

About the author

Rahul-Mark Fonseca was born in Pune, India, on 25th November 1987. Following a Bachelor degree in Petroleum Engineering from *Maharashtra Institute of Technology, Pune*, he headed to the windy plains of the *Netherlands* to pursue a Masters in Applied Earth Sciences with a specialization in Petroleum Engineering at *Delft University of Technology*, from which he graduated, *cum laude*. Upon graduation he pursued his doctoral studies at the *Delft University of Technology* under supervision of Prof.dr.ir. Jan-Dirk Jansen and Prof.dr.ir. Paul Van den Hof. During the past 4 years he has served as Vice-President of the SPE Delft student chapter, and in the organizing committee of the World Petroleum Council Europe Youth Conference. He is an active participant and founding member of the PhD committee of the *Mijnbouwkundige Vereeniging*. As of 1st September 2015, he started working as a reservoir engineering scientist at the Netherlands Organisation for Applied Scientific Research, TNO (*Toegepast Natuurwetenschappelijk Onderzoek*).



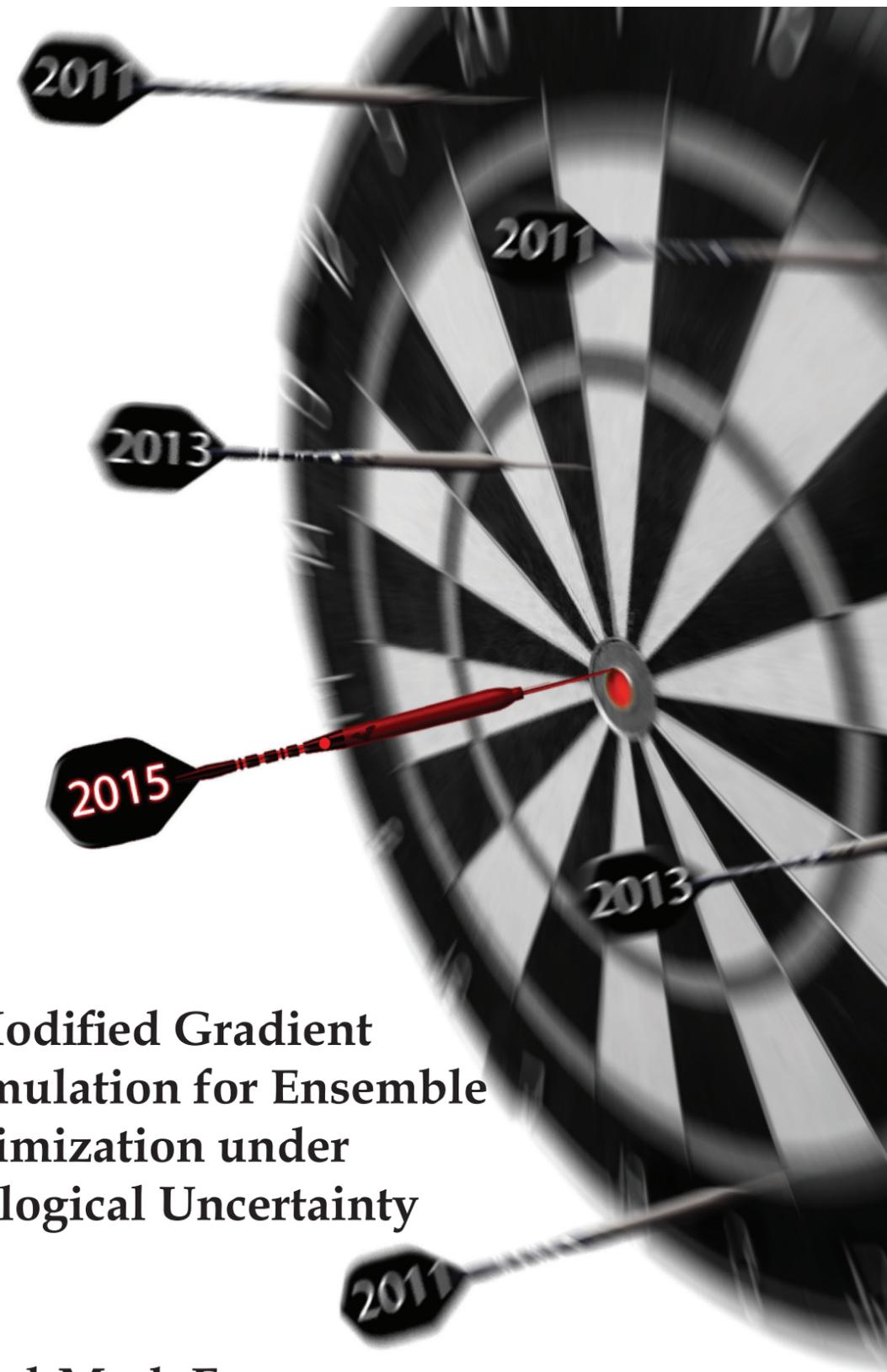
A Modified Formulation for Robust Ensemble Optimization

Rahul-Mark Fonseca

2015

A Modified Gradient Formulation for Ensemble Optimization under Geological Uncertainty

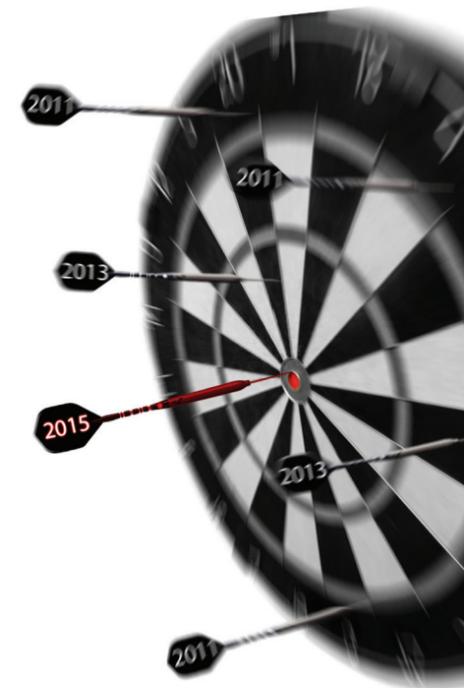
Rahul-Mark Fonseca



Invitation

It's my pleasure to invite you to attend the public defense of my doctoral thesis

A Modified Gradient Formulation for Ensemble Optimization under Geological Uncertainty



Tuesday, December 8th 2015
at 10:00 A.M. in the Frans van Hasseltzaal of the Aula, Delft University of Technology,
Mekelweg 5, Delft.

You are most welcome to attend the reception that will follow the defense at PSOR café, CiTG.

Rahul-Mark Fonseca

Propositions

accompanying the thesis:

A Modified Gradient Formulation for Ensemble Optimization under Geological Uncertainty

Rahul-Mark Fonseca

1. Ensemble-based optimization workflows are a very good alternative to adjoint-based optimization workflows for realistic reservoir models when accounting for geological uncertainty (Chapter 3 & 7).
2. The modified robust ensemble gradient formulation which has been developed in Chapter 2 of this thesis is not only theoretically more 'robust' compared to the original formulation, but also computationally as efficient.
3. Flexibility to different control types such as injection rates, bottom hole pressures, inflow control valve settings, etc., and different varieties of reservoir simulators is the biggest advantage of ensemble-based techniques for optimization workflows.
4. When a certain number of models is used to provide a description of the uncertainties present, utilizing all the uncertainty (models) for the optimization is more 'robust' than selecting a subset of models to describe uncertainty (Chapter 5).
5. The biggest hindrance in the real world towards the application of Closed Loop Reservoir Management techniques is not technological, but mental barriers towards the application of new techniques.
6. In today's scientific world a lot of 'new' research is merely a combination of pre-existing ideas and less of new 'original' ideas (Chapter 4).
7. Procrastination is the art of habitually delaying the execution of more important tasks for significantly less important tasks. Every PhD student should be awarded a double degree: 1. Doctor of Philosophy and 2. Doctor of Procrastination. The order being a point for debate.
8. The anonymity provided by social media has given birth to a behavioural anomaly: "Tigers on Twitter, Mice in Reality".
9. "Religion keeps the poor from murdering the rich." (Napoleon Bonaparte). This is especially true in under-developed countries like India.
10. The grass is never greener on the other side, it is only a different shade.

These propositions are regarded as opposable and defensible, and have been approved as such by the supervisor, Prof. dr. ir. J.D. Jansen and Prof. dr. ir. P.M.J. Van den Hof

Stellingen

Behorende bij het proefschrift:

A Modified Gradient Formulation for Ensemble Optimization under Geological Uncertainty

Rahul-Mark Fonseca

1. Op ensembles gebaseerde optimalisatiemethoden vormen een zeer goed alternatief voor op adjungatie gebaseerde optimalisatiemethoden wanneer rekening moet worden gehouden met geologische onzekerheden in realistische reservoirmodellen (Hoofdstuk 3 & 7).
2. De aangepaste formulering voor de robuuste ensemble gradiënt die werd ontwikkeld in Hoofdstuk 2 van dit proefschrift is niet alleen theoretisch ‘robuuster’ dan de originele formulering, maar ook net zo efficiënt in de berekening.
3. Het grootste voordeel van op ensembles gebaseerde optimalisatiemethoden is de flexibiliteit om een verscheidenheid aan controletypen te gebruiken zoals injectie debiet, druk op de putbodem, de instellingen van instroomcontrolekleppen, etc., alsmede verschillende reservoirsimulatorotypen.
4. Wanneer een bepaald aantal modellen gebruikt wordt om de bestaande onzekerheden weer te geven is het gebruiken van al deze modellen voor de optimalisatie ‘robuuster’ dan het selecteren van een subset om de onzekerheden te beschrijven (Hoofdstuk 5).
5. Het grootste opstakel in de daadwerkelijke toepassing van terugkoppelingsprincipes in reservoirmanagement is niet technologisch van aard, maar wordt gevormd door de metalen weerstand jegens de toepassing van nieuwe technieken.
6. In de hedendaagse wetenschappelijke wereld bestaat ‘nieuw’ onderzoek veelal uit het combineren van reeds bestaande ideeën en minder uit ‘originele’ ideeën (Hoofdstuk 4).
7. Procrastinatie is de kunst van het uit gewoonte uitstellen van belangrijke taken ten bate van minder belangrijke taken. Elke promovendus zou een dubbele graad moeten verdienen: 1. Doctor in de Wetenschappen en 2. Doctor in de Procrastinatie. Over de volgorde kan gediscussieerd worden.
8. De anonimiteit op sociale media heeft geleid tot het ontstaan van een gedragsanomalie: “Tijgers op Twitter, Wezels in Werkelijkheid”.
9. “Religie weerhoudt de armen ervan de rijken te vermoorden.” (Napoleon Bonaparte). Dit geldt vooral in onderontwikkelde landen zoals India.
10. Het gras is nooit groener aan de overkant, maar heeft slechts een andere tint.

Deze stellingen worden oponeerbaar en verdedigbaar geacht en zijn als zodanig goedgekeurd door de promotor, Prof. dr. ir. J.D. Jansen and Prof. dr. ir. P.M.J. Van den Hof.

A Modified Gradient Formulation for Ensemble Optimization under Geological Uncertainty

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op dinsdag 8 december 2015 om 10:00 uur

door

Rahul-Mark FONSECA

Master of Science in Applied Earth Sciences,
Delft University of Technology
geboren te Pune, India.

Dit proefschrift is goedgekeurd door de:

Promotor: Prof.dr.ir. J.D. Jansen

Promotor: Prof.dr.ir. P.M.J. Van den Hof

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. ir. J.D. Jansen,	Technische Universiteit Delft, promotor
Prof. dr. ir. P.M.J. Van den Hof,	Technische Universiteit Eindhoven, promotor

Onafhankelijke leden:

Prof. dr. ir. A.W. Heemink,	Technische Universiteit Delft,
Prof. dr. ir. S. Weiland,	Technische Universiteit Eindhoven,
Prof. dr. C. D. Macbeth,	Heriot Watt University,
Prof. dr. W.R. Rossen	Technische Universiteit Delft,
Dr. ir. O. Leeuwenburgh,	TNO,
Prof. dr. ir. J. Bruining	Technische Universiteit Delft, reservelid.

This research was carried out within the context of the Integrated Systems Approach to Petroleum Production (ISAPP) knowledge centre. ISAPP is a joint industry project between Delft University of Technology, The Netherlands Research Organisation (TNO), ENI, Statoil and Petrobras.

Copyright © 2015, Rahul-Mark Fonseca

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without the written permission of the author.

