to either minimise this recoil or apply a maximum limit to it.

The final stresses (after unloading) in the stent, plaque, and the artery are shown in Figure C.2. For the stent, it can be seen that the peak stresses are located in the curved parts of the circumferential rings as they form the key feature in a stent design to allow for expansion. In the plaque, maximum stresses are observed where contact occurs between the stent struts and plaque surface. These contact stresses lead to an observable stent imprint on the plaque both in terms of geometrical deformation and stress contours. The stresses in the artery are considerably lower than the stresses in the plaque (a difference of roughly one order of magnitude). It is interesting to note that even though the stent imprint can be seen in the form of stresses along the length of both plaque and the artery, the stresses in the central part are higher than on the distal ends. This is attributed to the shape of the stenosis. Since the stenosis is highest in the central region (owing to the Hicks-Henne parameter x_p=0.5), the stent faces higher expansion resistance in the central region, resulting in higher stresses. The volume averaged von-Mises stress in the plaque and the artery combined (calculated using Eq. 5.12) is 45.8×10^{-3} MPa.

Figure C.3 shows the distribution of max. principal plastic strains on the stent after

expansion. As with the stresses it can be readily observed that the curved regions of the circumferential rings are the regions where maximum plastic strains occur. These plastic deformations restrict recoil of the stent. A geometry which allows more plastic deformation in these areas has less acute recoil. It can also be observed from Fig. C.3 that the curved regions of the links have relatively higher plastic strains than the straight segments of both the circumferential rings and the links. When the circumferential rings expand their axial length decreases, to account for the increased angle between the struts of the circumferential rings. This causes an axial stretching of the links which leads to plastic deformation at the curved tips of the links.

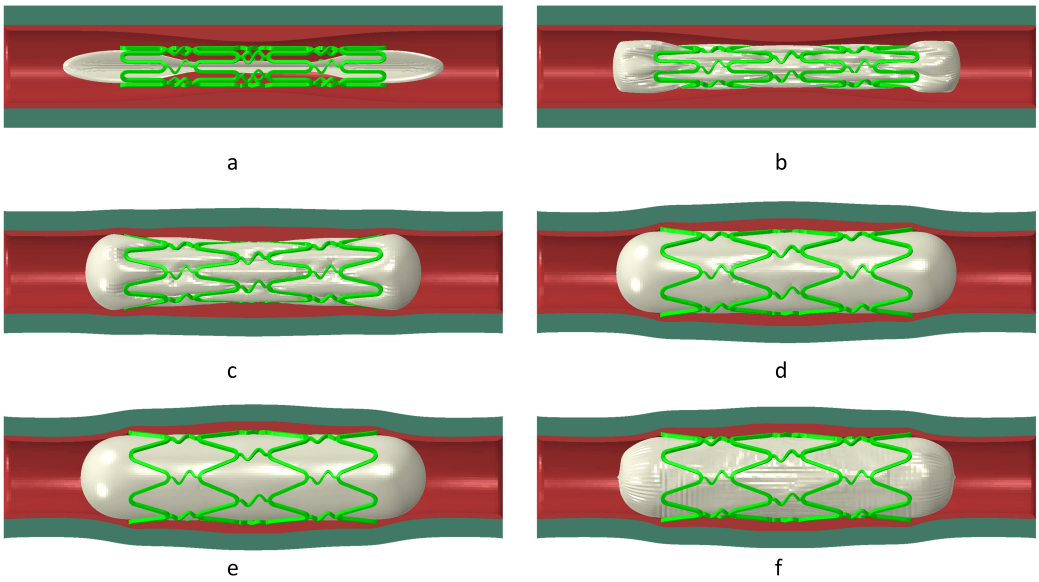


Figure C.1: Transient balloon expansion of the baseline geometry stent

Figure C.4 shows the results of the flexibility analysis, in particular the deformed state

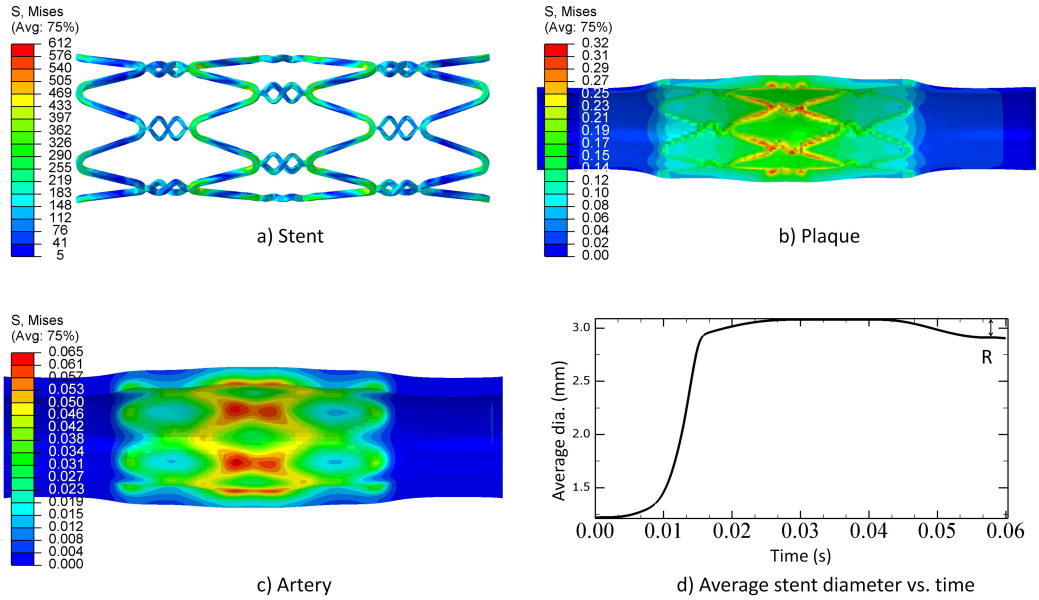


Figure C.2: Stent, plaque, and artery final stresses, and average radius vs. time plot for the baseline geometry

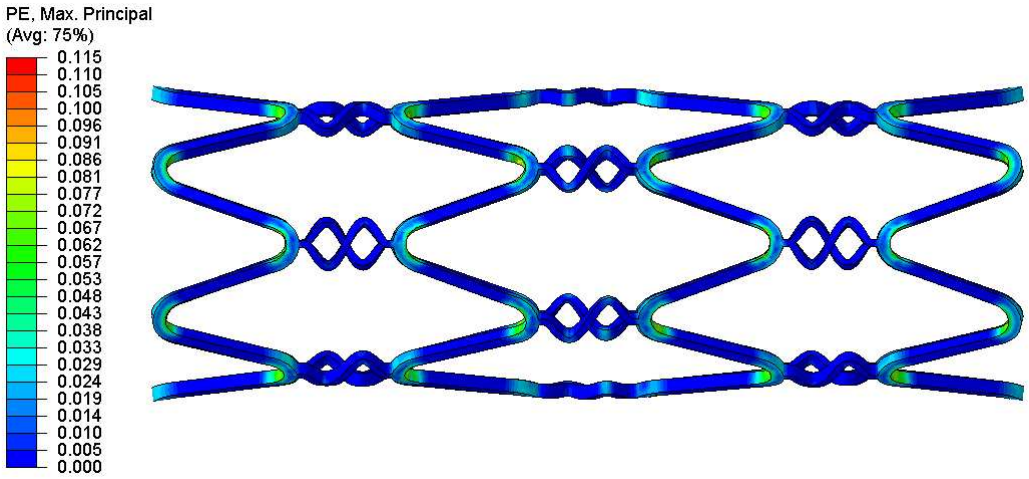


Figure C.3: Max. principal plastic strains on the stent post-expansion

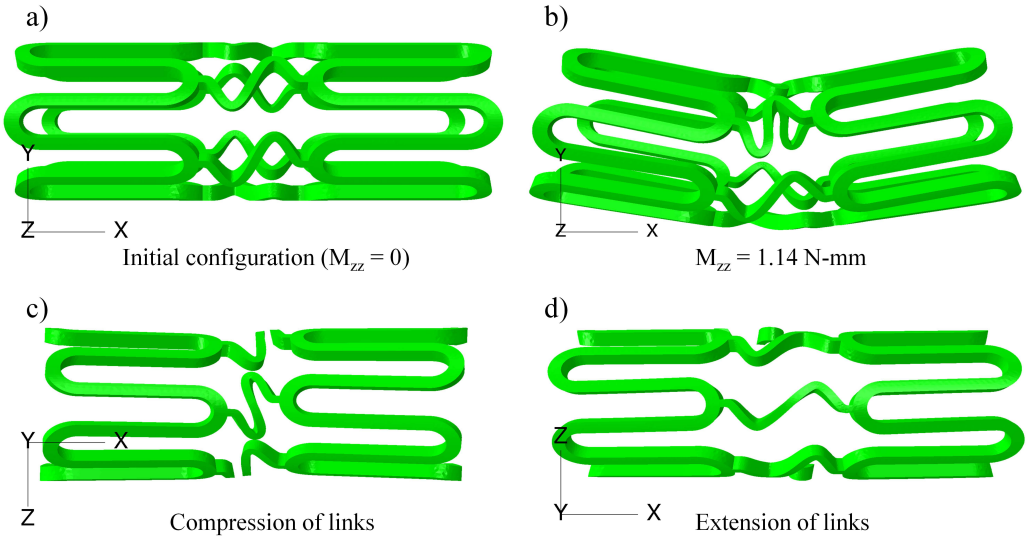


Figure C.4: Flexibility analysis for the baseline geometry: a) initial shape; b) deformed shape; c) deformed shape (y^+ view); deformed shape (y^- view)

of the stent with a moment of 1.14 N-mm is shown. As can be observed in this figure, the links play an important role in determining the flexibility of the stent. As the stent bends, the links on one side compress while the those on the other side expand to allow for the bending curvature. This can be seen in the bottom two images of Fig. C.4. It is apparent that the links on the compression side will come into self contact after a certain level of curvature. Such contact, which depends on the specific design of the links, can result in increased resistance to bending. A plot of the moment vs. the curvature index, the $M - \chi$ curve, is shown in Fig. C.5. This curve has two parts – the initial linear part which shows the regime when the deformations are only elastic in nature, and a curved (relatively flat) part when the deformation reach the plastic domain. The area under the $M - \chi$ curve, a lower value of which implies a more flexible stent, calculated using Eq. 6.1 is 50.45×10^{-2} N-rad.

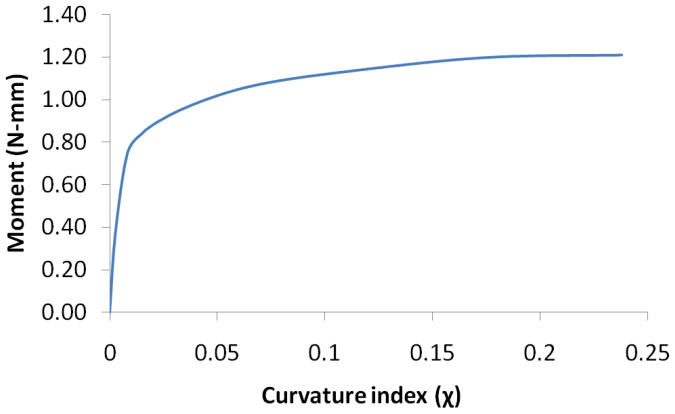


Figure C.5: Moment-curvature index curve for the baseline geometry

Appendix D

Extension of KNRS to multiobjective optimisation

In this appendix, an extension to KNRS algorithm, see section 10.6, is proposed. This extension is then tested on a few biobjective test problems.

D.1 Formulation of the multiobjective KNRS algorithm

The formulation of the multiobjective KNRS algorithm is based on the normalised normal constraint (NC) method proposed by Messac et. al. [17]. This extension is referred to as M-KNRS method, the 'M' referring to both multiobjective and Messac.

In this appendix, the development of the M-KNRS method is shown only for a bi-objective problem. The notation used in chapters 3 and 10 is adopted here. The Kriging predictors for the two objectives are referred to as $y_1(\mathbf{x})$ and $y_2(\mathbf{x})$, which have been created by sampling the real function at p_1 and p_2 points respectively. Hence the Kriging predictors for the two responses are

$$y_1(\mathbf{x}) = \beta_1 + \mathbf{r}_1(\mathbf{x})^T \mathbf{w}_1, \tag{D.1}$$

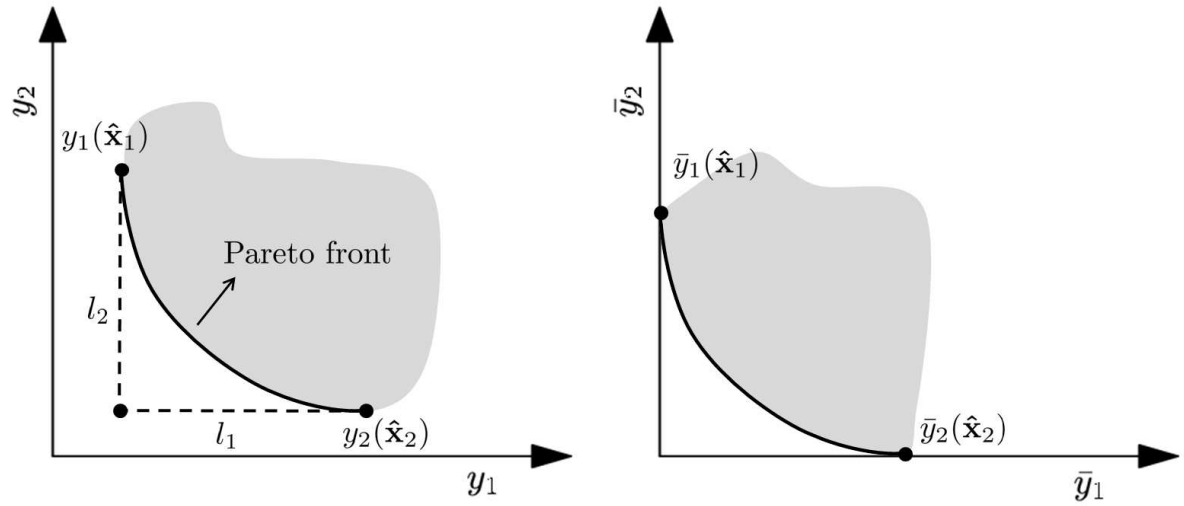
$$y_2(\mathbf{x}) = \beta_2 + \mathbf{r}_2(\mathbf{x})^T \mathbf{w}_2, \tag{D.2}$$

where $\mathbf{r}_1(\mathbf{x})$ and $\mathbf{r}_2(\mathbf{x})$ are the corresponding correlation vectors between the point $\mathbf{x}$ and the sampled points, respectively, $\mathbf{w_1}$ and $\mathbf{w_2}$ are the associated weight vectors (see equation 3.38), and β_1 and β_2 are the corresponding hyperparameters for the Kriging models.

The algorithm begins by finding, just like the normalised normal constraint method, the individual minima for the two functions. These can be found by any global optimisaton algorithm, including the KNRS algorithm. Let the points $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$, represent the individual minima for the two Kriging predictors, $y_1(\mathbf{x})$ and $y_2(\mathbf{x})$, respectively. These points are the extreme points on the Pareto front as shown in figure D.1a. The objective space is then normalised with respect to the distance l_1 and l_2 (shown in figure D.1a) as

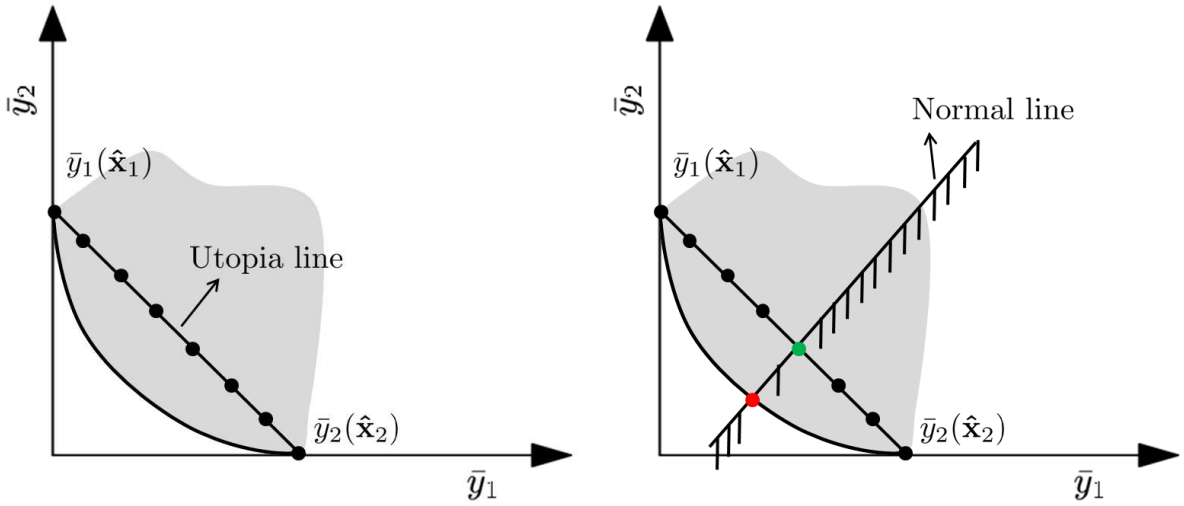
$$l_1 = y_1(\hat{\mathbf{x}}_2) - y_1(\hat{\mathbf{x}}_1), \tag{D.3}$$

$$l_2 = y_2(\hat{\mathbf{x}}_1) - y_2(\hat{\mathbf{x}}_2), \tag{D.4}$$



(a) Pareto front of the objective functions ($y_1(\mathbf{x})$ and $y_2(\mathbf{x})$) and their respective individual minimums

(b) The normalised space of the objective functions ($\bar{y}_1(\mathbf{x})$ and $\bar{y}_2(\mathbf{x})$)



(c) The utopia line in the normalised space and points on the utopia line

(d) The normal line from a point (shown in green) on the utopia line and the corresponding feasible space for constrained optimisation problem; the optimal solution is shown in red

Figure D.1: Steps in the normalised normal constrained method [17] which forms the basis of M-KNRS method

$$\bar{y}_1(\mathbf{x}) = \frac{y_1(\mathbf{x}) - y_1(\hat{\mathbf{x}}_1)}{l_1}, \tag{D.5}$$

$$\bar{y}_2(\mathbf{x}) = \frac{y_2(\mathbf{x}) - y_2(\hat{\mathbf{x}}_2)}{l_2}, \tag{D.6}$$

where $\bar{y}_1(\mathbf{x})$ and $\bar{y}_2(\mathbf{x})$ represent the normalised objective space for $y_1(\mathbf{x})$ and $y_2(\mathbf{x})$, respectively (see figure D.1b). For mathematical convenience $y_1(\hat{\mathbf{x}}_1)$ and $y_2(\hat{\mathbf{x}}_2)$ are written as z_1 and z_2, respectively,that is

$$z_1 = y_1(\hat{\mathbf{x}}_1), \tag{D.7}$$

$$z_2 = y_2(\hat{\mathbf{x}}_2). \tag{D.8}$$

$$\tag{D.9}$$

Consequently,

$$\bar{y}_1(\mathbf{x}) = \frac{y_1(\mathbf{x}) - z_1}{l_1}, \tag{D.10}$$

$$\bar{y}_2(\mathbf{x}) = \frac{y_2(\mathbf{x}) - z_2}{l_2}. \tag{D.11}$$

The *utopial line* is referred as the line which joins the individual minima of the two objective functions in the normalised objective space. This utopia line is shown in figure D.1c. Messac et. al. proposed generating uniformly distributed points in this line, as shown in figure D.1c and solving a single objective constrained optimisation problem from each of the generated point. The constrained optimisation problem, minimises one objective subject to all other objectives formulated as constraints. The constraints are formulated using a *normal line* from the chosen point on the utopia line. Figure D.1d shows this constraint in the bi-objective case, where the normal line is generated from one chosen point (shown in green). This normal line when posed as a constraint reduces the feasible space. A minimisation of one objective, $\bar{y}_2(\mathbf{x})$, in this reduces feasible space leads to a point on the Pareto front, shown in red in figure D.1d. This constraint when formulated by choosing a different point on the utopia line, and hence a different normal line, leads to different points on the Pareto front. In this way, the uniformity of points on the Pareto front is determined by the uniformity of points chosen on the utopia line.

The normal line vector, $\mathbf{N}$, is defined as

$$\mathbf{N} = \begin{bmatrix} n_1 \\ n_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}. \tag{D.12}$$

Let the starting point on the utopia line be $\mathbf{s}$, defined in terms of its components as

$$\mathbf{s} = [s_1, s_2]. \tag{D.13}$$

The single objective constrained minimisation problem then becomes

$$\text{Minimise} \quad \bar{y}_2(\mathbf{x}) \tag{D.14}$$

$$\text{subject to} \quad \mathbf{g}(\mathbf{x}) \leq 0 \tag{D.15}$$

$$\text{where} \quad \mathbf{g}(\mathbf{x}) = (\bar{y}_1 - s_1)n_1 + (\bar{y}_1 - s_2)n_1. \tag{D.16}$$

It should be noted that conceptually the constraint can also be posed as an equality constraint. The optimisation problem then becomes

$$\text{Minimise} \quad \bar{y}_2(\mathbf{x}) \tag{D.17}$$

$$\text{subject to} \quad \mathbf{h}(\mathbf{x}) = 0 \tag{D.18}$$

$$\text{where} \quad \mathbf{h}(\mathbf{x}) = (\bar{y}_1 - s_1)n_1 + (\bar{y}_1 - s_2)n_1. \tag{D.19}$$

Here, the latter equality constraint formulation is adopted for demonstration. From equations D.10, D.11, D.1, and D.2, $\bar{y}_1(\mathbf{x})$ and $\bar{y}_2(\mathbf{x})$ can be written as

$$\bar{y}_1(\mathbf{x}) = \frac{\beta_1 + \mathfrak{r}_1(\mathbf{x})^T \mathbf{w}_1 - z_1}{l_1}, \tag{D.20}$$

$$\bar{y}_2(\mathbf{x}) = \frac{\beta_2 + \mathfrak{r}_2(\mathbf{x})^T \mathbf{w}_2 - z_2}{l_2}. \tag{D.21}$$

The constraint equation, equation D.19, then becomes

$$\mathbf{h}(\mathbf{x}) = \left(\frac{\beta_1 + \mathfrak{r}_1(\mathbf{x})^T \mathbf{w}_1 - z_1}{l_1} - s_1\right) n_1 + \left(\frac{\beta_2 + \mathfrak{r}_2(\mathbf{x})^T \mathbf{w}_2 - z_2}{l_2} - s_2\right) n_1 = 0. \tag{D.22}$$

The Lagrangian for the equality constrained optimisation problem, equations D.17, D.18, and D.19, can be written as

$$\mathcal{L}(\lambda, \mathbf{x}) = \bar{y}_2 + \lambda \mathbf{h}(\mathbf{x}), \tag{D.23}$$

where λ is the lagrange multiplier for the equality constraint. Using equation D.22, the above can be written as

$$\mathcal{L}(\lambda, \mathbf{x}) = \left(\frac{\beta_2 + \mathfrak{r}_2(\mathbf{x})^T \mathbf{w}_2 - z_2}{l_2}\right) + \lambda \left[\left(\frac{\beta_1 + \mathfrak{r}_1(\mathbf{x})^T \mathbf{w}_1 - z_1}{l_1} - s_1\right) n_1 + \left(\frac{\beta_2 + \mathfrak{r}_2(\mathbf{x})^T \mathbf{w}_2 - z_2}{l_2} - s_2\right) n_1\right] \tag{D.24}$$

$$\mathcal{L}(\lambda, \mathbf{x}) = \left(\frac{1}{l_2} + \frac{\lambda n_2}{l_2}\right) \mathfrak{r}_2(\mathbf{x})^T \mathbf{w}_2 + \left(\frac{\lambda n_1}{l_1}\right) \mathfrak{r}_1(\mathbf{x})^T \mathbf{w}_1 + \lambda \left[\frac{\beta_1 - z_1 - s_1 l_1}{l_1} n_1 + \frac{\beta_2 - z_2 - s_2 l_1}{l_2} n_2\right] + \frac{\beta_2 - z_2}{l_2}. \tag{D.25}$$

The solution of the constrained optimisation is located at the stationary points of the Lagrangian. The stationary points can be obtained by equating the derivatives of the Lagrangian to zero, i.e.

$$\frac{\partial \mathcal{L}(\lambda, \mathbf{x})}{\partial \mathbf{x}} = 0 \tag{D.26}$$

$$\frac{\partial \mathcal{L}(\lambda, \mathbf{x})}{\partial \lambda} = 0. \tag{D.27}$$

Let us expand the i^{th} derivative of equation D.27

$$\frac{\partial \mathcal{L}(\lambda, \mathbf{x})}{\partial x_i} = -2\theta_{2_i} \left(\frac{1}{l_2} + \frac{\lambda n_2}{l_2}\right) \mathbf{D_{2_i}}(\mathbf{x}) \mathbf{Z_2}(\mathbf{x}) \mathbf{w_2} - 2\theta_{1_i} \left(\frac{\lambda n_1}{l_1}\right) \mathbf{D_{1_i}}(\mathbf{x}) \mathbf{Z_1}(\mathbf{x}) \mathbf{w_1}. \tag{D.28}$$

For mathematical convenience, m_1 and m_2 are defined as

$$m_1 = \frac{2\lambda m_1}{l_1}, \tag{D.29}$$

$$m_2 = -\frac{2}{l_2} - \frac{2\lambda n_2}{l_2}. \tag{D.30}$$

Equation D.28 then becomes

$$\frac{\partial \mathcal{L}(\lambda, \mathbf{x})}{\partial x_i} = 2\theta_{2_i} m_2 \mathbf{D_{2_i}}(\mathbf{x}) \mathbf{Z_2}(\mathbf{x}) \mathbf{w_2} - 2\theta_{1_i} m_1 \mathbf{D_{1_i}}(\mathbf{x}) \mathbf{Z_1}(\mathbf{x}) \mathbf{w_1}, \tag{D.31}$$

where, similar to equation 10.30 in chapter 10, $\mathbf{D_{1_i}(x)}$ and $\mathbf{D_{2_i}(x)}$, are defined as

$$\mathbf{D_{1_i}(x)} = x_i\mathbf{1_{p_1}} - \mathbf{X}_{1_i}, \tag{D.32}$$

$$\mathbf{D_{2_i}(x)} = x_i\mathbf{1_{p_2}} - \mathbf{X}_{2_i}, \tag{D.33}$$

where

$$\mathbf{1_{p_1}} = [1, 1, \ldots, 1] \in \mathbb{R}^{1\times p_1}, \tag{D.34}$$

$$\mathbf{1_{p_2}} = [1, 1, \ldots, 1] \in \mathbb{R}^{1\times p_2}, \tag{D.35}$$

and

$$\mathbf{X}_{1_i} = \begin{bmatrix} x_{1_i}^{(1)} & x_{1_i}^{(2)} & \cdots & x_{1_i}^{(p_1)} \end{bmatrix}, \tag{D.36}$$

$$\mathbf{X}_{2_i} = \begin{bmatrix} x_{2_i}^{(1)} & x_{2_i}^{(2)} & \cdots & x_{2_i}^{(p_2)} \end{bmatrix}. \tag{D.37}$$

Equation D.31 then becomes

$$\frac{\partial\mathcal{L}(\lambda, \mathbf{x})}{\partial x_i} = 2\theta_{2_i}m_2\left(x_i\mathbf{1_{p_1}} - \mathbf{X}_{1_i}\right)\mathbf{Z_2(x)w_2} - 2\theta_{1_i}m_1\left(x_i\mathbf{1_{p_2}} - \mathbf{X}_{2_i}\right)\mathbf{Z_1(x)w_1}, \tag{D.38}$$

where

$$\mathbf{Z_1(x)} = \begin{bmatrix} R(\mathbf{x}, \mathbf{x_1^{(1)}}) & 0 & \cdots & 0 \\ 0 & R(\mathbf{x}, \mathbf{x_1^{(2)}}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R(\mathbf{x}, \mathbf{x_1^{(p_1)}}) \end{bmatrix} \in \mathbb{R}^{p_1\times p_1}, \tag{D.39}$$

and

$$\mathbf{Z_2(x)} = \begin{bmatrix} R(\mathbf{x}, \mathbf{x_2^{(1)}}) & 0 & \cdots & 0 \\ 0 & R(\mathbf{x}, \mathbf{x_2^{(2)}}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R(\mathbf{x}, \mathbf{x_2^{(p_2)}}) \end{bmatrix} \in \mathbb{R}^{p_2\times p_2}. \tag{D.40}$$