Cover design by K. Bisdorn, D. Kawale and R.M. Fonseca.

Printed by Gildeprint, The Netherlands

ISBN: 978-94-6233-185-3

“It takes team work to make a Dream work”
- *Dr. Martin Luther King Jr.*

To my Dream Team ,
My Mother, Father and Sister

TABLE OF CONTENTS

1 INTRODUCTION	1
1.1 Closed Loop Reservoir Management	2
1.2 Optimization	3
1.3 Research Objectives	4
1.4 Thesis Outline	5
1.4.1 Ensemble Optimization (EnOpt) : Chapter 2	5
1.4.2 Impact of Ensemble Size on Gradient Quality: Chapter 3	6
1.4.3 Covariance Matrix Adaptation (CMA-EnOpt): Chapter 4	7
1.4.4 Hierarchical Bi-Objective Production Optimization: Chapter 5	8
1.4.5 Generate a Pareto Front using Ensemble Optimization : Chapter 6	9
1.4.6 Robust Bi-Objective Optimization of On/Off Inflow Control Devices	9
2 ENSEMBLE OPTIMIZATION: A THEORETICAL VIEW	11
2.1 Ensemble Optimization (EnOpt)	12
2.1.1 Objective Function.....	12
2.1.2 Ensemble-Based Deterministic Formulation	13
2.2 Robust Optimization	16
2.2.1 Original Ensemble-based Robust Formulation	16
2.2.2 Modified Ensemble-Based Robust Formulation.....	17
2.3 Theoretical Understanding of EnOpt for Robust Optimization	18
2.3.1 Modified Formulation : Theoretical Reasoning.....	20
2.3.2 Justification for the ‘Simplex Gradient’	22
3 EFFECT OF ENSEMBLE SIZE ON GRADIENT QUALITY	25
3.1 Introduction	26
3.2 Theory	27
3.2.1 Adjoint method	28
3.2.2 Hypothesis testing.....	28
3.3 Numerical Example: Rosenbrock Function	29
3.3.1 Deterministic Case.....	30
3.3.2 Rosenbrock Function With Uncertainty	32
3.4 3D Reservoir Model: “Egg Model”	37
3.4.1 Deterministic case.....	38
3.4.2 Egg Model with Uncertainty.....	42
3.5 Conclusions	47

4	ENOPT WITH COVARIANCE MATRIX ADAPTATION	49
4.1	Introduction	50
4.2	Update rules	51
4.3	Covariance Matrix Adaptation	52
4.3.1	Rank- μ Update.....	52
4.3.2	Rank-One Update	53
4.3.3	Combined Rank Update.....	53
4.4	Covariance Matrix Adapted-EnOpt (CMA-EnOpt) Algorithm.....	53
4.5	Results: 5-Spot Synthetic Reservoir Model.....	54
4.5.1	Comparison between EnOpt and CMA-EnOpt.....	56
4.5.2	Improved Robustness.....	57
4.5.3	Learning Rates.....	58
4.5.4	Correlations and Block-Diagonal Update.....	60
4.5.5	Full Matrix Update	62
4.5.6	Comparison of Update types.....	64
4.5.7	Comparison with CMA-ES.....	64
4.5.8	Effect of Different Random Number Sequences	65
4.6	Modified Brugge Model	66
4.6.1	Results	67
4.6.2	Robustness to the Initial Covariance Matrix.....	68
4.6.3	Update Comparison	68
4.7	Conclusions	69
5	HIERARCHICAL BI-OBJECTIVE OPTIMIZATION.....	71
5.1	Introduction	72
5.2	Life Cycle Optimization	72
5.3	Hierarchical Bi-objective Optimization.....	73
5.3.1	Null-space based optimization.....	73
5.3.2	Null-space-based optimization algorithm	75
5.3.3	Switching algorithm.....	76
5.4	Results : Bi-Objective (nominal) Optimization.....	77
5.4.1	Objective functions.....	78
5.4.2	Unconstrained optimization	78
5.4.3	Null-space based optimization.....	79
5.4.4	Switching algorithm.....	80
5.4.5	Computational aspects.....	82
5.4.6	Discussion.....	83
5.4.7	Reactive control.....	83
5.5	Results: Bi-Objective Optimization under Uncertainty.....	84
5.5.1	Life-cycle optimization.....	84
5.5.2	Hierarchical switching optimization.....	85
5.6	Conclusions	86

6	PARETO FRONTS FOR BI-OBJECTIVE OPTIMIZATION	89
6.1	Introduction	90
6.2	Theory.....	91
6.2.1	Update Rules.....	91
6.3	Multi-Objective Optimization	91
6.4	Weighted Sum Method.....	92
6.5	Adjusted Weighted Sum Method	93
6.6	Normal Boundary Intersection (NBI) Method.....	93
6.6.1	Tracking the Pareto front using NBI.....	95
6.6.2	Hierarchical Switching Method	95
6.7	Results : 5-Spot Reservoir Model.....	96
6.7.1	Control Comparison.....	99
6.7.2	Comparison of Weighted Sum Techniques	100
6.7.3	Comparison of Weighted Sum and NBI	101
6.7.4	Hierarchical Switching Optimization Method	103
6.8	Results : Egg Model.....	104
6.8.1	Discussion.....	107
6.9	Conclusions	108
7	ROBUST BI-OBJECTIVE OPTIMIZATION OF ON-OFF INFLOW CONTROL DEVICES: A REALISTIC MODEL.....	109
7.1	Introduction	110
7.2	Theory.....	111
7.3	Control Parameterization	111
7.4	Results: Field Case Inspired Application	112
7.4.1	Life Cycle Optimization	114
7.4.2	Effect of Discount Factor.....	115
7.4.3	Reactive Control	117
7.4.4	Optimized Control Strategy	118
7.4.5	Comparison of Different Gradient Formulations.....	118
7.4.6	Reduction in Number of ICDs.....	119
7.4.7	Hierarchical Switching Optimization	120
7.4.8	Discussion.....	122
7.5	Conclusions	123
8	CONCLUSIONS & FUTURE PERSPECTIVES	125
8.1	Conclusions	125
8.2	Future Perspectives	130
9	APPENDIX.....	137
10	BIBLIOGRAPHY.....	139

11 SYMBOLS AND NOTATION	146
12 LIST OF PUBLICATIONS	147
13 SUMMARY	149
14 SAMENVATTING.....	151
15 ACKNOWLEDGEMENTS.....	154

1

INTRODUCTION

Energy security to sustain the demands of rapidly growing developing nations has been one of the driving forces for the ever increasing demand of hydrocarbons i.e. petroleum and natural gas, which are finite natural resources. The energy consumption of the emerging economies such as India, China, Brazil etc. has grown at nearly 100% over the past decade, see; *BP Statistical Review of World Energy* (2014). At the same time, the evolution of alternative energy sources has not been keeping pace with rising energy demand. Additionally, while applications of alternative energy sources such as solar, wind, biofuels etc. might be economically feasible in developed countries in the near future, in developing countries practically little or no infrastructure exists to efficiently use these alternative energy sources. Thus the dependency on traditional energy sources such as petroleum and natural gas will continue. This coupled with an increasing world population and a quest for higher standards of living will further increase the energy demand. The increased energy demand together with the decreasing availability of ‘easy oil’, i.e. hydrocarbons which are relatively easy to extract from deep below the earth’s surface, has led to an increasing number of applications which use advanced technologies to improve and increase recovery of hydrocarbons. A class of these advanced technologies are enhanced oil recovery methods which use a range of complex substances like chemicals, polymers etc. to increase the total volumetric recovery of hydrocarbons from existing reservoirs. These techniques have also been successfully applied to more complex and challenging reservoirs, commercial development of which was not possible till recently, and is heavily dependent on oil prices. On the other hand, with traditional (reactive) reservoir

management practices, currently the worldwide recovery factor, i.e. the ratio between the volume of hydrocarbons produced to the total hydrocarbon volume present, is approximately 30%; see; *BP Statistical Review of World Energy* (2014). Thus, on average, we produce only 30% of the oil actually present in the subsurface while leaving behind approximately 70%, i.e. we leave behind more than we produce. This is partly due to economic and technical viability but also many times due to an inefficient production process. Thus the question arises as to whether using a proactive reservoir management strategy instead of the traditional reactive strategy will aid in increasing the total volume of hydrocarbons produced.

1.1 Closed Loop Reservoir Management

Recent advances in technology have made it possible to control the production of hydrocarbons through a system of wells. Such devices, in general terminology, are called Inflow Control Devices (ICDs) which are installed in wells and can be operated either mechanically, hydraulically or even remotely. Wells which have been installed with such technology are called Smart Wells. In addition to the ability to control the production process nowadays, wells can be equipped with downhole measurement devices such as pressure and temperature sensors or also devices to measure flow rates, fluid compositions etc. These measurement devices provide significantly more information about the production process in the well from the reservoir compared to traditional wells, Jansen et al. (2005). In addition with an increase in computing power we are able to build models that are assumed to be representative of the system i.e. the real reservoir being produced. To generate a model of the reservoir, information is obtained from various sources such as seismic data, well logs, geological insight etc. which are highly uncertain. This coupled with the fact that we cannot “see” any reservoir, as it is buried deep below the surface of the earth, these models are usually not representative of the true reservoir description. Thus information obtained from the measurement devices can be used to update/improve the reservoir model. In the petroleum industry this has traditionally been done manually in batch mode, i.e. once every 4-5 years, and is known as "history matching". Oliver and Chen (2011) provide an extensive review of various computer-assisted history matching techniques available, their advantages and drawbacks. Thus in essence while history matching updates the uncertain parameters to “match” the history, it really aims to improve the predictive capability of the models being used.

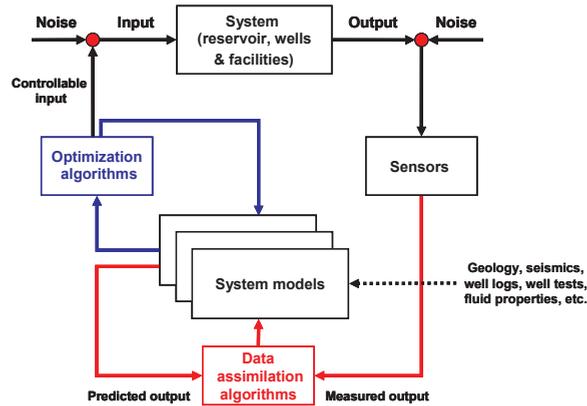


Fig. 1.1: Illustration of the closed-loop reservoir management framework. Taken from Jansen et al. (2009).

In addition to improving predictability, the models can also be used to optimize the volumes of hydrocarbons produced or an economic objective based on the volumes produced. For example, the settings over time of the ICDs or injection and production rates are variables which are often used in optimization. Traditionally, reservoir management strategies have been reactive in nature. E.g., in scenarios where water is injected into the reservoir to increase the volume of oil produced, the wells will be stopped from producing only when it is no longer commercially viable to continue production. Using optimal control theory in conjunction with reservoir models, such a reactive reservoir management strategy can be made “proactive”. There are many techniques which can be used to find a set of optimal controls, references of which can be found in [Brouwer and Jansen (2004), Sarma et al. (2005), Jansen (2011), Chen et al. (2009), Do and Reynolds (2013), etc.]

Jansen et al. (2005) proposed to combine the history matching step with the optimization step together in a closed loop framework to achieve an improved reservoir management strategy. **Fig. 1.1** shows a block-diagram representation of the closed loop reservoir management process taken from Jansen et al. (2009). As can be observed in Fig. 1.1, there are two distinct loops, an optimization loop in blue and the history matching/model updating loop in red. In the remainder of this thesis we will focus on the application of different techniques for optimization (blue loop) and do not consider the history matching (red loop).

1.2 Optimization

Several studies have shown that there is considerable scope to improve the economic life-cycle performance of oil fields through the use of formal optimization

methods in conjunction with reservoir simulation models. For such problems, gradient-based techniques, in terms of accuracy and computational efficiency, are the most successful and widely applied. A very efficient way to perform such model-based life-cycle optimization is with the aid of gradient-based methods where the gradient is obtained through an adjoint technique. An overview of the adjoint method used in Brouwer and Jansen (2004), Sarma et al. (2005) amongst others, and a large number of references can be found in the review paper by Jansen (2011). Oliver et al. (2008) provides a review and relevant references detailing the application of the adjoint method in reservoir engineering for computer-assisted history matching. The adjoint method not only provides the most accurate gradient, it is also computationally very efficient. However, it is an intrusive method, requiring access to the simulator source code as well as extensive implementation efforts. Because it is practically impossible to access commercial simulator source codes for implementation of the adjoint there is a need for alternative methods for model-based production optimization in which the simulator is treated as a black-box. Additionally the adjoint method is not very flexible in adaptation to different control types such as water injection rates, oil production rates, bottom hole pressures, settings of inflow control devices or valves etc.,. These limitations of the adjoint method have led to the development of alternative gradient-based techniques. One such alternative gradient-based technique and its applicability to a range of different problems is investigated in this thesis.