Equating the above to zero to find the stationary point of the Lagrangian, which represent the solution to the constrained optimisation problem, yields

$$\frac{\partial\mathcal{L}(\lambda, \mathbf{x})}{\partial x_i} = 0 \tag{D.41}$$

$$2\theta_{2_i}m_2\left(x_i\mathbf{1_{p_1}} - \mathbf{X}_{1_i}\right)\mathbf{Z_2(x)w_2} - 2\theta_{1_i}m_1\left(x_i\mathbf{1_{p_2}} - \mathbf{X}_{2_i}\right)\mathbf{Z_1(x)w_1} = 0 \tag{D.42}$$

$$x_i = \frac{\theta_{2_i}m_2\mathbf{X}_{2_i}\mathbf{Z_2(x)w_2} - \theta_{1_i}m_1\mathbf{X}_{1_i}\mathbf{Z_1(x)w_1}}{\theta_{2_i}m_2\mathbf{1}_{p_2}\mathbf{Z_2(x)w_2} - \theta_{1_i}m_1\mathbf{1}_{p_1}\mathbf{Z_1(x)w_1}} \tag{D.43}$$

$$\boxed{x_i = \mathfrak{g}_i(\mathbf{x}, \lambda)}, \tag{D.44}$$

where

$$\mathfrak{g}_i(\mathbf{x}, \lambda) = \frac{\theta_{2_i}m_2\mathbf{X}_{2_i}\mathbf{Z_2(x)w_2} - \theta_{1_i}m_1\mathbf{X}_{1_i}\mathbf{Z_1(x)w_1}}{\theta_{2_i}m_2\mathbf{1}_{p_2}\mathbf{Z_2(x)w_2} - \theta_{1_i}m_1\mathbf{1}_{p_1}\mathbf{Z_1(x)w_1}} \tag{D.45}$$

or

$$\boxed{\mathfrak{F}_i((\mathbf{x},\lambda)) = x_i - \mathfrak{g}_i(\mathbf{x},\lambda) = 0}. \tag{D.46}$$

Similarly, equating the derivative of Lagrangian with respect to λ to zero yields

$$\frac{\partial \mathcal{L}(\lambda,\mathbf{x})}{\partial \lambda} = 0 \tag{D.47}$$

$$\frac{\partial \mathcal{L}(\lambda,\mathbf{x})}{\partial \lambda} = \left(\frac{n_2}{l_2}\right)\mathfrak{r}_2(\mathbf{x})^T\mathbf{w}_2 + \left(\frac{n_1}{l_1}\right)\mathfrak{r}_1(\mathbf{x})^T\mathbf{w}_1 + \left[\frac{\beta_1 - z_1 - s_1 l_1}{l_1}n_1 + \frac{\beta_2 - z_2 - s_2 l_1}{l_2}n_2\right] \tag{D.48}$$

or

$$\mathfrak{n}(\mathbf{x}) = \left(\frac{n_2}{l_2}\right)\mathfrak{r}_2(\mathbf{x})^T\mathbf{w}_2 + \left(\frac{n_1}{l_1}\right)\mathfrak{r}_1(\mathbf{x})^T\mathbf{w}_1 + \left[\frac{\beta_1 - z_1 - s_1 l_1}{l_1}n_1 + \frac{\beta_2 - z_2 - s_2 l_1}{l_2}n_2\right] \tag{D.49}$$

or

$$\boxed{\mathfrak{n}(\mathbf{x}) = 0}. \tag{D.50}$$

The system of equations represented by equations D.44 and D.50 represents a solution to the constrained optimisation problem. Observe that equation D.44 is a fixed point iterative sequence derived in chapter 10 (equation 10.37). Consequently, the methods proposed in chapter 10, viz. the SFP, MFP, and the NR sequences could potentially be employed to obtain a solution for these. However, the equation for λ, i.e. equation D.50, can not be represented in a form $\lambda = \mathfrak{f}(\mathbf{x},\lambda)$, where $\mathfrak{f}$ is some function of $\mathbf{x}$ and λ, primarily because equation D.50 does not have the term λ in it. If this were possible, then the set of equations represented by equation D.44 and $\lambda = \mathfrak{f}(\mathbf{x},\lambda)$ would form a simple fixed point iterative sequence (analogous to the SFP of chapter 10). Introducing λ as a pseudo-variable in equation D.50, for example by multiplying both sides of equation D.50 by λ or adding λ to both sides, to convert D.50 into the form of $\lambda = \mathfrak{f}(\mathbf{x},\lambda)$, as found by the author, is of no use and results in divergence of the fixed point iterative sequence. This apparent difficulty is resolved by thinking that equation D.50 represents an equation in $\mathbf{x}$, which in turn are dependent on λ through equation D.44. Hence, a Newton-Raphson sequence in λ can be formed using equation D.50 as

$$\lambda_{new} = \lambda_{old} - \frac{\mathfrak{n}(\mathbf{x})}{(\partial \mathfrak{n}(\mathbf{x})/\partial\lambda)}, \tag{D.51}$$

where the term $(\partial \mathfrak{n}(\mathbf{x})/\partial\lambda)$ can be evaluated using the chain rule as

$$\frac{\partial \mathfrak{n}(\mathbf{x})}{\partial \lambda} = \sum_{i=1}^{n} \frac{\partial \mathfrak{n}(\mathbf{x})}{\partial x_i}\frac{\partial x_i}{\partial \lambda}. \tag{D.52}$$

In the above summation, the partial derivatives of $\mathfrak{n}$ can be written as

$$\frac{\partial \mathfrak{n}(\mathbf{x})}{\partial x_i} = -2\theta_{1_i}\left(\frac{\lambda n_1}{l_1}\right)\mathbf{D_{1_i}(x)Z_1(x)w_1} - 2\theta_{2_i}\left(\frac{\lambda n_2}{l_2}\right)\mathbf{D_{2_i}(x)Z_2(x)w_2}, \tag{D.53}$$

and the partial derivative of x_i with respect to λ need to be calculated from equation D.44. These derivatives can be evaluated analytically using equation D.44, or estimated using a finite difference method. The results presented in this appendix employ a forward finite difference formulation. Another approach to solve the system of equations represented by D.46 and D.50 is to employ a Newton-Raphson scheme for all the equations. This results in more robust convergence in finding the Pareto front. With this background, the M-KNRS method, for a biobjective problem, can be written in the following steps

Inputs

1. The number of points required on the Pareto front: N_p
2. The number of starting points to solve the constrained optimisation problem that is used to generated one point in the Pareto front: N_s.
3. An upper bound on the Lagrange multiplier value: λ_{max}.

Begin M-KNRS

Step 1: Construct appropriate Krigs for the two objective functions involved.

Step 2: Find the individual global minima of the two objectives using KNRS or any other preferred algorithm.

Step 3: Using the individual minima normalise the objective functions and construct the utopia line in the normalised space.

Step 4: Generate N_p uniformly distributed points in the utopia line. This is the number of points required on the Pareto front. Let these points be represented by $\mathbf{s}^1, \mathbf{s}^2, \ldots, \mathbf{s}^{N_p}$

Step 5: Initialise the list containing the Pareto solutions, **PF** = Φ.

For each $\mathbf{s} \in [\mathbf{s}^1, \mathbf{s}^2, \ldots, \mathbf{s}^{N_p}]$

Inner Step 1: Initialise the list $P(\mathbf{s}) = \Phi$

Inner Step 2: Formulate the constrained optimisation problem, say minimise $\bar{y}_2(\mathbf{x})$ subject to $\mathbf{h(x)} = 0$, using $\mathbf{s}$ as the chosen point in the utopia line.

Inner Step 3: Generate SOBOL sequence of N_s points in the space of $[\mathbf{x}, \lambda]$, $\mathbf{x}$ varies between the design space bounds and λ varies between 10^{-06} and λ_{max}. Let this SOBOL sequence be represented by $\mathbf{S_1}, \mathbf{S_2}, \ldots, \mathbf{S_{N_s}}$. Set $\kappa = 1$;

Inner Step 4: Starting from $[(\mathbf{x}_0, \lambda_0)] = \mathbf{S}_\kappa$, use equation D.44, i.e. $x_{new} = \mathfrak{g}_i(\mathbf{x}_{old}, \lambda)$, and equation D.51 to update the values of $\mathbf{x}$ and λ, respectively. Repeat this update procedure, from the new values, until the solution converges or is declared diverging.

Inner Step 5: If the above iteration converged then store then append the converged solution to the list $P(\mathbf{s})$

Inner Step 6: Set $\kappa = \kappa + 1$; If $\kappa \leq N_s$ go to **Inner Step 7**; otherwise go to **Inner Step 4**

Inner Step 7: Find the solution in $P(\mathbf{s})$ for which the objective minimised in the constrained optimisation problem, i.e. $\bar{y}_2(\mathbf{x})$ has the minimum value. Append this solution to the list **PF**.

End for loop

Step 6: Return the list **PF** as the list containing Pareto solutions.

End M-KNRS

As mentioned before, the workings of the for loop can be changed in the above algorithm to solve equations D.44 and D.50 using a combined Newton-Raphson iteration.

D.2 Application of M-KNRS on test functions

The performance of M-KNRS on a few test functions is tested in this section. The test functions chosen are

1. Schaffer's function (SCH) [74, 232] defined as

$$f_1(x) = x^2 \tag{D.54}$$
$$f_2(x) = (x-2)^2 \tag{D.55}$$
$$x \in [-10^3, 10^3].$$

   The SCH function has a convex Pareto front. The optimal solutions lie in $[0, 2]$.

2. Two variable Fonseca and Fleming's (FON) function [74, 233] defined as

$$f_1(x) = 1 - \exp\left(-\sum_{i=1}^{2}\left(x_i - \frac{1}{\sqrt{3}}\right)^2\right) \tag{D.56}$$
$$f_2(x) = 1 - \exp\left(-\sum_{i=1}^{2}\left(x_i + \frac{1}{\sqrt{3}}\right)^2\right) \tag{D.57}$$
$$x \in [-4, 4].$$

   The FON function has a non-convex Pareto front. The optimal solutions lie in $x_1, x_2, x_3 \in [-\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}]$.

3. Poloni's function (POL) [74, 234] defined as

$$f_1(x) = 1 + (A_1 - B_1)^2 + (A_2 - B_2)^2 \tag{D.58}$$
$$f_2(x) = (x_1 + 3)^2 + (x_2 + 1)^2 \tag{D.59}$$
$$A_1 = 0.5\sin 1 - 2\cos 1 + \sin 2 - 1.5\cos 2 \tag{D.60}$$
$$A_2 = 1.5\sin 1 - \cos 1 + 2\sin 2 - 0.5\cos 2 \tag{D.61}$$
$$B_1 = 0.5\sin x_1 - 2\cos x_1 + \sin x_2 - 1.5\cos x_2 \tag{D.62}$$
$$B_2 = 0.5\sin x_1 - \cos x_1 + 2\sin x_2 - 0.5\cos x_2 \tag{D.63}$$
$$x \in [-\pi, \pi].$$

   The POL function has a non-convex disconnected Pareto front.

Figures D.2 and D.3 show, respectively, the feasible objective space and the results of the M-KNRS algorithm on the SCH function. In the latter figure, the green circles show the points on the utopia line (transformed back to original space from the normalised space), the red circles show the soultion(s) for the inner constrained optimisation problem, and the arrows point from the point on the utopia line used to form the constrained optimisation problem to the corresponding solution(s). It can be seen from this figure that for the relatively easy SCH function, a well distributed set of points on the convex Pareto front are found.

Figures D.4 and D.5 show, respectively, the feasible objective space and the results of the M-KNRS algorithm on the FON function. In the latter figure, for the inner constrained optimisation problem from each point on the utopia line, all solutions (as opposed to only the minimum) are shown. This is done to understand the workings of M-KNRS, so that the local Pareto fronts, found by the algorithm can also be seen. For the FON function, it can be seen that the non-convex Pareto front is found by the M-KNRS algorithm with a well distributed set of points. It can also be seen that one solution of the constrained optimisation problem, converged to top-right end of the feasible space. This is not unreasonable, as the M-KNRS algorithm tries, depending on the parameter N_s, to converge to all the stationary points of the Lagrangian. These stationary points can be either a minimum, maximum, or a saddle point, one of the saddle points representing the real solution to the constrained optimisation problem.

Figures D.6 and D.7 show, respectively, the feasible objective space and the results of the M-KNRS algorithm on the POL function. It can be seen from these figures that the M-KNRS algorithm does not have difficulty in finding the Pareto front, non-convex and disconnected, for the PON problem. Figure D.7 also shows the various local Pareto fronts that the inner loop of the algorithm converged to. Of particular interest are the solutions which lie outside the constraint bounds shown in this figure. It is well known [36] that for the POL problem, the existence of the Pareto front labeled 'Region A' is purely due to limiting the decision variable bounds to $[-\pi, \pi]$. In particular, if the first decision variable is allowed to take values below $-\pi$, then 'Region A' does not form part of the global Pareto front. This is because, the solutions that are marked 'solutions outside bounds' shown in figure D.7 become feasible, thereby dominating the solutions of 'Region A'. The M-KNRS algorithm, similar to the KNRS algorithm, in its current form is not designed to honour variable bounds. This, although not a problem, for the POL function, as both fronts are found, is a limitation of the M-KNRS algorithm. Thus, similar to KNRS algorithm, the M-KNRS algorithm needs modification so that the constraints, including variable bounds, can be satisfied.

D.3 Conclusions

In this appendix, an extension of the KNRS algorithm for multiobjective optimisation, the M-KNRS algorithm, is proposed. Although the algorithm is tested for three multiobjective test functions, and its performance is deemed convincing, further investigation of its performance on problems with both higher number of decision variables and higher number of objective functions is needed. Furthermore, it is identified that the M-KNRS algorithm has a limitation of converging to solutions that lie outside the variable bounds similar to the KNRS algorithm. This issue needs to be addressed for further development of the algorithm.

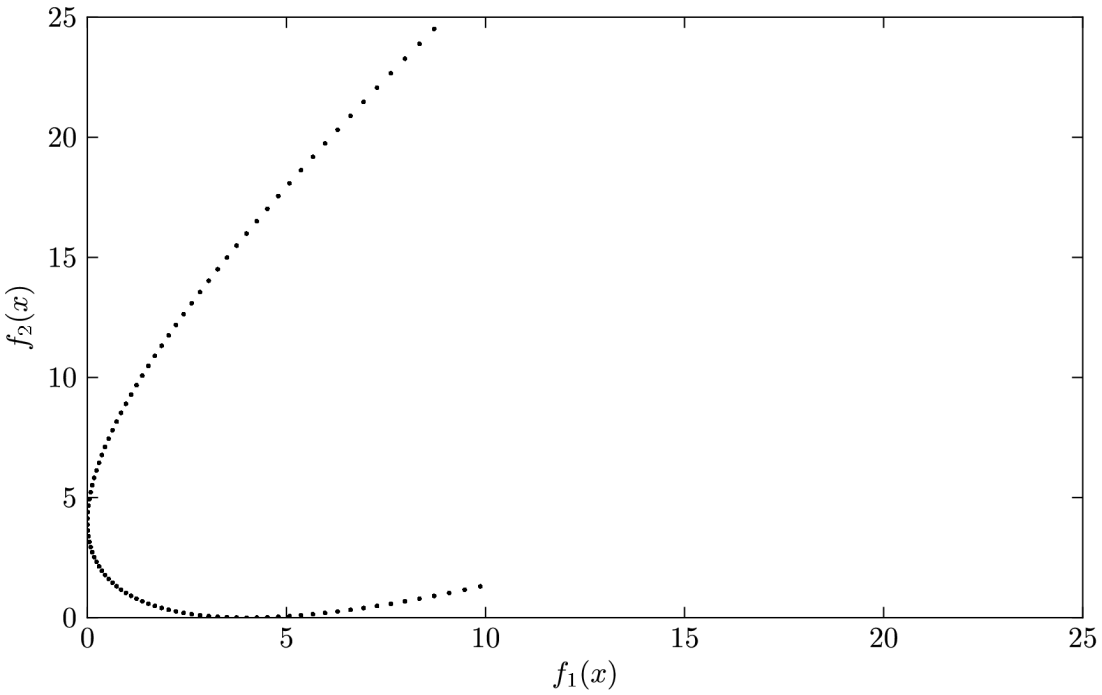


Figure D.2: Feasible space for the SCH function

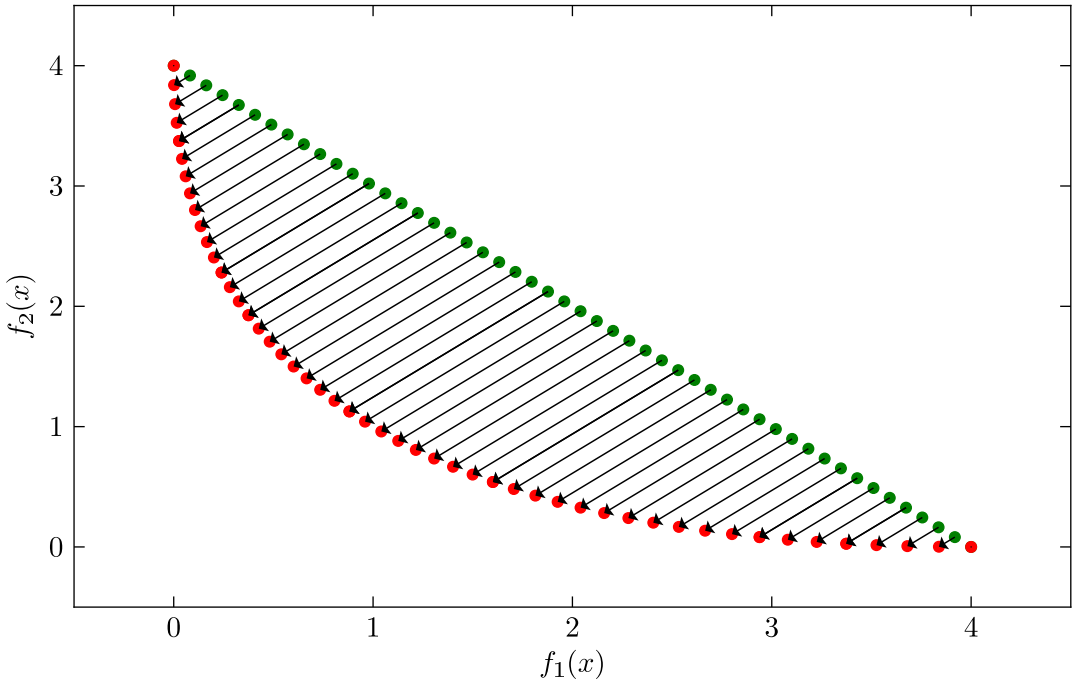


Figure D.3: M-KNRS algorithm on SCH function; $N_p = 50$, $N_s = 5$, $\lambda_{max} = 1.0$; green points show the uniformly distributed points in the utopia line and red points show the corresponding solution(s) for the constrained optimisation problem

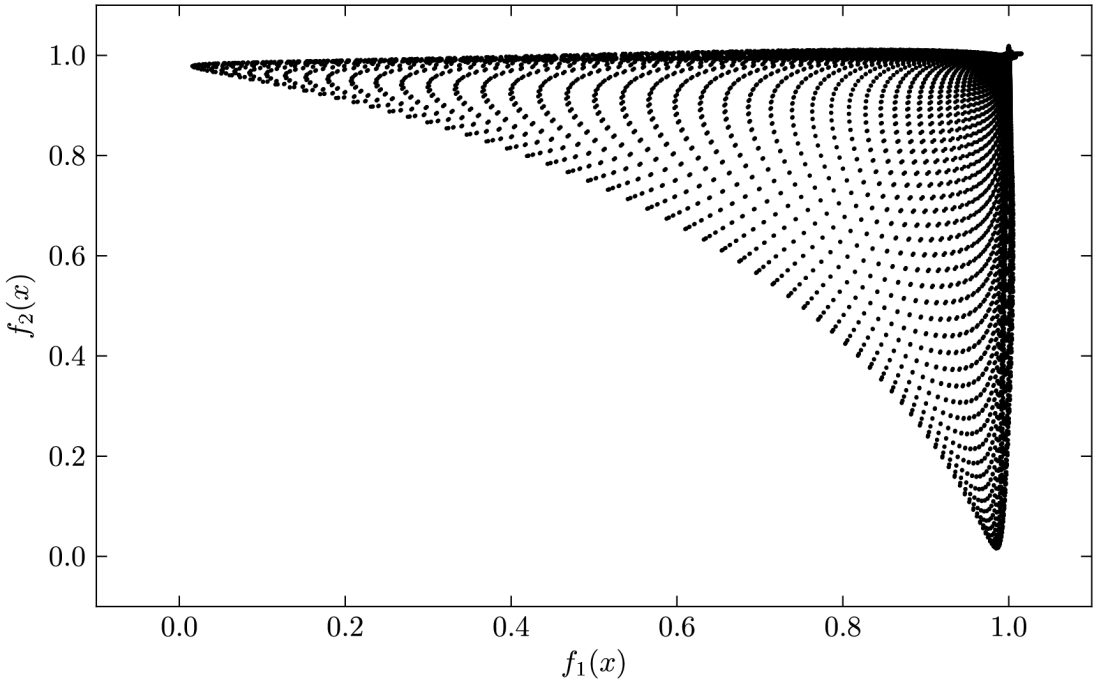


Figure D.4: Feasible space for the FON function

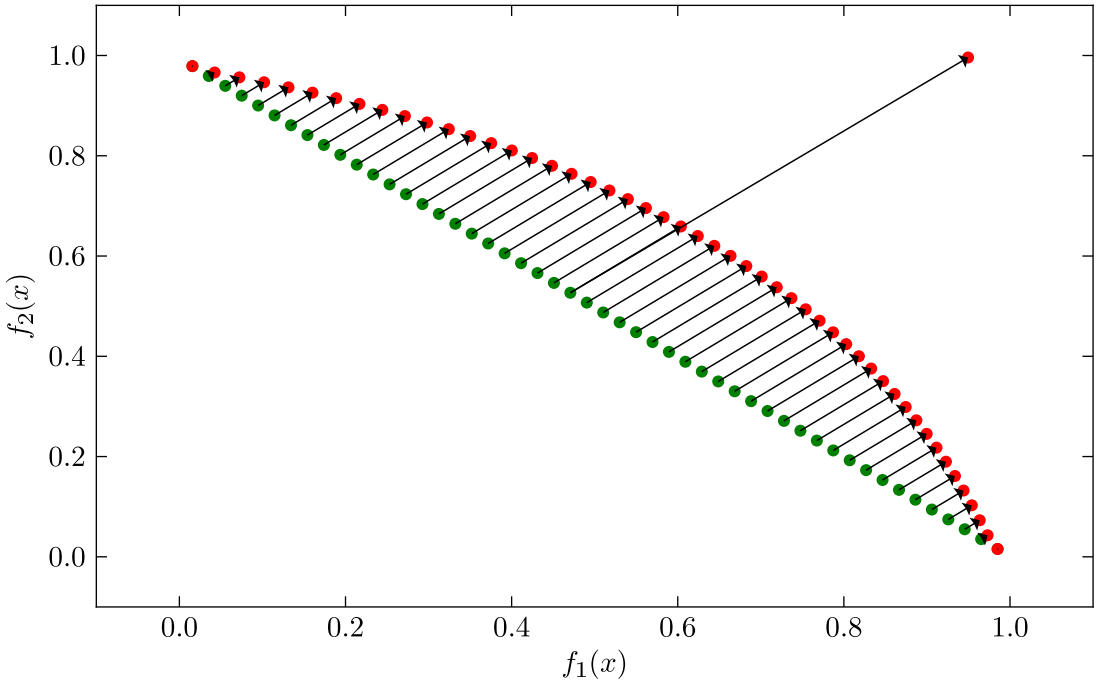


Figure D.5: M-KNRS algorithm on FON function; $N_p = 50$, $N_s = 5$, $\lambda_{max} = 1.0$; green points show the uniformly distributed points in the utopia line and red points show the corresponding solution(s) for the constrained optimisation problem

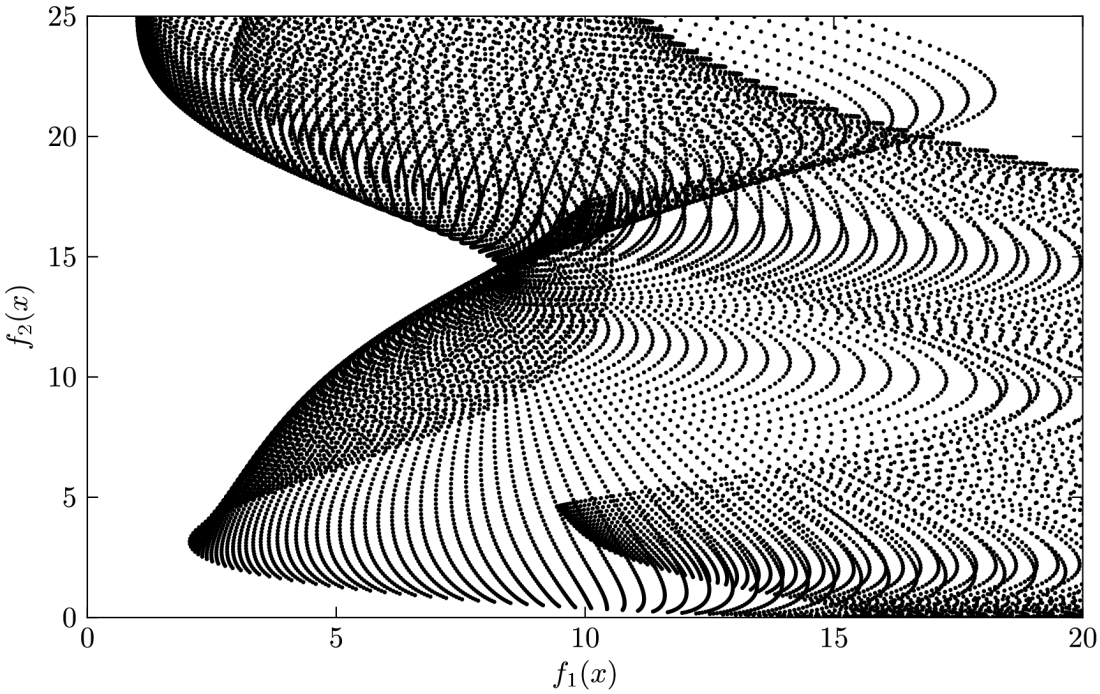


Figure D.6: Feasible space for the POL function

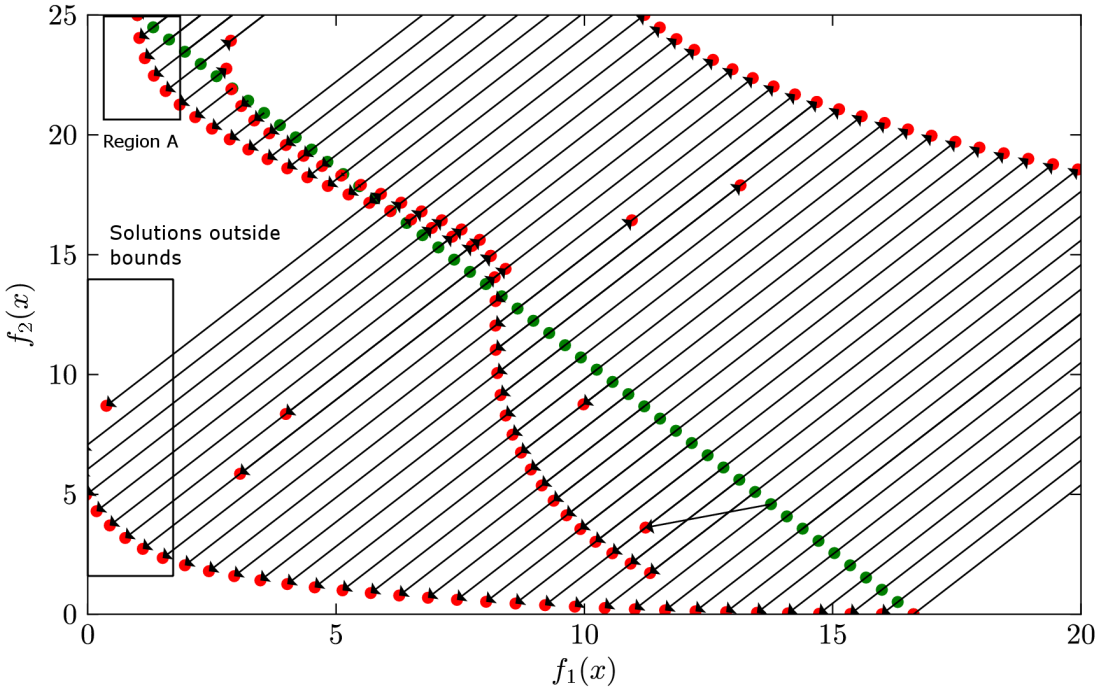


Figure D.7: M-KNRS algorithm on POL function; $N_p = 50$, $N_s = 50$, $\lambda_{max} = 1.0$; green points show the uniformly distributed points in the utopia line and red points show the corresponding solution(s) for the constrained optimisation problem