1.3 Research Objectives

The main objective of this research is to

Investigate the applicability of an approximate gradient technique to balance long and short-term production optimization targets with and without uncertainty.

For this purpose we :

- Investigate the applicability of using a computationally efficient method for ensemble optimization under geological uncertainty and develop a theoretical understanding of the method.
- Investigate the impact of ensemble size on the quality of an approximate gradient to be used for optimization with and without geological uncertainty using principles of hypothesis testing and statistical principles.
- Investigate the impact of adaptively updating the covariance matrix which is used to generate the ensemble of controls used in the estimate of the approximate gradient during the optimization process.

- Investigate the applicability of various bi-objective optimization techniques using an approximate gradient to achieve balanced long and short-term production optimization.
- Investigate the applicability of the ensemble optimization technique incorporating geological uncertainty for bi-objective optimization to optimize inflow control device settings of a sector model inspired from a real field case.

1.4 Thesis Outline

In this section we provide an introduction of the general concepts investigated based on the research objectives in the different chapters included in this thesis.

1.4.1 Ensemble Optimization (EnOpt) : Chapter 2

The approximate gradient technique considered in this thesis is the Ensemble Optimization (EnOpt) method, inspired by the Ensemble Kalman Filter (EnKF) method. Predecessors to the EnOpt method were proposed by Lorentzen et al. (2006) and Nwaozo (2006), where after Chen (2008) and Chen et al. (2009) gave systematic descriptions of the method as mostly used today. Chen (2008) proposed the now standard formulation of the EnOpt method which uses an ensemble of randomly perturbed control vectors to approximate a gradient of the objective function with respect to some specific controls. Thereafter, several publications addressed applications and computational aspects of the method; see e.g. Chaudhri et al. (2009), Chen and Oliver (2010), Su and Oliver (2010), Leeuwenburgh et al. (2010), and Chen and Oliver (2012). In a recent paper, Do and Reynolds (2013) demonstrate that EnOpt can be interpreted as a member of a broader class of approximate-gradient methods that also includes the simultaneous perturbation stochastic approximation (SPSA) method. The major advantages of EnOpt are its ease of implementation, flexibility to adapt to different control types and ability to be used with any reservoir simulator. The major drawback of this method, relative to the adjoint method, is its computational inefficiency and inaccuracy of the gradient approximation. Most of the publications about EnOpt have focused on large-scale water flooding production optimization problems starting from a single reservoir model. However, in reality the geological and reservoir modeling process is fraught with uncertainties since a reservoir is modeled using uncertain interpretations based on uncertain data sources such as seismic, well logs etc. Incorporating these uncertainties into the optimization framework is vital to achieve results of any practical significance. This uncertainty will translate into a

distribution of possible objective function values which cannot be adequately characterized by a single model outcome. Yeten et al. (2003) described an approach to account for geological uncertainty during well-location optimization with the aid of multiple models. Van Essen et al. (2009) provided a list of references of non-petroleum engineering applications that incorporated uncertainty within the modeling and control framework. They introduced a ‘robust optimization’ methodology in conjunction with the adjoint method to include the effect of uncertainties into the optimization framework. They used an ensemble of equi-probable reservoir models with differing geology and maximized the expectation of the objective function over this ensemble of models. Chen (2008) introduced this robust optimization concept within the ensemble optimization framework. They proposed the use of an ensemble of controls of equal size as the ensemble of geological models. Coupling of one member from the control ensemble with one member of the geological ensemble, a mean gradient can be approximated with the EnOpt formulation. This formulation, while computationally very attractive for robust optimization, has received scant attention with respect to its theoretical understanding. Recently Fonseca et al. (2014) demonstrated a case wherein the original formulation for ensemble-based robust optimization leads to inferior results and suggested a modified gradient formulation, which through the aid of numerical experiments showed that the modified formulation achieved significantly better results, i.e. higher objective function values with the same computational efficiency.

This chapter will first give the theory of the Ensemble Optimization (EnOpt) method as used in this thesis for deterministic (nominal) optimization problems. Following this, the theory underlying EnOpt for robust optimization is provided along with theoretical reasoning which provides insights into the superior performance of the modified robust EnOpt gradient formulation.

1.4.2 Impact of Ensemble Size on Gradient Quality: Chapter 3

For EnOpt the two main inputs that influence the quality of the approximate gradient are the covariance matrix used to create the ensemble of perturbed controls and the number of control samples generated, i.e. the ensemble size. The effect of the covariance matrix has been investigated recently in Fonseca et al. (2015) and a theoretical foundation for the use of a varying covariance matrix has been provided in Stordal et al. (2014). However none of those studies have performed a detailed investigation into the effect of ensemble size on the estimated ensemble gradient quality. In this chapter we aim to quantify the ensemble size required to approximate a gradient comparable to the adjoint gradient especially for robust optimization

problems, using principles from hypothesis testing and statistical analysis. We first provide an introduction of the test statistics used to validate our results using a hypothesis testing methodology. This will be followed by a detailed set of experiments on a widely used optimization test function, the Rosenbrock function, for cases with and without model uncertainty. Finally we test the proposed methodology on a medium-sized reservoir model, again with and without geological uncertainty. We also show through numerical experiments the effect of a poor quality gradient on the optimization process for both deterministic i.e. single model realisation and robust optimization cases.

1.4.3 Covariance Matrix Adaptation (CMA-EnOpt): Chapter 4

When using EnOpt the gradient of the objective function with respect to the vector of control variables is approximated by first evaluating the objective function values for an ensemble of control vectors. This ensemble of control vectors is generated from a multi-Gaussian random distribution with a constant prescribed covariance matrix. The gradient is then estimated through an optimal regression model using a least-squares approach. As an alternative to exact or approximate gradient-based optimization methods one can revert to gradient-free methods such as genetic algorithms or evolutionary strategies as developed in the ‘machine-learning’ community. One of the latter, called the Covariance Matrix Adapted-Evolutionary Strategy (CMA-ES), which was developed by Hansen and co-workers (Hansen and Ostermeier 1996, 2001, Hansen 2006), has recently been used for well placement optimization by Ding (2008) and Bouzarkouna et al. (2011), in a flooding optimization problem by Schulze-Riebert et al. (2011), and for a smart well optimization problem by Pajonk et al. (2011). The main idea in CMA-ES is to systematically adapt the variance of the control vector sample in directions that have proven to be successful. In this chapter we propose an improvement to EnOpt in which the use of a constant prescribed covariance matrix throughout the optimization is replaced with a covariance matrix that is constantly adapted using the CMA-ES logic. We will refer to the resulting hybrid scheme as Covariance Matrix Adapted EnOpt (CMA-EnOpt). CMA-EnOpt combines the advantage of explicitly using gradient information to achieve faster convergence (EnOpt) with the continuous adaptation of the covariance matrix (CMA) to improve the gradient estimate via improved sampling using “local” knowledge of the nature of the objective function search space. In this chapter we first provide an application of CMA-EnOpt to a small synthetic 3D reservoir model and a modified version of the Brugge benchmark model. A comparison of the results to those obtained

with EnOpt will illustrate the advantages of CMA-EnOpt for relatively large-scale model-based production optimization.

1.4.4 Hierarchical Bi-Objective Production Optimization: Chapter 5

Various model-based optimization studies have shown the potential of using optimal control theory for dynamic optimization of petroleum systems, thereby improving overall reservoir management. Most of these studies used a single reservoir model to optimize a single life-cycle (i.e. long-term) objective, for example, water flooding strategies that typically aim to prevent early water breakthrough at the production wells. The resulting operating strategies are suboptimal, since in reality well and field operation strategies are typically based on operational production criteria such as delivery contracts, which are governed by much shorter time horizons (days to weeks or months) than life-cycle objectives (years to decades). In fact, strategies that optimize short-term operational objectives are often in conflict with optimal long-term strategies. Jansen et al. (2009) observed that significantly different optimized long-term water flooding strategies result in nearly equal values of the objective function, defined as net present value (NPV). Thus, there exist multiple solutions to the optimization problem, and different initial starting points may lead to different solutions in an optimal subset of the decision variable space. They concluded that the life cycle optimization problem is ill-posed as a result of the non-uniqueness of the solution. A similar non-uniqueness in minimizing the mismatch between measured and simulated data during computer-assisted history matching was demonstrated by Oliver et al. (2008). This implies the presence of redundant degrees of freedom (DOFs) in the high dimensional optimization problems which could possibly be exploited to optimize multiple objectives. This idea, together with the (indirect) indication that the DOFs appear to manifest themselves as ridges in the objective function, formed the basis for the hierarchical multi-objective optimization structure proposed by Van Essen et al. (2011). They suggested two variants of a hierarchical optimization scheme to include secondary objectives into the life cycle optimization using the adjoint formulation. They observed a significant increase in short-term objectives with minimal change to the primary objective function for both the variants. Similar results were obtained by Chen et al. (2012) and Suwartadi et al. (2012). In this chapter the applicability of EnOpt instead of the adjoint method when using the hierarchical structure for bi-objective optimization is investigated for cases with and without geological uncertainty.

1.4.5 Generate a Pareto Front using Ensemble Optimization : Chapter 6

When dealing with conflicting objectives decision makers usually prefer to have multiple strategies to choose from. Isebor and Durlofsky (2014) applied an evolutionary algorithm to generate points along a “Pareto” front for a bi-objective water flooding problem. Yasari et al. (2014) applied the popular NSGA-II evolutionary algorithm also to generate a Pareto front. Liu and Reynolds (2014) applied the normal boundary intersection method (NBI) first introduced in Das and Dennis (1998) to a bi-objective water flooding problem for cases with and without geological uncertainty using the adjoint technique. In this chapter the applicability of EnOpt to generate points along a “Pareto” front with acceptable computational effort, and its applicability to solve constrained optimization problems using the augmented Lagrangian method are investigated for two problems, one where ICD settings are the controls and the other where injection rates are the controls.

1.4.6 Robust Bi-Objective Optimization of On/Off Inflow Control Devices

A second reason for sub-optimality resulting from life-cycle optimization is the inherent uncertainty present in the geological and petro-physical modeling that forms the basis for the reservoir model. In an example Van Essen et al. (2009) showed that robust optimization increased the expected value and reduced the variance of the optimized strategy applied to the different geological realizations in comparison to a reactive strategy using the adjoint formulation for gradient-based optimization. Chen et al. (2009) reported a successful application of this approach to the SPE Brugge benchmark case. Recently Raniolo et al. (2013) and Li et al. (2012), have investigated the applicability of approximate gradient techniques for life-cycle robust water flooding optimization while Yang et al. (2011) applied the robust optimization principle to a Steam-Assisted Gravity Drainage (SAGD) application. A limited number of optimization studies such as Bailey et al. (2005), Sarma et al. (2008), Alhutali et al. (2009), Chaudhri et al. (2010), Forouzanfar et al. (2013), Van Essen et al. (2010) and Raniolo et al. (2013) amongst others have used realistic field scale or sector models. Most of the studies that use realistic real field models concerned single-objective optimization on a single geological realization, with the exception of Alhutali et al. (2009) and Raniolo et al. (2013) performed single-objective optimization using an ensemble of geological realizations. In this chapter we use an ensemble of sector models inspired from a real field case for the optimization. The settings of Inflow Control Devices (ICDs) which have discrete settings, i.e. either 0 or 1, are the controls used. However, EnOpt, like any other gradient-based technique, cannot efficiently

handle discrete control problems. Through the use of a re-parameterization of the controls into switching times the controls are transformed into continuous variables which can be efficiently handled by EnOpt. In this chapter the applicability of EnOpt incorporating geological uncertainty for hierarchical bi-objective optimization over an ensemble of sector models inspired from a real field case is investigated.

The thesis is based on publications written during the completion of the PhD project. Chapter specific conclusions are provided at the end of each chapter. The thesis is concluded with a short overview of the general conclusions from this thesis followed by a list of future perspectives and further research areas regarding the applicability of approximate gradient techniques such as EnOpt for balanced long and short term optimization of oil recovery.