References

[1] A.M.R. Agur and A.F. Dalley. *Grant's atlas of anatomy*. Lippincott Williams & Wilkins and Walters Kluwer business, 13 edition, 2012. [cited at p. ix, 8]

[2] P.J. Lynch and C.C. Jaffe. Advanced instructional media medical illustrations, 2010. School of medicine, Yale University, http://commons.wikimedia.org/wiki/File:Heart_coronary_artery_lesion.jpg (last accessed June 17, 2010). [cited at p. ix, 8]

[3] A. Colombo and G. Stankovic. *Colombo's Tips & Tricks with Drug-Eluting Stents*. Taylor and Francis Group, 2005. [cited at p. ix, 9]

[4] A.J. Lusis. Atherosclerosis. *Nature*, 407(6801):233–241, 2000. [cited at p. ix, 7, 9]

[5] D. Faxon. *Restenosis: a guide to therapy*. Informa Health Care, illustrated edition, 2001. [cited at p. ix, 11, 12]

[6] D. Stoeckel, C. Bonsignore, and S. Duda. A survey of stent designs. *Minimally Invasive Therapy & Allied Technologies*, 11:137–47, 2002. [cited at p. ix, 12, 14, 15]

[7] Various websites for informationon stent images.
http://www.bostonscientific.com ; last accessed June 17, 2010
http://www.bostonscientific-international.com; last accessed June 28, 2012
http://www.abbottvascular.com/ ; last accessed June 17, 2010
http://www.xiencev.com/ ; last accessed June 17, 2010
http://www.medtronic.com/; last accessed June 28, 2012
http://www.cypherstent.com/; last accessed June 29, 2012
http://www.biosensors.com/intl/;last accessed June 29, 2012. [cited at p. ix, xi, 17, 132, 133]

[8] Art-stent website. http://www.art-stent.com/ ; last accessed June 17, 2010. [cited at p. ix, 17, 52]

[9] NASA. 'Borg' Computer Collective Designs NASA Space Antenna. News release, World Wide Web, February 2006. http://www.nasa.gov/centers/ames/research/exploringtheuniverse/borg.html; last accessed 11 June, 2012. [cited at p. ix, 24]

[10] B. P. Murphy, P. Savage, P. E. McHugh, and D. F. Quinn. The stress–strain behavior of coronary stent struts is size dependent. *Annals of Biomedical Engineering*, 31(6):686–691, 2003. [cited at p. x, 77, 79, 80, 94]

[11] M. De Beule. *Finite Element Stent Design*. PhD thesis, University of Gent, 2008. [cited at p. xi, 5, 77, 78, 104, 148]

[12] R. Fattori and T. Piva. Drug-eluting stents in vascular intervention. *The Lancet*, 361(9353):247–249, 2003. [cited at p. xii, 140, 141]

[13] M. De La Maza and D. Yuret. Dynamic hill climbing. *AI expert*, 9:26–26, 1994. [cited at p. xiv, 210, 212, 213, 235]

[14] A. Kumar. Multimodal and multiobjective optimization using real-coded genetic algorithms. Master's thesis, Kanpur, India: Indian Institute of Technology Kanpur, 1996. [cited at p. xv, 234]

[15] K. Deb and A. Kumar. Real-coded genetic algorhithms with simulated binary crossover: Studies on multimodal and multiobjective problems. *Complex Systems*, 9(6):431–454, 1995. [cited at p. xv, 38, 232, 234, 235]

[16] V.B. Kolachalama, A.R. Tzafriri, D.Y. Arifin, and E.R. Edelman. Luminal flow patterns dictate arterial drug deposition in stent-based delivery. *Journal of Controlled Release*, 133(1):24–30, 2009. [cited at p. xv, 109, 142, 143, 255, 256]

[17] A. Messac, A. Ismail-Yahaya, and C.A. Mattson. The normalized normal constraint method for generating the pareto frontier. *Structural and multidisciplinary optimization*, 25(2):86–98, 2003. [cited at p. xv, 45, 261, 262]

[18] A. Keane and P.B. Nair. *Computational approaches for aerospace design: the pursuit of excellence*. John Wiley & Sons, Ltd., 2005. [cited at p. xvii, 25, 26, 27, 28, 32, 34, 37, 42, 47, 48, 188, 198]

[19] F. Gervaso, C. Capelli, L. Petrini, S. Lattanzio, L. Di Virgilio, and F. Migliavacca. On the effects of different strategies in modelling balloon-expandable stenting by means of finite element method. *Journal of biomechanics*, 41(6):1206–1212, 2008. [cited at p. xvii, 74, 75, 79, 80, 81, 82, 86, 89]

[20] P. Zunino, C. D'Angelo, L. Petrini, C. Vergara, C. Capelli, and F. Migliavacca. Numerical simulation of drug eluting coronary stents: Mechanics, fluid dynamics and drug release. *Computer Methods in Applied Mechanics and Engineering*, 198(45-46):3633–3644, 2009. [cited at p. xvii, 74, 79, 80, 86, 141, 142, 143]

[21] P. Scarborough, P. Bhatnagar, K. Wickramasinghe, K. Smolina, C. Mitchell, and M. Rayner. *Coronary heart disease statistics*. British Heart Foundation Statistics Database (www.heartstats.org), British Heart Foundation Health Promotion Research Group and Department of Public Health, University of Oxford, 2010 edition, October 2010. [cited at p. 7]

[22] D. Lloyd-Jones, R.J. Adams, T.M. Brown, M. Carnethon, S. Dai, G. De Simone, T.B. Ferguson, E. Ford, K. Furie, C. Gillespie, et al. Heart Disease and Stroke Statistics–2010 Update. A Report From the American Heart Association. *Circulation*, 121:e46–e215, 2010. [cited at p. 7]

[23] Patient Information Guide, Abbott Vascular. http://www.xiencev.com/documents/patient_guide_int_EN.pdf; last accessed June 17, 2010. [cited at p. 8, 9]

[24] J.D. Humphrey. Towards a theory of vascular growth and remodelling. *Mechanics of Biologocal Tissue*, pages 3–15, 2006. [cited at p. 9]

[25] British Heart Foundation. http://www.bhf.org.uk/heart-health/treatment/coronary-bypass-surgery.aspx ; last accessed 09 June, 2010. [cited at p. 10]

[26] National Heart Lung and Blood Institute (NLHBI), Bethesda, MD. http://www.nhlbi.nih.gov/health/health-topics/topics/angioplasty/howdone.html; last accessed: 09, June, 2012. [cited at p. 10]

[27] G. Dangas and V. Fuster. Management of restenosis after coronary intervention. *American Heart Journal*, 132(2):428–436, 1996. [cited at p. 11]

[28] G.D. Dangas, B.E. Claessen, A. Caixeta, E.A. Sanidas, G.S. Mintz, and R. Mehran. In-stent restenosis in the drug-eluting stent era. *Journal of the American College of Cardiology*, 56(23):1897–1907, 2010. [cited at p. 11]

[29] A. Caixeta, P. Généreux, G. Dangas, and R. Mehran. *In-stent restenosis in the drug-eluting stent era*, chapter 28. Oxford Textbook of Interventional Cardiology. Oxford University Press, 2010. [cited at p. 11]

[30] R. Balossino. *A computational study of minimally-invasive devices for the treatment of coronary disease*. PhD thesis, Politecnico di Milano, 2007. [cited at p. 11, 12]

[31] M.G. Rabbat, A.A. Bavry, D.L. Bhatt, and S.G. Ellis. Understanding and minimizing late thrombosis of drug-eluting stents. *Cleveland Clinic journal of medicine*, 74(2):129–136, 2007. [cited at p. 12]

[32] M.J. Eisenberg. Drug-eluting stents: some bare facts. *The Lancet*, 364:1466–1467, 2004. [cited at p. 12]

[33] D.A. Brown, E.W. Lee, C.T. Loh, and S.T. Kee. A new wave in treatment of vascular occlusive disease: Biodegradable stents - clinical experience and scientific principles. *Journal of Vascular and Interventional Radiology*, 20:315–325, 2009. [cited at p. 12]

[34] P.W. Serruys. *Handbook of Coronary Stents*. Martin Dunitz Publishers, fourth edition, 1997. [cited at p. 12, 52, 75, 76, 111]

[35] Johnson & Johnson company. Cordis announces discontinuation of nevo sirolimus-eluting coronary stent. Press release, June 2011. http://www.jnj.com/connect/news/all/cordis-announces-discontinuation-of-nevo-sirolimus-eluting-coronary-stent; last accessed: 16 June 2012. [cited at p. 16, 75]

[36] K. Deb. *Multi-objective optimization using evolutionary algorithms*, volume 16. Wiley, 2001. [cited at p. 23, 33, 37, 38, 44, 45, 210, 269]

[37] L.A. Schmit and B. Farshi. Some approximation concepts for structural synthesis. *AIAA Journal*, 12:692–699, 1974. [cited at p. 25]

[38] L.A. Schmit and H. Miura. Approximation concepts for efficient structural synthesis. *NASA*, (CR-2552), 1976. [cited at p. 25]

[39] G.E.P. Box and N.R. Draper. *Empirical model-building and response surfaces*. John Wiley & Sons, 1987. [cited at p. 26]

[40] R.H. Myers and D.C. Montgomery. *Response surface methodology: process and product optimization using designed experiments*. John Wiley & Sons Inc, 1995. [cited at p. 26]

[41] C.A. Micchelli. Interpolation of scattered data: distance matrices and conditionally positive definite functions. *Constructive Approximation*, 2(1):11–22, 1986. [cited at p. 27]

[42] B.P. Wang. Parameter optimization in multiquadric response surface approximations. *Structural and Multidisciplinary Optimization*, 26(3):219–223, 2004. [cited at p. 27]

[43] D.G. Krige. A statistical approach to some basic mine valuation problems on the witwatersrand. *Journal of the Chemical, Metallurgical and Mining Engineering Society of South Africa*, 52(6):119–139, 1951. [cited at p. 28]

[44] N. Cressie. The origins of kriging. *Mathematical Geology*, 22(3):239–252, 1990. [cited at p. 28]

[45] J. Sacks, W.J. Welch, T.J. Mitchell, and H.P. Wynn. Design and analysis of computer experiments. *Statistical science*, 4(4):409–423, 1989. [cited at p. 28, 32]

[46] T.J. Santner, B.J. Williams, and W. Notz. *The design and analysis of computer experiments*. Springer Verlag, 2003. [cited at p. 28, 151]

[47] D.J.C. MacKay. *Information theory, inference, and learning algorithms*. Cambridge University Press, 2003. [cited at p. 28]

[48] C.E. Rasmussen and C.K.I. Williams. Gaussian processes for machine learning. *The MIT Press, Cambridge, MA, USA*, 38:715–719, 2006. [cited at p. 28, 29]

[49] E. Parzen. *Stochastic processes*, volume 24. Society for Industrial Mathematics, 1999. [cited at p. 28, 180]

[50] D.R. Jones, M. Schonlau, and W.J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, 13(4):455–492, 1998. [cited at p. 32, 48, 118, 132, 152, 180]

[51] M. Schonlau. *Computer experiments and global optimization*. PhD thesis, University of Waterloo, Waterloo, Ontario, Canada, 1997. [cited at p. 32]

[52] V.B. Kolachalama, N.W. Bressloff, and P.B. Nair. Mining data from hemodynamic simulations via Bayesian emulation. *Biomedical Engineering Online*, 6(1), 2007. [cited at p. 32]

[53] I.M. Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271–280, 2001. [cited at p. 32]

[54] J. Nocedal and S.J. Wright. *Numerical optimization*. Springer verlag, 1999. [cited at p. 33, 35, 39, 40, 41, 43]

[55] K. Deb. *Optimization for engineering design: Algorithms and examples*. PHI Learning Pvt. Ltd., 2004. [cited at p. 34, 35, 36, 39, 41, 42]

[56] W.C. Davidon. Variable metric method for minimization. Technical Report ANL-599, AEC Research Development Report, 1959. [cited at p. 35]

[57] R. Fletcher and M.J.D. Powell. A rapidly convergent descent method for minimization. *The Computer Journal*, 6(2):163–168, 1963. [cited at p. 35]

[58] R. Fletcher and CM Reeves. Function minimization by conjugate gradients. *The computer journal*, 7(2):149–154, 1964. [cited at p. 36]

[59] R. Hooke and T.A. Jeeves. "direct search"solution of numerical and statistical problems. *Journal of the ACM (JACM)*, 8(2):212–229, 1961. [cited at p. 36]

[60] J.A. Nelder and R. Mead. A simplex method for function minimization. *The computer journal*, 7(4):308–313, 1965. [cited at p. 36]

[61] D.E. Goldberg. *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley, 1989. [cited at p. 37]

[62] K. Deb and R.B. Agrawal. Simulated binary crossover for continuous search space. *Complex Systems*, 1(9):115–148, 1994. [cited at p. 38]

[63] W.M. Spears. *The role of mutation and recombination in evolutionary algorithms*. PhD thesis, Fairfax, VA: George Mason University, 1998. [cited at p. 38]

[64] K. Deb and M. Goyal. A combined genetic adaptive search (geneas) for engineering design. *Computer Science and Informatics*, 26:30–45, 1996. [cited at p. 38]

[65] G.B. Dantzig. *Linear programming and extensions*. Princeton University Press, 1998. [cited at p. 39, 42]

[66] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorics*, 4:373–395, 1984. [cited at p. 39]

[67] G. Zoutendijk. *Methods of feasible directions*. Elsevier, Amsterdam, 1960. [cited at p. 42]

[68] P.T. Boggs and J.W. Tolle. Sequential quadratic programming. *Acta numerica*, 4(1):1–51, 1995. [cited at p. 43]

[69] Y.Y. Haimes, L.S. Lasdon, and D.A. Wismer. On a bicriterion formulation of the problems of integrated system identification and system optimization. *IEEE Transactions on Systems, Man, and Cybernetics*, 1(3):296–297, 1971. [cited at p. 44]

[70] I. Das and J.E. Dennis. Normal-boundary intersection: a new method for generating the pareto surface in nonlinear multicriteria optimization problems. *SIAM Journal on Optimization*, 8(3):631–657, 1998. [cited at p. 44]

[71] A. Ismail-Yahaya and A. Messac. Effective generation of the pareto frontier using the normal constraint method. In *AIAA 40th Aerospace Sciences Meeting and Exhibit*, 2002. [cited at p. 44]

[72] 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. [cited at p. 45]

[73] P.K. Shukla and K. Deb. On finding multiple pareto-optimal solutions using classical and evolutionary generating methods. *European Journal of Operational Research*, 181(3):1630–1652, 2007. [cited at p. 45]

[74] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *Evolutionary Computation, IEEE Transactions on*, 6(2):182–197, 2002. [cited at p. 45, 46, 151, 268]

[75] E. Zitzler, M. Laumanns, and L. Thiele. SPEA2: Improving the strength pareto evolutionary algorithm. Technical report 103, Eidgenössische Technische Hochschule Zürich (ETH), Institut für Technische Informatik und Kommunikationsnetze (TIK), Zurich, 2001. [cited at p. 45]

[76] D. Corne, J. Knowles, and M. Oates. The pareto envelope-based selection algorithm for multiobjective optimization. In *Parallel Problem Solving from Nature PPSN VI*, pages 839–848. Springer, 2000. [cited at p. 45]

[77] E. Zitzler and L. Thiele. Multiobjective evolutionary algorithms: A comparative case study and the strength pareto approach. *Evolutionary Computation, IEEE Transactions on*, 3(4):257–271, 1999. [cited at p. 45]

[78] N. Srinivas and K. Deb. Muiltiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary computation*, 2(3):221–248, 1994. [cited at p. 46]

[79] A. Forrester, A. Sobester, and A. Keane. *Engineering design via surrogate modelling: a practical guide*. John-Wiley and Sons, 2008. [cited at p. 48]

[80] A. Kastrati, J. Mehilli, J. Dirschinger, J. Pache, K. Ulm, H. Schühlen, M. Seyfarth, C. Schmitt, R. Blasini, F.J. Neumann, and A. Schömig. Restenosis after coronary placement of various stent types. *American Journal of Cardiology*, 87:34–39, 2001. [cited at p. 50]

[81] C. Rogers and E.R. Edelman. Endovascular stent design dictates experimental restenosis and thrombosis. *Circulation*, 91:2995–3001, 1995. [cited at p. 50, 72, 96]

[82] A. Kastrati, J. Mehilli, J. Dirschinger, F. Dotzer, H. Schuhlen, F.J. Neumann, M. Fleckenstein, C. Pfafferott, M. Seyfarth, and A. Schömig. Intracoronary stenting and angiographic results: Strut thickness effect on restenosis outcome (isar-stereo) trial. *Circulation*, 103:2816–2821, 2001. [cited at p. 50, 72]

[83] D.N. Ku. Blood flow in arteries. *Annual Review of Fluid Mechanics*, 29:399–434, 1997. [cited at p. 50, 59]

[84] D.N. Ku, C.K. Zarins, D.P. Giddens, and S. Glagov. Pulsatile flow and atherosclerosis in the human carotid bifurcation: positve correlation between plaque localization and low and oscillating shear stress. *Arteriosclerosis*, 5:292–302, 1985. [cited at p. 50]

[85] J.E. Moore Jr., C. Xu, S. Glagov, C.K. Zarins, and D.N. Ku. Fluid wall shear stress measurements in a model of the human abdominal aorta: oscillatory behavior and relationship to atherosclerosis. *Atherosclerosis*, 110(2):225–40, 1994. [cited at p. 50]

[86] J.J Wentzel, R. Krams, J.C.H Schuurbiers, J.A Oomen, J. Kloet, W.J.V.D. Giessen, P.W. Serruys, and C.J. Slager. Relationship between neointimal thickness and shear stress after wallstent implantation in human coronary arteries. *Circulation*, 103:1740–1745, 2001. [cited at p. 50]

[87] J.L. Berry, A. Santamaria, J.E.Jr. Moore, S. Roychowdhury, and W.D. Routh. Experimental and computational flow evaluation of coronary stents. *Annals of Biomedical Engineering*, 28:386–98, 2000. [cited at p. 51, 52, 62, 65]

[88] J.F. LaDisa Jr., I. Guler, L.E. Olson, D.A. Hettrick, J.R. Kersten, D.C. Warltier, and P.S. Pagel. Three-dimensional computational fluid dynamics modeling of alterations in coronary wall shear stress produced by stent implantation. *Annals of biomedical Engineering*, 31:972–980, 2003. [cited at p. 51, 55, 59, 62]

[89] J.F. LaDisa Jr., G. Ismail, L.E. Olson, S.H. Audi, D.A. Hettrick, J.R. Kersten, D.C. Warltier, and P.S. Pagel. Stent design properties and deployment ratio influence indexes of wall shear stress: a three-dimensional computational fluid dynamics investigation within a normal artery. *Journal of Applied Physiology*, 97:424–430, 2004. [cited at p. 51, 62]

[90] D. Rajamohan, R.K. Banerjee, L.H. Back, A.A. Ibrahim, and M.A. Jog. Developing pulsatile flow in a deployed coronary stent. *Transactions of the ASME. Journal of Biomedical Engineering*, 128:347–359, 2006. [cited at p. 51, 55, 59, 62, 65]

[91] T. Seo, L.G. Schachter, and A.I. Barakat. Computational study of fluid mechanical disturbance induced by endovascular stents. *Annals of Biomedical Engineering*, 33:444–456, 2005. [cited at p. 51]

[92] I. Faik, R. Mongrain, R.L. Leask, J. Rodes-Cabau, E. Larose, and O. Bertrand. Time-dependent 3d simulations of the hemodynamics in a stented coronary artery. *Biomedical Materials*, 2:S28–S37, 2007. [cited at p. 51, 62]

[93] V. Dehlaghi, M.T. Shadpoor, and S. Najarian. Numerical analysis of pulsatile blood flow in a stented human coronary artery with a flow divider. *American Journal of Applied Sciences*, 4(6):397–404, 2007. [cited at p. 51]

[94] N. Benard, R. Perrault, and D. Coisne. Computational approach to estimating the effects of blood properties on changes in intra-stent flow. *Annals of Biomedical Engineering*, 34:1259–1271, 2006. [cited at p. 51]

[95] R. Balossino, F. Gervaso, F. Migliavacca, and G. Dubini. Effects of deifferent stent designs on local hemodynamics in stented arteries. *Journal of Biomechanics*, 41:1053–1061, 2008. [cited at p. 51, 59]

[96] J. Peacock, S. Hankins, T. Jones, and R. Lutz. Flow instabilities induced by coronary artery stents: assessment with an in vitro pulse duplicator. *Journal of Biomechanics*, 28(1):17–26, 1995. [cited at p. 51]

[97] N. Benard, D. Cosine, E. Donal, and R. Perrault. Experimental study of laminar blood flow through an artery treated by a stent implantation: characterisation of intra-stent wall shear stress. *Journal of Biomechanics*, 36:991–998, 2003. [cited at p. 52]

[98] N. DePaola, M.A. Gimbrone, P.F. Davies, and C.F. Dewey Jr. Vascular endothelium responds to fluid shear stress gradients. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 12:1254–1257, 1992. [cited at p. 52, 66]

[99] T. Nagel, N. Resnick, C.F. Dewey Jr., and M.A. Gimbrone. Vascular endothelial cells respond to spatial gradients in fluid shear stress by enhanced activation of transcription factors. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 19:1825–1834, 1999. [cited at p. 52]

[100] H.I. Yeh, S.K. Lu, T.Y. Tian, R.C. Hong, W.H. Lee, and C.H. Tsai. Comparison of endothelial cells grown on different stent materials. *Journal of Biomedical Materials Research Part A*, 76:835–841, 2006. [cited at p. 52]

[101] Biomatrix stent website. http://www.biomatrix.com/ ; last accessed June 17, 2010. [cited at p. 52]

[102] K. Perktold, M. Hofer, G. Rappitsch, M. Loew, B.D. Kuban, and M.H. Friedman. Validated computation of physiologic flow in a realistic coronary artery artery branch. *Journal of Biomechanics*, 31:217–28, 1998. [cited at p. 55]

[103] CD-adapco. *User guide, Star-CCM+ version 3.06.006.* CD-adapco, Melville, NY USA, 2008. [cited at p. 56, 57]

[104] F. S. Henry. *Flow in stented arteries. In: Vendonck P., Perktold K. (Ed.), Intra- and Extracorporeal Cardiovascular Fluid Dynamics.*, volume 2, pages 333–64. Southampton, UK: WIT Press, 2000. [cited at p. 59]

[105] J. M. Jimenez and P. F. Davies. Hemodynamically driven stent strut design. *Annals of Biomedical Engineering*, 37,7:1483–1494, 2009. [cited at p. 66]

[106] N.W. Bressloff. Parametric geometry exploration of the human carotid artery bifurcation. *Journal of Biomechanics*, 40:2483–2491, 2007. [cited at p. 68]

[107] A. Farb, G. Sangiorgi, A. J. Carter, V. M. Walley, W. D. Edwards, R. S. Schwartz, and R. Virmani. Pathology of acute and chronic coronary stenting in humans. *Circulation*, 99:44–52, 1999. [cited at p. 72, 96]

[108] R.S. Schwartz, K. C. Huber, W.D. Murphy, J. G.and Edwards, A. R. Camrud, R. R. Vlietstra, and D.R. Holmes. Restenosis and the proportional neointimal response to coronary artery injury: results in a porcine model. *Journal of the American College of Cardiology*, 19:267–274, 1992. [cited at p. 72, 96]

[109] A. J. Carter, J. R. Laird, A. Farb, W. Kufs, D. C. Wortham, and R. Virmani. Morphologic characteristics of lesion formation and time course of smooth muscle cell proliferation in a porcine proliferative restenosis model. *Journal of the American College of Cardiolog*, 24:1398–1405, 1994. [cited at p. 72, 96]

[110] A. König, T. M. Schiele, J. Rieber, K. Theisen, H. Mudra, and V. Klauss. Influence of stent design and deployment technique on neointima formation and vascular remodeling. *Zeitschrift fr Kardiologie*, 91:98–102, 2002. [cited at p. 72, 96]

[111] R. Hoffmann, G. S. Mintz, R. Mehran, K. M. Kent, A. D. Pichard, L. F. Satler, and M. B. Leon. Tissue proliferation within and surrounding palmaz-schatz stents is dependent on the aggressiveness of stent implanation technique. *American Journal of Cardiology*, 83:1170–1174, 1999. [cited at p. 72]

[112] R. Kornowski, M. K. Hong, F. O. Tio, O. Bramwell, H. Wu, and M. B. Leon. In-stent restenosis: Contributions of inflammatory responses and arterial injury to neointimal hyperplasia. *Journal of the American College of Cardiology*, 31:224–230, 1998. [cited at p. 72]

[113] J. Pache, A. Kastrati, J. Mehilli, H. Schuhlen, F. Dotzer, J. Hausleiter, M. Fleckenstein, F. Neumann, U. Sattelberger, C. Schmitt, M. Muller, J. Dirschinger, and A. Schomig. Intracoronary stenting and angiographic results: Strut thickness effect on restenosis outcome (isar-stereo-2) trial. *Journal of the American College of Cardiology*, 41(8):1283–1288, 2003. [cited at p. 72]

[114] A. C. Morton, D. Crossman, and Gunn. J. The influence of physical stent parameters upon restenosis. *Pathologie biologie*, 52:196–205, 2004. [cited at p. 72]

[115] M. D. Perry and R. T. Chang. Finite element analysis of NiTi alloy stent deployment. In *SMST-97: Proceedings of the Second International Conference on Shape Memory and Superelastic Technologies,(AR Pelton, D. Hodgson, S. Russell and TW Duerig eds.), Pacific Grove, California, USA*, pages 601–606, 1997. [cited at p. 73]

[116] F. Auricchio and R.L. Taylor. Shape-memory alloys: modelling and numerical simulations of the finite-strain superelastic behavior. *Computer Methods in Applied Mechanics and Engineering*, 143(1-2):175–194, 1997. [cited at p. 73]

[117] F. Trochu and P. Terriault. Finite element stress analysis of a shape memory medical stent. In *SMST-97: Proceedings of the Second International Conference on Shape Memory and Superelastic Technologies,(AR Pelton, D. Hodgson, S. Russell and TW Duerig eds.), Pacific Grove, California, USA*, pages 596–600, 1997. [cited at p. 73]

[118] F. Trochu and P. Terriault. Nonlinear modelling of hysteretic material laws by dual kriging and application. *Computer Methods in Applied Mechanics and Engineering*, 151(3-4):545–558, 1998. [cited at p. 73]

[119] C. Rogers, D.Y. Tseng, J.C. Squire, and E.R. Edelman. Balloon-artery interactions during stent placement: a finite element analysis approach to pressure, compliance, and stent design as contributors to vascular injury. *Circulation research*, 84(4):378–383, 1999. [cited at p. 73, 96]