2

ENSEMBLE OPTIMIZATION: A THEORETICAL VIEW

There are numerous methods for model-based optimization of hydrocarbon recovery. These methods can be classified into two general classes, derivative-based and derivative free techniques. Derivative/Gradient-based methods have been shown to be computationally more efficient than derivative free methods. Among the gradient-based methods the adjoint approach, see Jansen (2011) for an overview, provides the most accurate gradient and is computationally very efficient. However the adjoint approach has the disadvantage that it requires access to the simulation code to be implemented. Chen (2008) and Chen et al. (2009) introduced the ensemble-based optimization method (EnOpt)¹, an approximate gradient-based method which is computationally less attractive than the adjoint method but does not require simulator access and has proven to achieve good results. An earlier, somewhat different version of EnOpt was proposed by Lorentzen et al. (2006) and Nwaozo (2006), and there exist other ensemble-based methods such as the simultaneous perturbation stochastic approximation (SPSA; Spall et al., 1998), or the covariance matrix adaptation evolutionary strategy (CMA-ES; Hansen, 2006) to perform model based optimization studies. Do and Reynolds (2013) provided theoretical connections between the various approximate gradient techniques. This chapter details the formulations of the ensemble optimization (EnOpt) technique for deterministic and robust optimization problems investigated in this thesis and outlines a more ‘robust’ theoretical understanding of the method.

¹ This chapter is based on Fonseca, R.M., Chen, B., Jansen, J.D. and Reynolds, A.C. 2015: Theoretical Understanding of an Approximate Ensemble-based Gradient incorporating Geological Uncertainty. Submitted to *International Journal of Numerical Methods in Engineering*

2.1 Ensemble Optimization (EnOpt)

This technique, as proposed by Chen (2008) and Chen et al. (2009), is a stochastic gradient-based optimization method, which utilizes an ensemble of control vectors to estimate a gradient. EnOpt approximates the gradient based on the sensitivity of the ensemble averaged over the objective function with respect to the controls. Distinct characteristics of the EnOpt method are (Chen, 2008; Chen et al., 2009):

- The search direction (gradient) is obtained from the ensemble of controls generated.
- It is largely independent of simulator specifics, and requires minimal code development.
- It has been shown to work with high-dimensional control vectors, for e.g. the Brugge test case, see Peters et al. (2013) for an overview of the results.
- It can be applied to maximize the expected objective function based on multiple geological realizations.

Approximating the gradient from the sensitivity of the ensemble enables the use of any type of control variable without modification to the existing algorithm, thus another advantages of the method is its flexibility. Additionally with increases in computing power and resources the method is inherently parallelizable as each ensemble member can be evaluated independently thus the computational efficiency is another attractive feature of this technique.

2.1.1 Objective Function

The two most commonly used objective functions for production optimization are ultimate recovery or an economic objective such as Net Present Value (NPV). The flexibility of EnOpt allows any objective function to be used, in this thesis we chose the objective function J to be the NPV, defined in the usual fashion as

$$J = \sum_{k=1}^K \left(\frac{\{[(q_{o,k}) \cdot r_o - (q_{wp,k}) \cdot r_{wp}] - [(q_{wi,k}) \cdot r_{wi}]\} \cdot \Delta t_k}{(1+b)^{t_k/\tau_t}} \right), \quad (2.1)$$

where $q_{o,k}$ is the oil production rate in bbl/day, $q_{wp,k}$ is the water production rate in bbl/day, $q_{wi,k}$ is the water injection rate in bbl/day, r_o is the price of oil produced in \$/bbl, r_{wp} is the cost of water produced in \$/bbl, r_{wi} is the cost of water injected in \$/bbl, Δt_k is the difference between consecutive time steps in days, b is the discount factor expressed as a fraction per year, t_k is the cumulative time in days corresponding to time step k , and τ_t is the reference time period for discounting, typically one year.

2.1.2 Ensemble-Based Deterministic Formulation

In model-based reservoir management applications, for a given model \mathbf{m} and a vector of control inputs \mathbf{u} , $y(\mathbf{u}, \mathbf{m})$ is the solution of a system of nonlinear partial differential equations for an initial-boundary-value problem (IBVP). As we assume this system of equations is deterministic, J is a function of \mathbf{u} and y only, i.e., $J = J(\mathbf{u}, \mathbf{m})$. Often the solution of the IBVP can only be approximated via the solution of a system of nonlinear discrete equations generated from a finite volume formulation as is the case for the problems considered in this thesis and in many other studies such as Chen et al. (2009), Do and Reynolds (2013), Fonseca et al. (2015) etc.

In this section we outline the standard formulation of the EnOpt algorithm for deterministic (i.e., single model realization) optimization as proposed by Chen et al. (2009). We take \mathbf{u} to be a single control vector containing all the control variables to be optimized. This vector has length N equal to the product of the controllable well parameters (number of well settings like bottom hole pressures, rates or valve settings) and the number of control time steps. Chen et al. (2009) sample the initial mean control vector from a Gaussian distribution while, at later iteration steps the final control vector of the previous iteration is taken as the mean control. However the initial controls can also be chosen by the user, as will be done in our experiments.

$$\mathbf{u} = [u_1 \quad u_2 \quad \cdots \quad u_N]^T. \quad (2.2)$$

To estimate the EnOpt gradient, a multivariate, Gaussian distributed ensemble $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_M\}$ is generated with a distribution mean \mathbf{u} and a predefined distribution covariance matrix $\tilde{\mathbf{C}}$ where M is the ensemble size. During the iterative optimization process, \mathbf{u} is updated until convergence, whereas $\tilde{\mathbf{C}}$ is, traditionally, kept constant. [An alternative procedure, in which $\tilde{\mathbf{C}}$ is updated during the optimization process, is described in Fonseca et al. (2015), see chapter 4]. In our implementation of EnOpt the ensemble members \mathbf{u}_i , $i = 1, 2, \dots, M$, are created using

$$\mathbf{u}_i = \mathbf{u} + \tilde{\mathbf{C}}^{1/2} \mathbf{z}_i, \quad (2.3)$$

and

$$\bar{\mathbf{u}} = \frac{1}{M} \sum_{i=1}^M \mathbf{u}_i \quad (2.4)$$

where we use a Cholesky decomposition to calculate $\tilde{\mathbf{C}}^{1/2}$, and draw \mathbf{z}_i from a univariate Gaussian distribution, this is not a requirement and other distributions can be used, see, Sarma and Chen (2014) who used Sobol sampling to generate the

ensemble of control vectors. To estimate the gradient, a mean-shifted ensemble matrix is defined as

$$\mathbf{U} = [\mathbf{u}_1 - \bar{\mathbf{u}} \quad \mathbf{u}_2 - \bar{\mathbf{u}} \quad \cdots \quad \mathbf{u}_M - \bar{\mathbf{u}}]. \quad (2.5)$$

A mean-shifted objective function vector is defined as

$$\mathbf{j} = [J(\mathbf{u}_1) - \bar{J} \quad J(\mathbf{u}_2) - \bar{J} \quad \cdots \quad J(\mathbf{u}_M) - \bar{J}]^T, \quad (2.6)$$

where the average of the objective function is given by

$$\bar{J} = \frac{1}{M} \sum_{i=1}^M J(\mathbf{u}_i). \quad (2.7)$$

The approximate gradient as proposed by Chen (2008) and Chen et al. (2009) is given by

$$\mathbf{g} = \mathbf{C}_{uu}^{-1} \mathbf{c}_{uJ}, \quad (2.8)$$

where

$$\mathbf{C}_{uu} = \frac{1}{M-1} (\mathbf{U}\mathbf{U}^T), \quad (2.9)$$

and

$$\mathbf{c}_{uJ} = \frac{1}{M-1} (\mathbf{U}\mathbf{j}), \quad (2.10)$$

are ensemble (sample) covariance and cross-covariance matrices respectively. (Note that \mathbf{c}_{uJ} is a one-dimensional matrix, i.e. a vector.) For the usual case where $M < N$, matrix \mathbf{C}_{uu} is rank-deficient, and Chen (2008) and Chen et al. (2009) therefore propose not to use expression (2.8) but, instead, to use

$$\mathbf{g}' = \mathbf{C}_{uu} \mathbf{C}_{uu}^{-1} \mathbf{c}_{uJ} = \mathbf{c}_{uJ}, \quad (2.11)$$

or

$$\mathbf{g}'' = \mathbf{C}_{uu} \mathbf{c}_{uJ}. \quad (2.12)$$

Alternatively, the pre-multiplication in equation (2.12) can be performed with $\tilde{\mathbf{C}}$, leading to

$$\mathbf{g}''' = \tilde{\mathbf{C}} \mathbf{c}_{uJ}. \quad (2.13)$$

All three expressions (2.11), (2.12) and (2.13) can be interpreted as modified (regularized or smoothed) approximate gradients. In this thesis usually a straight gradient is used, i.e. expression (2.8), computed as the underdetermined least squares solution

$$\mathbf{g} = (\mathbf{U}\mathbf{U}^T)^{-1}\mathbf{U}\mathbf{j} = \mathbf{U}^\dagger\mathbf{j}. \quad (2.14)$$

where the superscript \dagger indicates the Moore-Penrose pseudo inverse, which is conveniently computed using a singular value decomposition (SVD); see, e.g., Golub et al. (1998) Strang (2006), where the SVD on \mathbf{U} has a truncation level of 0.999. Moreover, it is also possible to use single-smoothed and double-smoothed versions of equation (2.14) :

$$\mathbf{g}''' = \tilde{\mathbf{C}}(\mathbf{U}^\dagger\mathbf{j}), \quad (2.15)$$

$$\mathbf{g}'''' = \tilde{\mathbf{C}}\tilde{\mathbf{C}}(\mathbf{U}^\dagger\mathbf{j}), \quad (2.16)$$

Equation (2.14) was also described in Dehdari and Oliver (2012), while Do and Reynolds (2013) recently demonstrated that it is akin to what is known as a ‘Simplex gradient’ in, e.g., Conn et al. (2009). Do and Reynolds (2013) also provided theoretical connections between various ensemble methods such as simultaneous perturbation stochastic approximation (SPSA), Simplex gradient, EnOpt etc. Moreover, they proposed a modification to the gradient formulation which uses the current control vector \mathbf{u}^ℓ and the corresponding objective function value J^ℓ to calculate the control and objective function anomalies \mathbf{U} and \mathbf{j} :

$$\mathbf{U} = [\mathbf{u}_1 - \mathbf{u}^\ell \quad \mathbf{u}_2 - \mathbf{u}^\ell \quad \cdots \quad \mathbf{u}_M - \mathbf{u}^\ell], \quad (2.17)$$

$$\mathbf{j} = [J(\mathbf{u}_1) - J(\mathbf{u}^\ell) \quad J(\mathbf{u}_2) - J(\mathbf{u}^\ell) \quad \cdots \quad J(\mathbf{u}_M) - J(\mathbf{u}^\ell)]^T, \quad (2.18)$$

where the superscript ℓ is the optimization iteration counter. Equations (2.11)-(2.16) can all be used to estimate a gradient-based on either the original [equations (2.5) and (2.6)] or the modified [equations (2.17) and (2.18)] formulations. Thus we can estimate as many as twelve different gradient formulations for deterministic cases. Note that the regularized gradients \mathbf{g}' , \mathbf{g}'' and \mathbf{g}''' are dimensionally inconsistent, in the sense that their elements do not have the same magnitude as those of \mathbf{g} . Chen and Oliver (2012), Oliveira and Reynolds (2014), and Zhao et al. (2013) use equation (2.12) or equation (2.13) for the gradient estimate. When using diagonal covariance matrices, equations (2.11) and (2.13) or any of the regularized gradient varieties act only as a scaling of the

magnitude of the gradient and have no impact on the direction of the gradient. In chapter 4 we investigate the impact of the covariance matrix on the gradient estimate used during an optimization experiment for different cases.

Gradient-based optimization requires the gradient $\mathbf{g} = (dJ/d\mathbf{u})^T$ which is used within an optimization algorithm to iteratively optimize the objective function. For a detailed description of various available optimization algorithms see, e.g., Nocedal and Wright (2006). Individual chapters in this thesis contain more information about the optimization algorithm used. Usually the elements of the control vector are required to stay within upper and lower bounds, and different approaches for such bound control problems are available. Moreover, in addition to these constraints on the inputs, there may be constraints on the outputs of the simulator, which are much more difficult to handle. For all the problems considered in this thesis we only consider simple bound constraints.