[120] C. Dumoulin and B. Cochelin. Mechanical behaviour modelling of balloon-expandable stents. *Journal of Biomechanics*, 33(11):1461–1470, 2000. [cited at p. 73]

[121] F. Auricchio, M.D. Loreto, and E. Sacco. Finite-element analysis of a stenotic artery revascularization through a stent insertion. *Computer Methods in Biomechanics and Biomedical Engineering*, 4(3):249–263, 2001. [cited at p. 73, 79]

[122] F. Etave, G. Finet, M. Boivin, J.C. Boyer, G. Rioufol, and G. Thollet. Mechanical properties of coronary stents determined by using finite element analysis. *Journal of Biomechanics*, 34(8):1065–1075, 2001. [cited at p. 73, 79]

[123] F. Migliavacca, L. Petrini, M. Colombo, F. Auricchio, and R. Pietrabissa. Mechanical behavior of coronary stents investigated through the finite element method. *Journal of Biomechanics*, 35(6):803–811, 2002. [cited at p. 73, 79]

[124] L. Gu, S. Santra, R.A. Mericle, and A.V. Kumar. Finite element analysis of covered microstents. *Journal of biomechanics*, 38(6):1221–1227, 2005. [cited at p. 73, 79]

[125] G. A. Holzapfel, M. Stadler, and T.C. Gasser. Changes in the mechanical environment of stenotic arteries during interaction with stents: computational assessment of parametric stent designs. *Journal of biomechanical engineering*, 127:166–180, 2005. [cited at p. 73, 74, 96]

[126] F. Migliavacca, L. Petrini, V. Montanari, I. Quagliana, F. Auricchio, and G. Dubini. A predictive study of the mechanical behaviour of coronary stents by computer modelling. *Medical engineering & physics*, 27(1):13–18, 2005. [cited at p. 73, 79]

[127] G.A. Holzapfel, M. Stadler, and C.A.J. Schulze-Bauer. A layer-specific three-dimensional model for the simulation of balloon angioplasty using magnetic resonance imaging and mechanical testing. *Annals of Biomedical Engineering*, 30(6):753–767, 2002. [cited at p. 73, 74]

[128] D. K. Liang, D. Z. Yang, M. Qi, and W. Q. Wang. Finite element analysis of the implantation of a balloon-expandable stent in a stenosed artery. *International journal of cardiology*, 104(3):314–318, 2005. [cited at p. 74, 79]

[129] R.V. Marrey, R. Burgermeister, R.B. Grishaber, and R. O. Ritchie. Fatigue and life prediction for cobalt-chromium stents: A fracture mechanics analysis. *Biomaterials*, 27(9):1988–2000, 2006. [cited at p. 74, 99]

[130] W.Q. Wang, D.K. Liang, D.Z. Yang, and M. Qi. Analysis of the transient expansion behavior and design optimization of coronary stents by finite element method. *Journal of biomechanics*, 39(1):21–32, 2006. [cited at p. 74, 79, 108]

[131] K. Takashima, T. Kitou, K. Mori, and K. Ikeuchi. Simulation and experimental observation of contact conditions between stents and artery models. *Medical engineering & physics*, 29(3):326–335, 2007. [cited at p. 74, 79]

[132] W. Wu, W.Q. Wang, D.Z. Yang, and M. Qi. Stent expansion in curved vessel and their interactions: A finite element analysis. *Journal of biomechanics*, 40(11):2580–2585, 2007. [cited at p. 74, 79]

[133] F.J.H. Gijsen, F. Migliavacca, S. Schievano, L. Socci, L. Petrini, A. Thury, J.J. Wentzel, A.F.W. Van Der Steen, P.W.S. Serruys, and G. Dubini. Simulation of stent deployment

in a realistic human coronary artery. *BioMedical Engineering OnLine*, 7(1):23, 2008. [cited at p. 74, 79]

[134] M. De Beule, P. Mortier, S.G. Carlier, B. Verhegghe, R. Van Impe, and P. Verdonck. Realistic finite element-based stent design: The impact of balloon folding. *Journal of biomechanics*, 41(2):383–389, 2008. [cited at p. 74, 76, 77, 79, 81, 82, 83, 89]

[135] Cordis Corporation. *Insrtuctions for use: CYPHER Sirolimus-eluting Coronary Stent on RAPTOR Over-the-Wire Delivery System*. Johnson & Johnson company, Warren/Bridgewater Campus, 12 edition, March 2010. http://www.cordislabeling.com, last accessed 20th May, 2010, 1913 hrs. [cited at p. 75, 76, 77, 78, 79, 83, 86, 88, 149]

[136] R.M. Hicks and P.A. Henne. Wing design by numerical optimization. *Journal of Aircraft*, 15(7):407–412, 1978. [cited at p. 77]

[137] V.B. Kolachalama, N.W. Bressloff, P.B. Nair, C.P. Shearman, B.E. Center, and C. MIT. Predictive haemodynamics in a one-dimensional human carotid artery bifurcation. Part I: Application to stent design. *IEEE Trans. Biomed. Eng.*, 54(5):802–812, 2007. [cited at p. 77]

[138] J. P. McGarry, B. P. O'Donnell, P. E. McHugh, and J. G. McGarry. Analysis of the mechanical performance of a cardiovascular stent design based on micromechanical modelling. *Computational Materials Science*, 31(3-4):421–438, 2004. [cited at p. 77, 79]

[139] G.J. Hall and E.P. Kasper. Comparison of element technologies for modeling stent expansion. *Journal of biomechanical engineering*, 128:751–756, 2006. [cited at p. 79, 81]

[140] M. De Beule, R. Van Impe, B. Verhegghe, P. Segers, and P. Verdonck. Finite element analysis and stent design: Reduction of dogboning. *Technology and Health Care*, 14(4):233–241, 2006. [cited at p. 79]

[141] L. Petrini, F. Migliavacca, F. Auricchio, and G. Dubini. Numerical investigation of the intravascular coronary stent flexibility. *Journal of biomechanics*, 37(4):495–501, 2004. [cited at p. 79, 99, 100, 104]

[142] G.A. Holzapfel, G. Sommer, C.T. Gasser, and P. Regitnig. Determination of layer-specific mechanical properties of human coronary arteries with nonatherosclerotic intimal thickening and related constitutive modeling. *American Journal of Physiology-Heart and Circulatory Physiology*, 289(5):H2048–H2058, 2005. [cited at p. 80]

[143] G.A. Holzapfel, G. Sommer, and P. Regitnig. Anisotropic Mechanical Properties of Tissue Components in Human Atherosclerotic Plaques. *Journal of Biomechanical Engineering(Transactions of the ASME)*, 126(5):657–665, 2004. [cited at p. 80]

[144] H.C. Wong, K.N. Cho, and W.C. Tang. Bending of a stented atherosclerotic artery. In *Proceedings of the COMSOL Conference 2009 Boston*, 2009. [cited at p. 80, 81]

[145] S.A. Kock, J.V. Nygaard, N. Eldrup, E.T. Frund, A. Klærke, W.P. Paaske, E. Falk, and W. Yong Kim. Mechanical stresses in carotid plaques using MRI-based fluid–structure interaction models. *Journal of biomechanics*, 41(8):1651–1658, 2008. [cited at p. 80, 81]

[146] I. Pericevic, C. Lally, D. Toner, and D.J. Kelly. The influence of plaque composition on underlying arterial wall stress during stent expansion: The case for lesion-specific stents. *Medical engineering & physics*, 31(4):428–433, 2009. [cited at p. 81]

[147] D.E. Kiousis, T.C. Gasser, and G.A. Holzapfel. A numerical model to study the interaction of vascular stents with human atherosclerotic lesions. *Annals of Biomedical Engineering*, 35(11):1857–1869, 2007. [cited at p. 81]

[148] C. Lally, A.J. Reid, and P.J. Prendergast. Elastic behavior of porcine coronary artery tissue under uniaxial and equibiaxial tension. *Annals of biomedical engineering*, 32(10):1355–1364, 2004. [cited at p. 81]

[149] Dassault Systémes. *Abaqus 6.9.1 user manual*. Dassault Systèmes Simulia Corp., 2009. version 6.9.1. [cited at p. 82, 83, 86, 101]

[150] Engineers hand book. http://www.engineershandbook.com; last accessed 02 May, 2010. [cited at p. 83]

[151] J. Contiliano and Q. Zhang. Method of manufacturing a polymeric stent having a circumferential ring configuration, October 1 2009. WO Patent WO/2009/121048. [cited at p. 94, 110]

[152] S. Hur, A. Miyazawa, R. Sakurai, M. Yamasaki, Y. Hongo, W. Wijns, J. Fajadet, M.B. Leon, D. E. Kandzari, P. G. Yock, P. J. Fitzgerald, and Y. Honda. Impact of Final Stent Dimensions on Long-Term Stent Patency Following ABT-578 Eluting Phosphorylcholine Coated Cobalt Chromium Alloy Stent Implantation: A Serial 3-D IVUS Study. *Journal of American College of Cardiology*, 47(2B), 2006. [cited at p. 95]

[153] H. Otake, Y. Honda, and P.J. Fitzerald. *The role of intravascular ultrasound in percutaneous coronary intervention*, chapter 10. Oxford Textbook of Interventional Cardiology. Oxford University Press, 2010. [cited at p. 95]

[154] S. Kasaoka, J.M. Tobis, T. Akiyama, B. Reimers, C. Di Mario, N.D. Wong, and A. Colombo. Angiographic and intravascular ultrasound predictors of in-stent restenosis. *Journal of the American College of Cardiology*, 32:1630–1635, 1998. [cited at p. 95]

[155] Y. Ozaki, M. Okumura, T.F. Ismail, H. Naruse, K. Hattori, S. Kan, M. Ishikawa, T. Kawai, Y. Takagi, J. Ishii, et al. The fate of incomplete stent apposition with drug-eluting stents: an optical coherence tomography-based natural history study. *European heart journal*, 31(12):1470, 2010. [cited at p. 95]

[156] S. Cook, P. Wenaweser, M. Togni, M. Billinger, C. Morger, C. Seiler, R. Vogel, O. Hess, B. Meier, and S. Windecker. Incomplete stent apposition and very late stent thrombosis after drug-eluting stent implantation. *Circulation*, 115(18):2426–2434, 2007. [cited at p. 95]

[157] P. Szabaddíts, Z. Puskás, and J. Dobránszky. Flexibility and trackability of laser cut coronary stent systems. *Acta of Bioengineering and Biomechanics*, 11(3), 2009. [cited at p. 99]

[158] J.A. Ormiston, S.R. Dixon, M.W.I. Webster, P.N. Ruygrok, J.T. Stewart, I. Minchington, and T. West. Stent longitudinal flexibility: a comparison of 13 stent designs

before and after balloon expansion. *Catheterization and Cardiovascular Interventions*, 50(1):120–124, 2000. [cited at p. 99]

[159] K. Mori and T. Saito. Effects of stent structure on stent flexibility measurements. *Annals of biomedical engineering*, 33(6):733–742, 2005. [cited at p. 99]

[160] W. Wu, D.Z. Yang, M. Qi, and W.Q. Wang. An FEA method to study flexibility of expanded coronary stents. *Journal of Materials Processing Technology*, 184(1-3):447–450, 2007. [cited at p. 99, 100, 101]

[161] F. Ju, Z. Xia, and C. Zhou. Repeated unit cell (RUC) approach for pure bending analysis of coronary stents. *Computer Methods in Biomechanics and Biomedical Engineering*, 11(4):419–431, 2008. [cited at p. 100]

[162] J.A. Bryant, M.A. Atherton, and M.W. Collins. *Information Transfer in Biological Systems, Design in Nature Series*, volume 2. Southampton: WIT Press, 2006. [cited at p. 108]

[163] N. Li and Y. GU. Parametric design analysis and shape optimization of coronary arteries stent structure. In *6th World Congresses of Structural and Multidisciplinary Optimization, Rio de Janeiro, 30 May - 03 June, 2005, Brazil*. [cited at p. 108]

[164] N. Li, H. Zhanga, and H. Ouyang. Shape optimization of coronary artery stent based on a parametric model. *Finite Elements in Analysis and Design*, 45:468–475, 2009. [cited at p. 108]

[165] A. Blouza, L. Dumas, and I. M'Baye. Multiobjective optimization of a stent in a fluid-structure context. In *Proceedings of the 2008 GECCO conference companion on Genetic and evolutionary computation*, pages 2055–2060, 2008. [cited at p. 108]

[166] J. Bedoya, C.A. Meyer, L.H. Timmins, M.R. Moreno, and J.E. Moore Jr. Effects of stent design parameters on normal artery wall mechanics. *Journal of biomechanical engineering*, 128:757–765, 2006. [cited at p. 108]

[167] L.H. Timmins, M.R. Moreno, C.A. Meyer, J.C. Criscione, A. Rachev, and J.E. Moore. Stented artery biomechanics and device design optimization. *Medical and Biological Engineering and Computing*, 45(5):505–513, 2007. [cited at p. 108]

[168] W. Wu, L. Petrini, D. Gastaldi, T. Villa, M. Vedani, E. Lesma, B. Previtali, and F. Migliavacca. Finite Element Shape Optimization for Biodegradable Magnesium Alloy Stents. *Annals of Biomedical Engineering*, 2007. to appear. [cited at p. 108]

[169] J. Crank. *The mathematics of diffusion*. Oxford University Press, USA, 1979. [cited at p. 108, 143]

[170] D. R. Hose, A. J. Narracott, B. Griffiths, S. Mahmood, J. Gunn, D. Sweeney, and P. V. Lawford. A thermal analogy for modelling drug elution from cardiovascular stents. *Computer Methods in Biomechanics and Biomedical Engineering*, 7(5):257–264, 2004. [cited at p. 108, 141, 143]

[171] P.H. Feenstra and C.A. Taylor. Drug transport in artery walls: A sequential porohyperelastic-transport approach. *Computer Methods in Biomechanics and Biomedical Engineering*, 12(3):263–276, 2009. [cited at p. 108, 109, 143]

[172] R.B. Statnikov and J.B. Matusov. *Multicriteria analysis in engineering: using the PSI method with MOVI 1.0*. Kluwer Academic Pub, 2002. [cited at p. 115, 151]