2.2 Robust Optimization

The previous section outlines the mathematical formulations of the EnOpt method when a single model realization is used. In our application the model used is a geological model. Geological modeling and interpretation is an inherently uncertain process. Incorporating these uncertainties into the optimization framework reduces the uncertainty involved with the optimization; see Yeten et al. (2003). Van Essen et al. (2009) first presented for petroleum engineering applications a ‘robust optimization’, i.e. optimization over an ensemble of geological realizations which represents the geological uncertainty, methodology. In an example they showed that this technique increased the expected value and reduced the variance of the optimized strategy applied to the different geological realizations in comparison to a reactive strategy using adjoint-based optimization.

2.2.1 Original Ensemble-based Robust Formulation

An EnOpt formulation for robust optimization was introduced in Chen (2008) and Chen and Oliver (2010) wherein they use two ensembles, one of controls and another of geological models, which consist of M members each. Intuitively in such a scenario we would require M^2 function evaluations for a gradient estimate which is computationally not attractive. Thus, in order to make the method computationally more efficient, Chen (2008) provided an argumentation for the possibility of evaluating only M samples to approximate the robust EnOpt gradient. To estimate a ‘robust gradient’, she coupled one member from the control ensemble with one member of the geological ensemble, i.e. in a 1:1 ratio, which is referred to as the ‘original

formulation' in this thesis. Recently Stordal et al. (2014) reached a similar conclusion starting from a different mathematical viewpoint. Thus for robust optimization using the formulation proposed by Chen (2008) we have

$$\mathbf{U} = [\mathbf{u}_1 - \bar{\mathbf{u}} \quad \mathbf{u}_2 - \bar{\mathbf{u}} \quad \cdots \quad \mathbf{u}_M - \bar{\mathbf{u}}], \quad (2.19)$$

where $\bar{\mathbf{u}}$ is as defined in equation (2.4) and the objective function anomalies for robust optimization is given by

$$\mathbf{j} = [J(\mathbf{u}_1, \mathbf{m}_1) - \bar{J} \quad J(\mathbf{u}_2, \mathbf{m}_2) - \bar{J} \quad \cdots \quad J(\mathbf{u}_M, \mathbf{m}_M) - \bar{J}]^T, \quad (2.20)$$

where $\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_M$ represent the different geological realizations used to quantify the model uncertainty and where the expectation of the objective function is given by

$$\bar{J} = \frac{1}{M} \sum_{i=1}^M J(\mathbf{u}_i, \mathbf{m}_i). \quad (2.21)$$

2.2.2 Modified Ensemble-Based Robust Formulation

The formulation proposed by Chen (2008) was found to achieve optimized strategies the objective function of which was inferior to a traditional reactive control strategy, see Raniolo et al. (2013) and Fonseca et al. (2014). Additionally, the theoretical understanding of using this 1:1 ratio is still incomplete. As an alternative to this formulation, Fonseca et al. (2014) propose a modified formulation for the robust EnOpt gradient which no longer uses the mean-shifted control samples and objective values, equations (2.19) and (2.20). Instead, in equation (2.22) the control sample mean $\bar{\mathbf{u}}$ is replaced by the control vector of the current iteration step, \mathbf{u}^ℓ :

$$\mathbf{U} = [\mathbf{u}_1 - \mathbf{u}^\ell \quad \mathbf{u}_2 - \mathbf{u}^\ell \quad \cdots \quad \mathbf{u}_M - \mathbf{u}^\ell], \quad (2.22)$$

The new formulation replacing equation (2.20) is

$$\mathbf{j} = [J(\mathbf{u}_1, \mathbf{m}_1) - J(\mathbf{u}^\ell, \mathbf{m}_1) \quad \cdots \quad \cdots \quad J(\mathbf{u}_M, \mathbf{m}_M) - J(\mathbf{u}^\ell, \mathbf{m}_M)]^T. \quad (2.23)$$

Note that equation (2.17) is identical to equation (2.22) as used in the deterministic modified expression of Do and Reynolds (2013), but that equation (2.18) is different from equation (2.23). This modified gradient formulation [based on equations (2.22) and (2.23)] behaves distinctly different compared to the original robust formulation [based on equations (2.19) and (2.20)] for the following reasons:

- ✓ First, because the subtractions in the objective function values in equation (2.23) are with respect to the individual objective function values $J(\mathbf{u}^\ell, \mathbf{m}_i)$ and not with respect to the mean.
- ✓ Second, because for bound-constrained control problems, \mathbf{u}^ℓ and $\bar{\mathbf{u}}$ may be shifted with respect to each other.
- ✓ Thirdly, because the effect of outliers which may strongly influence the mean value used in our least-squares approach to estimate the gradient is reduced in the modified formulation. The vector \mathbf{j} can be viewed as a weighting factor in the gradient estimate which is significantly different in the modified formulation compared to the original formulation.

In the next section we provide a theoretical reasoning which justifies the use of the modified formulation for optimization problems.

2.2.2.1 Use of Different Ratio's

All the gradient formulations i.e. equations (2.11)-(2.16) for deterministic optimization are also applicable for robust optimization. Together with the robust modified formulation [equations (2.22) and (2.23)] this leads to a total of 18 potential robust gradient formulations for the 1:1 ratio (i.e. one control perturbation for each geological realization) approach. However, another distinction can be made if we use other ratios. E.g., Raniolo et al. (2013) suggest the use of 20 control perturbations for every model realization. For every model realization, using the 1:20 ratio, they estimate an individual gradient, where after they take the mean of the individual gradients to obtain the robust gradient. This formulation will hereafter be referred to as the ‘Mean of Individual Gradients’ (MIG). Alternatively, one can combine all the controls and objective function anomalies to estimate a single robust gradient, i.e. not estimate individual gradients for every model realization. This approach will hereafter be referred to as the ‘Hotch-Potch Gradient’ (HPG). This additional distinction leads to a total of 30 potential formulations [2 times 18 minus 6 because for the MIG approach there is no difference between using equations (2.18) and (2.23)].

2.3 Theoretical Understanding of EnOpt for Robust Optimization

In this section we provide a theoretical reasoning for the inferior results obtained with the original robust EnOpt formulation and a justification for the use of the modified robust EnOpt formulation described above. In this section we consider the problem of finding \mathbf{u} which maximizes the expectation over \mathbf{m} , which is a random vector with a known probability density function (pdf), of a nonlinear objective function J . Throughout we assume that the uncertainty in \mathbf{m} can be represented by sampling its pdf to obtain an ensemble of N_e realizations, \mathbf{m}_i , $i = 1, 2, \dots, N_e$ and

thus approximate the expectation of J with respect to \mathbf{m} as the mean of the set $[J(\mathbf{u}, \mathbf{m}_i)]_{i=1}^{N_e}$, i.e.,

$$J_E(\mathbf{u}) = \frac{1}{N_e} \sum_{k=1}^{N_e} J(\mathbf{u}, \mathbf{m}_k), \quad (2.24)$$

where $J_E(\mathbf{u})$ denotes the approximation of the expectation $E_{\mathbf{m}}[J(\mathbf{u}, \mathbf{m})]$. At iteration ℓ of EnOpt, we generate N_e control samples, using equation (2.5) the mean of the control samples is given by equation (2.6). For the robust original formulation, at each iteration ℓ , Chen et al. (2009) define $\bar{J}(\mathbf{u}^\ell, \mathbf{m})$ by

$$\bar{J}(\mathbf{u}^\ell, \mathbf{m}) \equiv \frac{1}{N_e} \sum_{i=1}^{N_e} J(\mathbf{u}_i^\ell, \mathbf{m}_i). \quad (2.25)$$

Although Chen et al. (2009) do not give a specific interpretation of equation (2.25), the right side of (2.25) can be interpreted as an approximation of the mean of J , \mathbf{m} and \mathbf{u} . Note that at each iteration, the original EnOpt formulation uses one perturbation of \mathbf{u}^ℓ per reservoir model. In this original formulation, Chen (2008) computed a singly-smoothed search direction as

$$\mathbf{d}^\ell = \frac{1}{N_e - 1} \sum_{i=1}^{N_e} (\mathbf{u}_i^\ell - \bar{\mathbf{u}}^\ell) (J(\mathbf{u}_i^\ell, \mathbf{m}_i) - \bar{J}(\mathbf{u}^\ell, \mathbf{m})). \quad (2.26)$$

The original formulation also suggest a pre-multiplication of equation (2.26) by $\tilde{\mathbf{C}}$, see equations (2.11)-(2.13). In an attempt to show that \mathbf{d}^ℓ defined in equation (2.26), represents a reasonable approximation of \mathbf{C}_{uu} times the true gradient \mathbf{g} , is desirable, Chen et al. (2009) make two assumptions, the first of which is that

$$\bar{\mathbf{u}}^\ell = \frac{1}{N_e} \sum_{i=1}^{N_e} \mathbf{u}_i^\ell \approx \mathbf{u}^\ell. \quad (2.27)$$

Because the set $\{\mathbf{u}_i^\ell\}_{i=1}^{N_e}$ are samples from $\mathcal{N}(\mathbf{u}^\ell, \tilde{\mathbf{C}})$, $\bar{\mathbf{u}}^\ell$ defined in equation (2.5) is an unbiased, consistent estimator of \mathbf{u}^ℓ which is a good approximation for N_e sufficiently large. However, this approximation may actually be inaccurate if the upper and/or lower bounds on \mathbf{u} force truncation of the samples. For example, suppose that the controls are simply the liquid rates at producers and at iteration ℓ some of these controls are close to the lower bound of zero and some are close to the upper bound. Then a sample \mathbf{u}_i^ℓ may have some negative entries which have to be truncated to zero in order to run the simulator to evaluate the objective function. Depending on the implementation and physics, it may also be necessary to truncate entries of \mathbf{u}_i^ℓ that exceed the upper bound to the upper bound, for e.g. when controls are Inflow Control Valve (ICV) settings. In these circumstances, equation (2.27) becomes an invalid approximation because the truncated versions of \mathbf{u}_i^ℓ are no longer samples from the distribution $\mathcal{N}(\mathbf{u}^\ell, \tilde{\mathbf{C}})$. The second assumption of Chen et al. (2009) is that for any \mathbf{m} ,

$$\bar{J}(\mathbf{u}^\ell, \mathbf{m}) = \frac{1}{N_e} \sum_{i=1}^{N_e} J(\mathbf{u}_i^\ell, \mathbf{m}_i) \approx J(\bar{\mathbf{u}}^\ell, \mathbf{m}) \approx J(\mathbf{u}^\ell, \mathbf{m}) \quad (2.28)$$

and the third assumption is that for any \mathbf{m} and \mathbf{u} ,

$$J(\mathbf{u}, \mathbf{m}) = J(\mathbf{u}^\ell, \mathbf{m}) + (\nabla J(\mathbf{u}^\ell, \mathbf{m}))^T (\mathbf{u} - \mathbf{u}^\ell). \quad (2.29)$$

The third assumption, equation (2.29) is simply a standard Taylor series approximation but the second assumption equation (2.28) is tenuous as it suggests that for any realization \mathbf{m}_i , $J(\mathbf{u}^\ell, \mathbf{m}_i) = \bar{J}(\mathbf{u}^\ell, \mathbf{m})$ from which it follows that

$$J(\mathbf{u}^\ell, \mathbf{m}_1) \approx J(\mathbf{u}^\ell, \mathbf{m}_2) \approx \dots \approx J(\mathbf{u}^\ell, \mathbf{m}_{N_e}), \quad (2.30)$$

which is clearly an invalid approximation unless the variance in the prior model for the random vector \mathbf{m} is sufficiently small so that when the vector of controls \mathbf{u}^ℓ is applied to each of these models, the same objective function value is obtained.