[173] J.L. Zhou, A.L. Tits, and C.T. Lawrence. User's guide for FFSQP version 3.7: A FORTRAN code for solving constrained nonlinear (minimax) optimization problems, generating iterates satisfying all inequality and linear constraints. *Institute for Systems Research, University of Maryland, College Park, MD*, 1998. [cited at p. 116, 130]

[174] B.L. Van Der Hoeven, N.M.M. Pires, H.M. Warda, P.V. Oemrawsingh, B.J.M. van Vlijmen, P.H.A. Quax, M.J. Schalij, E.E. van der Wall, and J.W. Jukema. Drug-eluting stents: results, promises and problems. *International journal of cardiology*, 99(1):9–17, 2005. [cited at p. 140]

[175] A. Finkelstein, D. McClean, S. Kar, K. Takizawa, K. Varghese, N. Baek, K. Park, M.C. Fishbein, R. Makkar, F. Litvack, et al. Local drug delivery via a coronary stent with programmable release pharmacokinetics. *Circulation*, 107(5):777–784, 2003. [cited at p. 140]

[176] Cordis corporation, Johnson & Johnson, RES technology. Web-link. http://www.res-technology.com/ ; last accessed: 12 June, 2010. [cited at p. 140]

[177] C.W. Hwang, D. Wu, and E.R. Edelman. Physiological transport forces govern drug distribution for stent-based delivery. *Circulation*, 104(5):600–605, 2001. [cited at p. 140]

[178] M.A. Lovich, M. Philbrook, S. Sawyer, E. Weselcouch, and E.R. Edelman. Arterial heparin deposition: role of diffusion, convection, and extravascular space. *American Journal of Physiology- Heart and Circulatory Physiology*, 275(6):H2236, 1998. [cited at p. 140]

[179] D.V. Sakharov, L.V. Kalachev, and D.C. Rijken. Numerical simulation of local pharmacokinetics of a drug after intravascular delivery with an eluting stent. *Journal of drug targeting*, 10(6):507–513, 2002. [cited at p. 141]

[180] G. Pontrelli and F. de Monte. Modelling of mass convection-diffusion in stent-based drug-delivery. In *XXV Congresso Nazionale UIT sulla Trasmssione del Calore, Trieste*, pages 18–20, 2007. [cited at p. 141]

[181] P. Zunino. Multidimensional pharmacokinetic models applied to the design of drug-eluting stents. *Cardiovascular Engineering*, 4(2):181–191, 2004. [cited at p. 141]

[182] A. Borghi, E. Foa, R. Balossino, F. Migliavacca, and G. Dubini. Modelling drug elution from stents: effects of reversible binding in the vascular wall and degradable polymeric matrix. *Computer Methods in Biomechanics and Biomedical Engineering*, 11(4):367–377, 2008. [cited at p. 141]

[183] B. Balakrishnan, A.R. Tzafriri, P. Seifert, A. Groothuis, C. Rogers, and E.R. Edelman. Strut position, blood flow, and drug deposition: implications for single and overlapping drug-eluting stents. *Circulation*, 111(22):2958–2965, 2005. [cited at p. 141]

[184] F. Migliavacca, F. Gervaso, M. Prosi, P. Zunino, S. Minisini, L. Formaggia, and G. Dubini. Expansion and drug elution model of a coronary stent. *Computer Methods in Biomechanics and Biomedical Engineering*, 10(1):63–73, 2007. [cited at p. 141]

[185] V.B. Kolachalama, E.G. Levine, and E.R. Edelman. Luminal flow amplifies stent-based drug deposition in arterial. *PLoS ONE*, 4, 2009. [cited at p. 142]

[186] M.A. Lovich and E.R. Edelman. Mechanisms of transmural heparin transport in the rat abdominal aorta after local vascular delivery. *Circulation research*, 77(6):1143, 1995. [cited at p. 143]

[187] S. Jeong and S. Obayashi. Efficient global optimization (ego) for multi-objective problem and data mining. In *Evolutionary Computation, 2005. The 2005 IEEE Congress on*, volume 3, pages 2138–2145. IEEE, 2005. [cited at p. 152, 247]

[188] A.J. Keane. Statistical improvement criteria for use in multiobjective design optimization. *AIAA journal*, 44(4):879–891, 2006. [cited at p. 152, 210, 247]

[189] J. Knowles. ParEGO: A hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. *Evolutionary Computation, IEEE Transactions on*, 10(1):50–66, 2006. [cited at p. 152, 247]

[190] W. Liu, Q. Zhang, E. Tsang, C. Liu, and B. Virginas. On the performance of metamodel assisted MOEA/D. In L. Kang, Y. Liu, and S. Zeng, editors, *Advances in Computation and Intelligence*, volume 4683 of *Lecture Notes in Computer Science*, pages 547–557. Springer Berlin / Heidelberg, 2007. [cited at p. 152, 247]

[191] M. Emmerich and J. Klinkenberg. The computation of the expected improvement in dominated hypervolume of pareto front approximations. Technical report, Technical Report 4-2008 Leiden Institute of Advanced Computer Science, LIACS, 2008. [cited at p. 152, 247]

[192] W. Ponweiser, T. Wagner, D. Biermann, and M. Vincze. Multiobjective optimization on a limited budget of evaluations using model-assisted $\mathcal{S}$-metric selection. *Parallel Problem Solving from Nature–PPSN X*, pages 784–794, 2008. [cited at p. 152, 247]

[193] Q. Zhang, W. Liu, E. Tsang, and B. Virginas. Expensive multiobjective optimization by moea/d with gaussian process model. *Evolutionary Computation, IEEE Transactions on*, 14(3):456–474, 2010. [cited at p. 152, 247]

[194] C.E. Rasmussen. Gaussian processes to speed up hybrid monte carlo for expensive bayesian integrals. In *Bayesian Statistics 7: Proceedings of the 7th Valencia International Meeting*, pages 651–659. Oxford University Press, 2003. [cited at p. 180]

[195] C.T. Kelley. *Iterative methods for linear and nonlinear equations*. Society for Industrial Mathematics, 1995. [cited at p. 182, 206]

[196] P.J. Olver. Nonlinear systems. In *Applied Mathematics Lecture Notes*. Available online at http://www.math.umn.edu/ olver/appl.html, 2006. [cited at p. 184, 206]

[197] L.C.W. Dixon and G.P. Szego. The global optimization problem: an introduction. *Towards Global Optimization*, 2:1–15, 1978. [cited at p. 186]

[198] M.D. McKay, R.J. Beckman, and WJ Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, pages 239–245, 1979. [cited at p. 188]

[199] P. Audze and V. Eglais. *New approach for planning out of experiments*, volume 35 of *Problems of Dynamics and Strengths*, pages 104–107. Zinatne Publishing House, Riga, 1977. (in Russian). [cited at p. 188]

[200] M. Molga and C. Smutnicki. Test functions for optimization needs. 2005. Available at http://www.zsd.ict.pwr.wroc.pl/files/docs/functions.pdf. [cited at p. 192]

[201] X.-S. Yang. Test problems in optimization. In Xin-She Yang, editor, *Engineering Optimization: An Introduction*. John Wiley & Sons, 2010. [cited at p. 192]

[202] I.M. Sobol. Uniformly distributed sequences with an additional uniform property. *USSR Computational Mathematics and Mathematical Physics*, 16:236–242, 1977. [cited at p. 206]

[203] I.A. Antonov and V.M. Saleev. An economic method of computing LP_τ-sequences. *USSR Computational Mathematics and Mathematical Physics*, 19(1):252–256, 1979. [cited at p. 206]

[204] C. Chisari. SOBOL: The Sobol Quasirandom Sequence. world wide web, February 2011. http://people.sc.fsu.edu/ jburkardt/py_src/sobol/sobol.html; last accessed: 31 May, 2012. [cited at p. 208]

[205] I.M. Sobol. On quasi-monte carlo integrations. *Mathematics and Computers in Simulation*, 47(2):103–112, 1998. [cited at p. 208]

[206] D. Yuret and M. de la Maza. Dynamic hill climbing: Overcoming the limitations of optimization techniques. In *The Second Turkish Symposium on Artificial Intelligence and Neural Networks*, pages 208–212. Citeseer, 1993. [cited at p. 210]

[207] E. Jones, T. Oliphant, P. Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. [cited at p. 210, 212, 233]

[208] E. Hughes. Multi-objective binary search optimisation. In *Second International Conference on Evolutionary Multi-Criterion Optimization*, pages 102–117. Springer, 2003. [cited at p. 212, 235]

[209] K.A. De Jong. *An analysis of the behavior of a class of genetic adaptive systems, 1975*. PhD thesis, University of Michigan, Ann Arbor, 1975. Dissertation Abstracts International 36(10), 5140B (University Microfilms No. 76-9381. [cited at p. 232]

[210] B.L. Miller and M.J. Shaw. Genetic algorithms with dynamic niche sharing for multimodal function optimization. In *Evolutionary Computation, 1996., Proceedings of IEEE International Conference on*, pages 786–791. IEEE, 1996. [cited at p. 232]

[211] O.J. Mengshoel and D.E. Goldberg. Probabilistic crowding: Deterministic crowding with probabilistic replacement. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-99)*, pages 409–416, 1999. [cited at p. 232]

[212] S.W. Mahfoud. Crowding and preselection revisited. In R. Manner and B. Manderick, editors, *Parallel problem solving from nature*, volume 2, pages 27–36. Amsterdam: The Netherlands: Elsevier Science, 1992. [cited at p. 232]

[213] D.E. Goldberg and J. Richardson. Genetic algorithms with sharing for multimodal function optimization. In J . J . Grefenstette, editor, *Proceedings of the Second International Conference on Genetic Algorithms on Genetic algorithms and their application*, pages 41–49. Lawrence Erlbaum Associates Inc., Hillsdale, NJ, 1987. [cited at p. 232, 233]

[214] K. Deb and D. E. Goldberg. An investigation of niche and species formation in genetic function optimization. In *Proceedings of the 3rd International Conference on Genetic Algorithms*, pages 42–50, San Francisco, CA, USA, 1989. Morgan Kaufmann Publishers Inc. [cited at p. 232, 233]

[215] S. W. Mahfoud. A comparison of parallel and sequential niching methods. In *Proceedings of the 6th International Conference on Genetic Algorithms*, pages 136–143, San Francisco, CA, USA, 1995. Morgan Kaufmann Publishers Inc. [cited at p. 232]

[216] David Beasley, David R. Bull, and Ralph R. Martin. A sequential niche technique for multimodal function optimization. *Evol. Comput.*, 1(2):101–125, June 1993. [cited at p. 232]

[217] P. Darwen and X. Yao. A dilemma for fitness sharing with a scaling function. In *Evolutionary Computation, 1995., IEEE International Conference on*, volume 1, pages 166–177. IEEE, 1995. [cited at p. 232]

[218] K. Deb and A. Saha. Multimodal optimization using a bi-objective evolutionary algorithm. *Evolutionary Computation*, 20(1):27–62, 2012. [cited at p. 232]

[219] K. Deb. Genetic algorithms in multimodal function optimization. Master's thesis, Tuscaloosa, AL: University of Alabama, 1989. [cited at p. 233]

[220] C. K. Oei, D. E. Goldberg, and S-J. Chang. Tournament selection, niching, and the preservation of diversity. IlliGAL report 91011, University of Illinois at Urbana Champaign, 1991. [cited at p. 233, 235]

[221] C.G.E. Boender, A.H.G. Rinnooy Kan, G.T. Timmer, and L. Stougie. A stochastic method for global optimization. *Mathematical programming*, 22(1):125–140, 1982. [cited at p. 239, 248]

[222] A.H.G. Rinnooy Kan and G.T. Timmer. Stochastic global optimization methods part I: Clustering methods. *Mathematical programming*, 39(1):27–56, 1987. [cited at p. 239, 248]

[223] A.H.G. Rinnooy Kan and G.T. Timmer. Stochastic global optimization methods part II: Multi level methods. *Mathematical Programming*, 39(1):57–78, 1987. [cited at p. 239, 248]

[224] F. Schoen. Random and quasi-random linkage methods in global optimization. *Journal of Global optimization*, 13(4):445–454, 1998. [cited at p. 239]

[225] S. Kucherenko and Y. Sytsko. Application of deterministic low-discrepancy sequences in global optimization. *Computational Optimization and Applications*, 30(3):297–318, 2005. [cited at p. 239]

[226] A. Eremeev and C. Reeves. On confidence intervals for the number of local optima. *Applications of Evolutionary Computing*, pages 115–115, 2003. [cited at p. 239]

[227] J. Garnier and L. Kallel. How to detect all attraction basins of a function? *Theoretical Aspects of Evolutionary Computation*, pages 343–365, 1999. [cited at p. 239]

[228] S. Kimura and K. Matsumura. Genetic algorithms using low-discrepancy sequences. In *Proceedings of the 2005 conference on Genetic and evolutionary computation*, pages 1341–1346. ACM, 2005. [cited at p. 239]

[229] O. Teytaud. When does quasi-random work? In *Parallel Problem Solving from Nature - PPSN X*, Lecture Notes in Computer Science, pages 325–336. Springer Berlin / Heidelberg, 2008. [cited at p. 239]

[230] B. Kulfan. A universal parametric geometry representation method - "CST". *AIAA-2007-62, 45th AIAA Aerospace Sciences Meeting and Exhibit*, 2007. [cited at p. 246]

[231] F. Yilmaz and M.Y. Gundogdu. A critical review on blood flow in large arteries; relevance to blood rheology, viscosity models, and physiologic conditions. *Korea-Australia Rheology Journal*, 20(4):197–211, 2008. [cited at p. 255]

[232] J.D. Schaffer. Multiple objective optimization with vector evaluated genetic algorithms. In *Proceedings of the 1st international Conference on Genetic Algorithms*, pages 93–100. L. Erlbaum Associates Inc., 1985. [cited at p. 268]

[233] C.M. Fonseca and P.J. Fleming. Multiobjective optimization and multiple constraint handling with evolutionary algorithms. II. application example. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 28(1):38–47, 1998. [cited at p. 268]

[234] C. Poloni. *Hybrid GA for multi objective aerodynamic shape optimization*, volume 33 of *Genetic algorithms in engineering and computer science*, pages 397–415. New York: Wiley, 1995. [cited at p. 268]