2.3.1 Modified Formulation : Theoretical Reasoning

Instead of relying on the potentially unreliable approximations given in equations (2.27) and (2.28), it would be far better to use a search direction that does not rely on these approximations. Following the modified formulation introduced in Fonseca et al. (2014), we define a foundational search direction as

$$\mathbf{d}^\ell = \frac{1}{N_e} \sum_{k=1}^{N_e} \left(\frac{1}{N_p} \sum_{i=1}^{N_p} (\mathbf{u}_i^\ell - \mathbf{u}^\ell) (J(\mathbf{u}_i^\ell, \mathbf{m}_k) - J(\mathbf{u}^\ell, \mathbf{m}_k)) \right), \quad (2.31)$$

where N_p is the number of independent random of samples of $(\mathbf{u}^\ell, \tilde{\mathbf{C}})$ used in equation (2.31). Thus at each optimization iteration, we use N_p random samples to generate an approximation to $\tilde{\mathbf{C}} \nabla_{\mathbf{u}} J(\mathbf{u}^\ell, \mathbf{m}_k)$. We could use a different set of N_p samples for each \mathbf{m}_k or for computational efficiency, we can use the same N_p samples to generate the approximation of $\tilde{\mathbf{C}} \nabla_{\mathbf{u}} J(\mathbf{u}^\ell, \mathbf{m}_k)$ for all k . If we use a different set of control perturbations for each \mathbf{m}_k , $k = 1, 2, \dots, N_e$, then we have to change notation slightly, i.e., we let $\mathbf{u}_{i,k}^\ell, j = 1, 2, \dots, N_e$ denote the set of samples used to generate an approximate gradient of $J(\mathbf{u}^\ell, \mathbf{m}_k)$. In this case, equation (2.31) should be replaced by

$$\mathbf{d}^\ell = \frac{1}{N_e} \sum_{k=1}^{N_e} \left(\frac{1}{N_p} \sum_{i=1}^{N_p} (\mathbf{u}_{i,k}^\ell - \mathbf{u}^\ell) (J(\mathbf{u}_{i,k}^\ell, \mathbf{m}_k) - J(\mathbf{u}^\ell, \mathbf{m}_k)) \right). \quad (2.32)$$

Assuming all second derivatives of J are continuous and bounded, the error in the first order Taylor series

$$J(\mathbf{u}_{i,k}^\ell, \mathbf{m}_k) = J(\mathbf{u}^\ell, \mathbf{m}_k) + (\nabla_{\mathbf{u}} J(\mathbf{u}^\ell, \mathbf{m}_k))^T (\mathbf{u}_{i,k}^\ell - \mathbf{u}^\ell) \quad (2.33)$$

satisfies

$$e_{i,k} = \frac{1}{2}(\mathbf{u}_{i,k}^\ell - \mathbf{u}^\ell) \mathbf{H}_k^\ell (\mathbf{u}_{i,k}^\ell - \mathbf{u}^\ell) = \mathcal{O} \|\mathbf{u}_{i,k}^\ell - \mathbf{u}^\ell\|_2^2$$

where \mathbf{H}_k^ℓ denotes the Hessian matrix given by $\mathbf{H}_k^\ell = \nabla_{\mathbf{u}}[(\nabla_{\mathbf{u}} J(\mathbf{u}^\ell, \mathbf{m}_k))^T]$. Using equation (2.33) in equation (2.32) gives the approximation

$$\mathbf{d}^\ell = \frac{1}{N_e} \sum_{k=1}^{N_e} \left(\frac{1}{N_p} \sum_{j=1}^{N_p} (\mathbf{u}_{i,k}^\ell - \mathbf{u}^\ell) (\mathbf{u}_{i,k}^\ell - \mathbf{u}^\ell)^T \nabla_{\mathbf{u}} J(\mathbf{u}^\ell, \mathbf{m}_k) \right). \quad (2.34)$$

We let $\mathbf{E}_{\mathbf{u}}$ denote the expectation with respect to \mathbf{u} . Then, taking the expectation of equation (2.34) assuming that all $\mathbf{u}_{i,k}^\ell$ are samples from $\mathcal{N}(\mathbf{u}^\ell, \tilde{\mathbf{C}})$ gives

$$\begin{aligned} \mathbf{E}_{\mathbf{u}}[\mathbf{d}^\ell] &= \frac{1}{N_e} \sum_{k=1}^{N_e} \left[\frac{1}{N_p} \sum_{j=1}^{N_p} (\mathbf{E}_{\mathbf{u}}[(\mathbf{u}_{i,k}^\ell - \mathbf{u}^\ell) (\mathbf{u}_{i,k}^\ell - \mathbf{u}^\ell)^T]) \nabla_{\mathbf{u}} J(\mathbf{u}^\ell, \mathbf{m}_k) \right], \\ &= \mathbf{C}_{uu} \left(\frac{1}{N_e} \nabla_{\mathbf{u}} J(\mathbf{u}^\ell, \mathbf{m}_k) \right) = \mathbf{C}_{uu} \nabla J_E(\mathbf{u}^\ell), \end{aligned} \quad (2.35)$$

where $J_E(\mathbf{u}^\ell)$ is the approximation of the expectation of $J(\mathbf{u}^\ell, \mathbf{m})$ with respect to \mathbf{m} as defined by equation (2.24). It is important to note that as \mathbf{m} is a constant we do not need any derivative w.r.t. \mathbf{m} and the error in equations (2.34) and (2.35) is $\mathcal{O}(\max_{i,k} \|\mathbf{u}_{i,k}^\ell - \mathbf{u}^\ell\|_2^2)$. We refer to the search direction of equation (2.32) [or equation (2.31)] as StoSAG (Stochastic Simplex Approximate Gradient). We see that the error in the StoSAG gradient goes to zero as $\|\mathbf{u}_{i,k}^\ell - \mathbf{u}^\ell\|_2^2 \rightarrow 0$ which is satisfied if $\mathbf{u}_{i,k}^\ell \sim \mathcal{N}(\alpha_\ell, \tilde{\mathbf{C}})$ where $\alpha_\ell \rightarrow 0$ as $\ell \rightarrow \infty$ with probability 1.

On the other hand, following the same reasoning, Fonseca et al. (2015) show that for the original EnOpt formulation $\mathcal{O}(\max_i \{(\|\mathbf{u}_i^\ell - \mathbf{u}^\ell\|_2 + \|\mathbf{m}_i - \bar{\mathbf{m}}\|_2)^2\})$ represents the error. Thus if the uncertainty is large, i.e., $(\|\mathbf{m}_i - \bar{\mathbf{m}}\|_2)^2$ is large, then the error will be large which is another reason why the original EnOpt formulation is less ‘robust’ compared to StoSAG. Here we have obtained a search direction which approximates $\mathbf{C}_{uu} \nabla J_E(\mathbf{u}^\ell)$ without invoking the potentially invalid assumptions of Chen et al. (2009) represented by equations (2.27) and (2.28). It is also important to note that one of the advantages of the equation (2.28) used by Chen et al. (2009) is its computational efficiency which arises that equation (2.28) uses only one perturbation of the control vector \mathbf{u}^ℓ per model realization and hence requires only N_e simulation runs of $J(\mathbf{u}, \mathbf{m})$. On the other hand, the theoretically sound search direction of equation (2.32) requires $N_p \times N_e$ evaluations of $J(\mathbf{u}, \mathbf{m})$ i.e. $N_p \times N_e$ simulation runs. However, one can select $N_p=1$ when applying equation (2.32) and in this case, StoSAG has the same computational efficiency as standard EnOpt. Note that equation (2.35) applies for $N_p=1$. Moreover, Fonseca et al. (2014) and Fonseca et al. (2015) show that equation (2.32) with $N_p=1$ leads to a methodology that results in a higher net-present-value than

is achievable with the original EnOpt formulation proposed by Chen (2008). Although Fonseca et al. (2015) indicate that far more accurate estimates of $\mathbf{C}_{uu} \nabla J_E(\mathbf{u}^\ell)$ can be obtained by using $N_p \gg 1$, when using $N_p \gg 1$ not only does the computational effort increase substantially it is practically impossible to know a-priori what value of N_p should be considered and thus a user defined choice needs to be made.

2.3.2 Justification for the ‘Simplex Gradient’

Fonseca et al. (2015) observe that higher net-present values can be obtained if instead of using equation (2.32) to form a steepest ascent search direction, one instead generates an approximation for $\nabla J_E(\mathbf{u}^\ell)$ to use as a search direction. This suggestion may be particularly important when the bounds on \mathbf{u} force the truncation of some components of a sample $\mathbf{u}_i^\ell \sim \mathcal{N}(\mathbf{u}^\ell, \tilde{\mathbf{C}})$ and whenever it is unclear what correlation length to choose in a covariance function in order to generate the covariance matrix $\tilde{\mathbf{C}}$ or when the optimal solution of the well control problem is such that the values of two temporally consecutive controls at a well are radically different, as occurs, for example, when the optimal solution is bang-bang see, Sudaryanto and Yortsos (2001), Zandvliet et al. (2007) etc. When the $\mathbf{u}_{i,k}^\ell$ used in equation (2.32) are truncated,

$$E_{\mathbf{u}} \left[\frac{1}{N_p} \sum_{i=1}^{N_p} (\mathbf{u}_{i,k}^\ell - \mathbf{u}^\ell)(\mathbf{u}_i^\ell - \mathbf{u}^\ell)^T \right] \neq \tilde{\mathbf{C}} \quad (2.36)$$

and equation (2.35) is no longer valid. However, one can still obtain an approximation of $\mathbf{C}_{uu} \nabla J_E(\mathbf{u}^\ell)$ by first using a simplex gradient, see Conn et al. (2009), to approximate $\nabla J_E(\mathbf{u}_\ell)$ and then multiplying this gradient by $\tilde{\mathbf{C}}$. In line with the notation used in the preceding section, we define the $N_u \times N_p$ matrix \mathbf{U}^ℓ and the $N_p \times 1$ vector $\mathbf{j}^{\ell,k}$, for $k=1, 2, \dots, N_e$ respectively, by

$$\mathbf{U}^{\ell,k} = [\mathbf{u}_{1,k}^\ell - \mathbf{u}^\ell \quad \mathbf{u}_{2,k}^\ell - \mathbf{u}^\ell \quad \dots \quad \mathbf{u}_{N_p,k}^\ell - \mathbf{u}^\ell], \quad (2.37)$$

$$\mathbf{j}^{\ell,k} = [J(\mathbf{u}_{1,k}^\ell, \mathbf{m}_k) - J(\mathbf{u}^\ell) \quad \dots \quad J(\mathbf{u}_{N_p,k}^\ell, \mathbf{m}_k) - J(\mathbf{u}^\ell)]. \quad (2.38)$$

Then using the first order Taylor series of equation (2.33) and the definitions of equations (2.37) and (2.38), it easily follows that for each $k, k=1, 2, \dots, N_e$.

$$\begin{aligned} \frac{1}{N_p} \sum_{i=1}^{N_p} (\mathbf{u}_{i,k}^\ell - \mathbf{u}^\ell)(J(\mathbf{u}_{i,k}^\ell, \mathbf{m}_k) - J(\mathbf{u}^\ell, \mathbf{m}_k)) &= \mathbf{U}^{\ell,k} (\mathbf{j}^{\ell,k})^T, \\ \mathbf{U}^{\ell,k} (\mathbf{j}^{\ell,k})^T &\approx [\mathbf{U}^{\ell,k} (\mathbf{U}^{\ell,k})^T] \nabla_{\mathbf{u}} J(\mathbf{u}^\ell, \mathbf{m}_k) \end{aligned} \quad (2.39)$$

Thus, as in Fonseca et al. (2015) and Do and Reynolds (2013) an approximate stochastic gradient, denoted by $\nabla_{\mathbf{u}} J(\mathbf{u}^\ell, \mathbf{m}_k)$ is given by

$$\nabla_{\mathbf{u}} J(\mathbf{u}^\ell, \mathbf{m}_k) = [\mathbf{U}^{\ell,k} (\mathbf{U}^{\ell,k})^T]^{-1} \mathbf{U}^{\ell,k} (\mathbf{j}^{\ell,k})^T \approx \nabla_{\mathbf{u}} J(\mathbf{u}^\ell, \mathbf{m}_k), \quad (2.40)$$

for $k = 1, 2, \dots, N_e$ and instead of computing the inverse we use $(\mathbf{U}^{\ell,k})^\dagger = [\mathbf{U}^{\ell,k} (\mathbf{U}^{\ell,k})^T]^{-1} \mathbf{U}^{\ell,k}$ where the superscript \dagger denotes the Moore-Penrose pseudo-inverse which can be obtained by singular value decomposition see Golub et al. (1998). Except for the fact that this gradient is stochastic because the \mathbf{u} 's are random vectors, the term $\nabla_{\mathbf{u}} J(\mathbf{u}^\ell, \mathbf{m}_k)$ is simply the simplex gradient of J at $(\mathbf{u}^\ell, \mathbf{m}_k)$, see Conn et al. (2009). It is important to note that the approximation of equation (2.39) applies regardless of how the perturbations around \mathbf{u}^ℓ are generated. Note that, this straight gradient can be multiplied by $\tilde{\mathbf{C}}$ to obtain a single smoothed search direction which is distinctly different compared to the search direction that is an approximation of \mathbf{C}_{uu} times the true gradient given by equation (2.35). In addition to properly accounting for truncation of samples from $\mathcal{N}(\mathbf{u}^\ell, \tilde{\mathbf{C}})$, the straight StoSAG gradient, i.e., equation (2.40) also avoids problems that can occur if we force temporal correlations on the controls at a well for a case where the true optimal controls vary significantly from one control step to the next as occurs for the bang-bang control case.

Thus we have shown that StoSAG is theoretically more sound compared to the original formulation and using the simplex gradient formulation can be beneficial for problems usually encountered in petroleum engineering application. In this thesis the StoSAG formulation in some chapters is referred to as the ‘modified’ EnOpt formulation. A comparison of the original EnOpt formulation for robust optimization and the StoSAG or modified EnOpt formulation for robust optimization for different models within an optimization context is shown in chapters 3,5 and 7.

3

EFFECT OF ENSEMBLE SIZE ON GRADIENT QUALITY

Ensemble Optimization (EnOpt) has been shown to be a practical method to solve reservoir production optimization problems. Rather than relying on the availability of exact gradients, it involves the evaluation of approximate gradients that are estimated from an ensemble of perturbed controls and the corresponding objective function values. Fonseca et al. (2014) and Forouzanfar et al. (2015) showed that the size of the control perturbations, and the correlations between them, have a significant impact on the efficiency of the optimization process, as well as on the final objective function value. Adaptive perturbation schemes furthermore suggest that, in order to improve convergence, different perturbation sizes may be needed during initial and late stages. The relationship between the choice of perturbation covariance matrix, gradient regularization method and gradient quality, however, has not been explicitly determined. It is known that also ensemble size plays an important role, with larger ensemble sizes leading to better approximations of the exact gradient, but at a higher computational cost. Here, we introduce an experimental framework² to numerically determine the ensemble gradient quality for variable ensemble size, perturbation size, and relative proximity of the control guess to the optimum. The gradient quality is established with a user-defined confidence level as the mean error with the exact gradient, as computed with e.g. the adjoint technique. This framework also enables us to investigate the impact of these parameters on the quality of the gradient in the presence of uncertainty, as represented by an ensemble of geological realizations.

² This chapter is from, Fonseca, R.M., Kahrobaei, S., Van Gastel, L.J., Leeuwenburgh, O. and Jansen, J.D., 2015. Quantification of Impact of Ensemble size on Ensemble Gradient Quality using principles from Hypothesis Testing. Under Review at *SPE Journal*

Various formulations of the robust gradient, that is, the gradient of the expected objective function, are compared. A motivation for the chosen approach is first presented through a set of experiments on the well-known Rosenbrock function, and on an extension with uncertainty in the function parameters. Deterministic optimization experiments are performed with variable degrees of gradient quality, as determined by a specified directional error. A numerical approach is subsequently introduced to determine the ensemble size required to obtain gradients with the desired quality. This approach is also applied to a reservoir model with large uncertainty in the geological properties. Different robust gradient estimators are compared in combination with variable ensemble and perturbation size, where we differentiate between early and late stages of the optimization process. It is shown that a recently proposed modification of the EnOpt scheme for robust cases leads to a strong reduction in the required ensemble size for the same gradient quality. Further improvements can be obtained when multiple controls samples are applied to each realization, especially during the near-convergence stages of the optimization, when the number of control samples is increased by a factor 10. These results provide practical guidelines for application of ensemble optimization to complex reservoir models at acceptable computational cost and illustrates the impact of gradient quality on an optimization experiment.

3.1 Introduction

Gradient-based algorithms have shown to be both effective and computationally efficient in model-based workflows for hydrocarbon recovery or net present value (NPV) optimization. In spite of drawbacks such as computational efficiency and inaccuracy of the gradient estimate, recently many studies for e.g. Chen et al. (2009), Chen and Oliver (2010), and Leeuwenburgh et al. (2010) amongst others have demonstrated the applicability of EnOpt for large-scale production optimization problems. The major advantages of EnOpt are its ease of implementation, flexibility to adapt to different control types such as injection rates, bottom hole pressures, settings of inflow control devices etc., and ability to be used with any reservoir simulator.

Van Essen et al. (2009) introduced a ‘robust optimization’ methodology in conjunction with the adjoint method to include the effect of uncertainties into the optimization framework using an ensemble of equi-probable reservoir models. Chen (2008) proposed a formulation of the EnOpt algorithm that could be applied to an ensemble of models i.e. for robust optimization. While this formulation is computationally very attractive in terms of the number of simulations required to obtain a gradient estimate for a given number of model realizations, it is not clear if these computational advantages outweigh possible disadvantages from the perspective of the gradient quality. Recently, Fonseca et al. (2014) presented a case wherein the

original formulation for ensemble-based robust optimization leads to inferior results and suggested a modified gradient formulation that was shown to produce improved results. These results were not analyzed in terms of the quality of the gradient, however, and alternative gradient formulations were not considered.

The impact of the covariance matrix on the optimization process when using EnOpt has been investigated recently by Fonseca et al. (2015). Their experiments suggested that different perturbation sizes may be required at early and late stages of the optimization process to achieve the best overall performance. A theoretical foundation for the use of a covariance matrix updating scheme was provided by Stordal et al. (2015). Sarma and Chen (2014) investigated the impact of alternative sampling techniques on the quality of the gradient estimate. However, none of these studies have performed a detailed investigation into the effect of ensemble size on the estimated ensemble gradient quality. Given the potentially large implications of poor choices for the optimization inputs in terms of computational cost and quality of the result, there is a clear need for practical guidelines, supported by analysis of underlying theory and of empirical results.

In this chapter we apply statistical analysis to quantify the relationship between ensemble size and gradient quality. We will focus especially on the robust optimization problem in which the number of control samples applied to each model realization may also be varied, leading to three different gradient algorithms, which will be compared. The analysis will be applied to both the initial stages of the optimization process, characterized by a fast increase in the objective function value, as well as to the late stages, typically characterized by slow convergence to the optimal value. In the remainder of this chapter we will first introduce the problem through a number of experiments performed with the well-known Rosenbrock function. A medium size reservoir model is subsequently introduced that forms the basis for further experiments in which an ensemble of realizations of geological properties is utilized to represent uncertainty. The quality of the estimated gradient resulting from the different robust gradient formulations will be compared, and related to the performance of an adjoint based optimization experiment. The impact of gradient quality on the optimization process is also investigated for deterministic optimization cases and a comparison of the optimization between adjoint based and EnOpt gradient formulation for robust optimization is investigated.

3.2 Theory

Chapter 2 contains a detailed overview of the ensemble optimization method used in this chapter. Most of the equations referred to in the results section can be found in chapter 2.

3.2.1 Adjoint method

The adjoint method has been investigated extensively for use in data assimilation and production optimization. Detailed derivations for the production optimization case can be found in, e.g., Brouwer and Jansen (2004), Sarma et al. (2005), Kraaijevanger et al. (2007) and Jansen (2011). The adjoint method is the most accurate and computationally efficient method for computing a gradient. Computation of the gradient only requires one forward simulation and one fast backward computation. Therefore the number of simulation runs is independent of the number of controls. However for robust optimization using the adjoint requires running the forward and backward simulation for every geological realization, thus to compute the robust gradient, the same number of simulation runs will be performed as required for the robust EnOpt gradient using the 1:1 ratio. For our experiments we assume that the adjoint gradient is the exact gradient, which the EnOpt method tries to approximate. In this study the adjoint module available in the Shell in-house simulator was used (Kraaijevanger et al. 2007).

3.2.2 Hypothesis testing

We use principles from hypothesis testing to validate the research goal of this paper, namely to test if the approximate EnOpt gradient is comparable in quality to the adjoint gradient. To be able to determine the difference in gradients we compute the angle between them by using the dot product:

$$\cos(\theta) = \frac{\mathbf{g}_{adj} \cdot \mathbf{g}_{ens}}{\|\mathbf{g}_{adj}\| \|\mathbf{g}_{ens}\|}. \quad (3.1)$$

the null hypothesis used is

$$H_0 : \theta \leq 10^\circ. \quad (3.2)$$

The statistical inference method used is based on pre-defined confidence intervals for the testing parameters defined above.

3.2.2.1 Confidence intervals

Creating a confidence interval is a method to define a range at and the certainty that the true value of an estimated parameter lies within it, based on the knowledge of the sampling distribution (Dekking et al., 2005). In our numerical experiments we create a dataset of the parameter, given in equation (3.1). The parameter of interest θ is the maximum allowable deviation of the EnOpt gradient with regard to the exact or adjoint gradient. As it is virtually impossible to achieve a 100% confidence, we apply a

confidence level of $\gamma = 0.95$. The general definition of a confidence interval assumes a two-sided interval, i.e. an upper and a lower limit. However it is also possible to have a one-sided interval. As the test parameters used for the numerical experiments are absolute values of deviations we only want to find the confidence interval of the maximum deviation, thus the upper limit. Using a one-sided interval, the confidence interval is just the integral of the probability distribution, i.e. the cumulative density function (CDF).

3.2.2.2 Methodology with Traditional Hypothesis Testing Principles

The EnOpt method samples random points from a normal distribution with a user-defined standard deviation. The distribution of the test parameter(θ) is, however, not normally distributed due to the non-linearity of the objective function and the function for the test statistics. In order to determine a confidence interval the distribution of the underlying parameters needs to be known. To test our null hypothesis we use a traditional hypothesis testing approach to either accept or falsify our hypothesis and estimate the necessary ensemble size as follows:

- Sample points in the control space and compute the adjoint and EnOpt gradients at each point.
- Count the number of points that satisfy the null hypothesis
- Compute the confidence interval for the predefined maximum allowable error.
- Repeat for varying ensemble sizes until the confidence interval is achieved.

We have chosen the 95% confidence interval to test our methodology, however any different confidence interval can be chosen within the same workflow. Varying the desired confidence interval will automatically vary the ensemble size needed to accept our hypothesis. In essence an ensemble size is found that gives an accurate gradient approximation in a number of times equal to the confidence interval, i.e. if the numerical experiment would be repeated many times, 95% of the experiments would result in an approximate gradient within the error margin compared to the true (adjoint) gradient. For a more detailed explanation of confidence intervals see, e.g. Dekking et al. (2005).

3.3 Numerical Example: Rosenbrock Function

The methodology is first applied to the non-linear Rosenbrock function named after the mathematician who first used it to demonstrate his optimization algorithm, see Rosenbrock (1960). This analytical function has since been used as a standard test case in mathematical optimization. The Rosenbrock function consists of a curved narrow valley which most algorithms have little difficulty finding. However once found, the

difficulty lies in finding the global optimum which is situated inside the valley. Equation (3.3) is a slightly altered version, as it is multiplied by -1 , making it a maximization problem opposed to a minimization problem:

$$J(u_1, u_2) = -100(u_2 - u_1^2)^2 - (1 - u_1)^2. \quad (3.3)$$

The Rosenbrock function has an optimal solution $J = 0$ at $(u_1, u_2) = [1 \ 1]$. This point lies on a long curved ridge; see **Fig. 3.1** for a contour plot. Since it is an analytical function it is possible to compute the exact gradient. The red dots in Fig. 3.1 are 50 randomly distributed points in space which will be used to test our methodology. All the numerical experiments will be carried out over the same set of points for different scenarios. Since we are working with approximate gradient techniques the effect of different random sequences also needs to be accounted for. Therefore, all the experiments are repeated for 50 different random sequences. The results presented below are the mean values of the angles at the 50 points in space over 50 different random sequences.

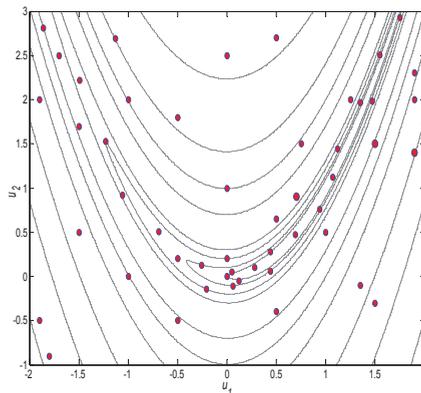


Fig. 3.1: Contour plot with value labels of Rosenbrock function given by equation (3.3). Red dots are 50 points randomly distributed in space.

3.3.1 Deterministic Case

We first test the hypothesis testing methodology which entails the calculation of the angle between the gradient estimated by EnOpt and the exact gradient for different ensemble sizes. The points are uniformly distributed so as to capture the effect of the spatial variability in the objective space on the gradient quality, with many points that are on the ridge or on the edge of the ridge. This will to some degree ensure that once the right ensemble size is determined the EnOpt gradient will be comparable to the true gradient almost everywhere in space. The other important factor in determining a high-quality ensemble gradient is the perturbation size used to generate the ensemble of

controls for the gradient estimate. **Fig. 3.2** is the mean angle over the 50 spatial points and 50 random seeds versus an increasing ensemble size which shows the impact of ensemble size for varying perturbation sizes σ . We observe that, as expected, as σ decreases the gradient quality increases for smaller ensemble sizes while for an increasing ensemble size the gradient quality always improves. The Rosenbrock function is a two-control problem, thus we would need a maximum of three (i.e. $n+1$, where n is the problem dimension) simulations to estimate a ‘good’ gradient with the finite difference method provided the perturbation size is sufficiently small. Fig. 3.2 is a reconfirmation that EnOpt, for a small perturbation, also would require an ensemble size equal to 3 to estimate a ‘good’ gradient. For the largest value of σ used we observe that even for an ensemble size of 300 we do not achieve a ‘good’ gradient at all the 50 points in space.

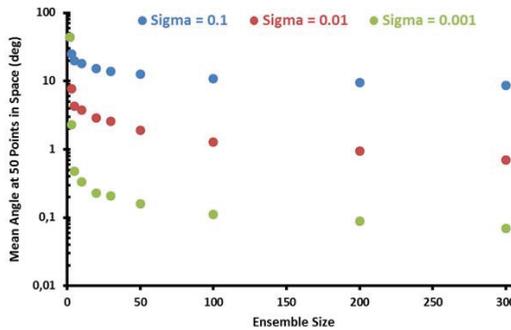


Fig. 3.2: Illustration of the trend in the mean angle over the 50 points in space and 50 random seeds for an increasing ensemble size and varying perturbation sizes (σ). Lower values (green & red dots) of σ give significantly better results compared to higher values (blue dots).

We count that 38 of the 50 points satisfy our null hypothesis (38/50), i.e. 76% of the points, satisfy our hypothesis, and hence, with an ensemble size of 300 and a perturbation size (σ) equal to 0.1, we have a confidence interval of 76%. We also observe that for an ensemble size equal to 2 we cannot achieve good results and this is also the case for the different perturbation sizes. To achieve the confidence interval defined in this work we would need approximately 900 samples when using $\sigma = 0.1$. We observe that for $\sigma = 0.01$ with an ensemble size of 5 samples we achieve a 100% confidence interval while for $\sigma = 0.001$, i.e. one order of magnitude smaller, we achieve a 100% confidence interval with an ensemble size of 3 samples.

The results show that gradient quality is strongly affected by the perturbation size which in turn determines the ensemble size needed to achieve good quality gradients. We observe that, for large perturbation sizes, we need significantly larger ensemble sizes to achieve the 95% confidence interval. For significantly smaller perturbation sizes we would need an ensemble size of 3, i.e. for an n dimensional problem we need

an ensemble size equal to $n+1$, just like if we were to estimate a finite difference gradient.

3.3.2 Rosenbrock Function With Uncertainty

The quality of the ‘robust’ EnOpt gradient has not been investigated before, so uncertainty in the Rosenbrock function is introduced through equation (3.4). The uncertainty is introduced to mimic geological uncertainty.

$$J(u_1, u_2) = -100(c_1 u_2 - u_1^2)^2 - \sin(c_2)(1 - u_1)^2. \quad (3.4)$$

The two constants c_1 and c_2 are drawn from a Gaussian distribution, where the standard deviation reflects the magnitude of the uncertainty introduced. This provides us a fast and accurate way to test the various ‘robust’ EnOpt gradient formulations to better understand these formulations. The value for c_1 mainly affects the steepness of the objective space, while c_2 rotates the space. **Fig. 3.3** and **Fig. 3.4** show five realizations each for cases which are representative of low and high uncertainty scenarios. This was done to investigate the impact of uncertainty on the gradient estimate.

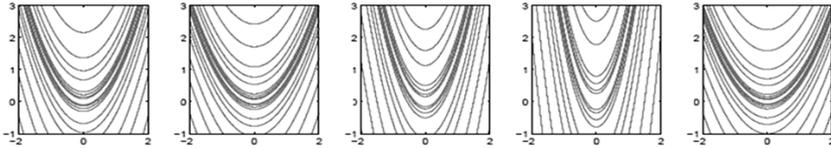


Fig. 3.3: Contour plot of five realizations of the Rosenbrock function given by equation (3.4) for a case mimicking a low degree of uncertainty.

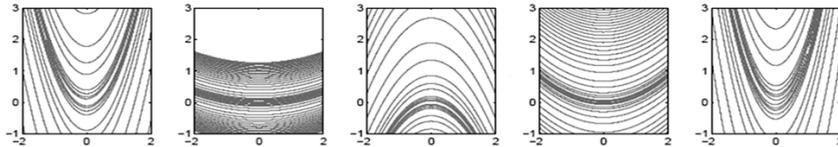


Fig. 3.4: Contour plot of five realizations of the Rosenbrock function given by equation (3.4) for a case mimicking a higher degree of uncertainty.

Chen (2008) first proposed the idea of the 1:1 ratio explained in the theory section to estimate a ‘robust’ gradient. The mathematical reasoning for the applicability of this 1:1 ratio provided by Chen (2008) is only applicable if we have very large ensembles for both the models and consequently the controls. In reality, however, we usually work with 100 models and in our experiments the original 1:1 formulation in many instances indeed produces results of insufficient quality. The poor results were reasoned to be the lack of a good quality gradient. **Fig. 3.5** illustrates the effect of using the 1:1 ratio for an increasing ensemble size of model realizations and, consequently,

also control realizations. We observe, in line with the theory, that the mean angle significantly reduces for an increasing ensemble size. However, even for an ensemble size of 1000 we do not satisfy our null hypothesis when using the original formulation (blue curve). On the other hand the modified formulation based on equations (2.22) and (2.23) (red curve) which still retains the computational attractiveness of the original formulation performs significantly better. We observe that for an ensemble size of 1000 we have a mean angle of approximately 4 degrees and we have satisfied our null hypothesis and achieved the 95% confidence interval.

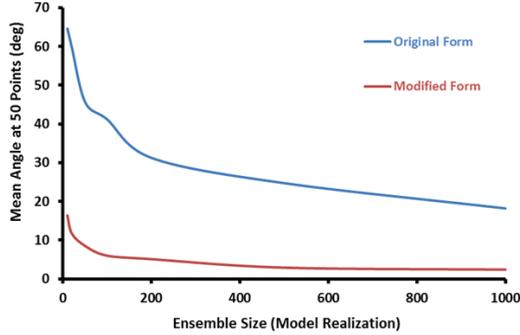


Fig. 3.5: Comparison of the trend in the mean angle over the 50 spatial points with increasing ensemble sizes for the original form (blue) and modified form (red) using the computationally attractive 1:1 ratio.

These results are based on the case representing the highest uncertainty in the models. In our reservoir optimization problem we usually use an ensemble size of 100 models to capture the uncertainty. For that size, the original formulation achieves a mean angle of 40 degrees while the modified formulation achieves a mean angle of 7 degrees. We recommend therefore that, when using the 1:1 ratio for robust optimization, the modified formulation based on equations (2.22) and (2.23) be applied.

3.3.2.1 Effect of Higher Ratios

The results in Fig. 3.5 were for the case representative of the highest model uncertainty using the 1:1 ratio. Some previous studies, e.g. Raniolo et al. (2013) and Li et al. (2013), did not achieve results of practical value with the 1:1 ratio and suggested the use of higher ratios (1:20 etc.) to find better gradient estimates. We investigate here the impact of varying the ratio between model and control perturbations on the quality of the gradient estimate. The results are obtained with the HPG formulation using a perturbation size equal to 0.01. **Fig. 3.6** consists of two plots, Fig. 3.6(a) which is the mean angle plotted against the ensemble size of model realization with the curves representing the ratio used for the gradient estimate with the original formulation, and Fig. 3.6(b) which displays the results using the modified formulation (note the

difference in the vertical scale). In both plots we observe that an increased ratio gives better gradient estimates for the different ensemble sizes of model realizations. Once again, the modified formulation outperforms the original formulation also for larger ratios. Thus in general better gradient estimates can be achieved while using higher ratios. While it is better to use higher ratios, the computational costs of using these ratios must also be accounted for, especially for large-scale high-dimensional problems.

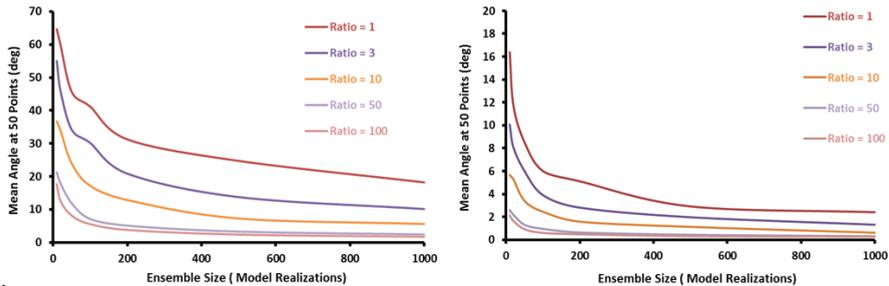


Fig. 3.6: (a). Illustrates the effect of using higher ratios on the gradient quality when using the original formulation with the HPG formulation. (b) Illustrates the effect when using the modified formulation with the HPG formulation. Note the different vertical scales.

3.3.2.2 Effect of Uncertainty

The effect of uncertainty is investigated in conjunction with the ratios needed to satisfy the null hypothesis and achieve the desired confidence interval. **Fig. 3.7** is an illustration of the impact that uncertainty has on the quality of the gradient estimate. Fig. 3.7(a) (left-side plot) depicts the results for the highest uncertainty case while Fig. 3.7(b) (right-side plot) depicts results for the lowest uncertainty scenario for both the original and modified formulations. In both plots we observe, in accordance with all the results thus far, that the modified formulation results in significantly better gradients than the original formulation. Also the modified formulation gradient estimate is less sensitive to the effect of uncertainty. The ensemble size of model realizations was kept constant at 100 for this exercise and a perturbation size of 0.01 was used. We also observe that for the original formulation using a larger perturbation size resulted in better mean angles, whereas for the modified formulation the gradient estimate improved for smaller perturbation sizes. We observe that, depending on the degree of uncertainty, the original formulation would need a ratio of 1:20 or 1:50 i.e. 2000 or 5000 function evaluations to find a mean angle less than 10 degrees. If the mean angle is less than 10 degrees it does not guarantee that we have achieved the necessary confidence interval.

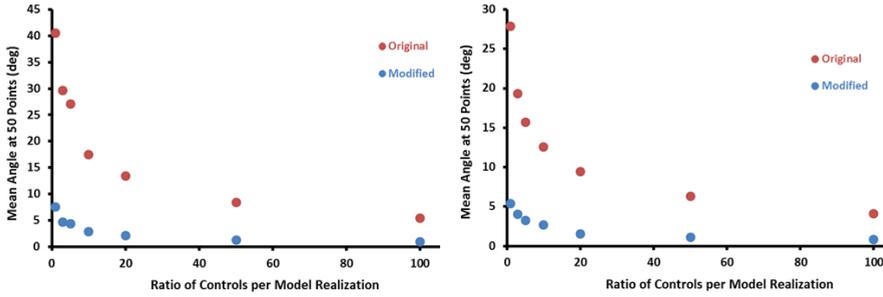


Fig. 3.7: (a) shows the effect of a high degree of uncertainty on the mean angle for increasing ratios with the HPG formulation. (b) Illustrates the results for the lowest uncertainty case. For both these plots an ensemble of 100 model realizations is used. The angles for the high uncertainty case are higher than the low uncertainty case. (Note the different vertical scales).

3.3.2.3 Hypothesis Test Results

Fig. 3.8 illustrates the ratio necessary to satisfy the null hypothesis used in this paper and achieve the desired 95% confidence interval for the different gradient formulations and the different ensemble sizes of model realizations. We observe that for a smaller ensemble size of model realizations, e.g. 10, we need significantly higher ratios (1:1100) when using the original formulation while for the same ensemble size, we would require a ratio of 1:10 with the modified formulation. We also observe that the required ratio decreases with an increase in the ensemble size of model realizations. All these results are for the highest uncertainty scenario. For the lowest uncertainty scenario we would require lower ratios: 1:800 with the original formulation for an ensemble size of 10 model realizations, and a 1:5 ratio for the modified formulation.

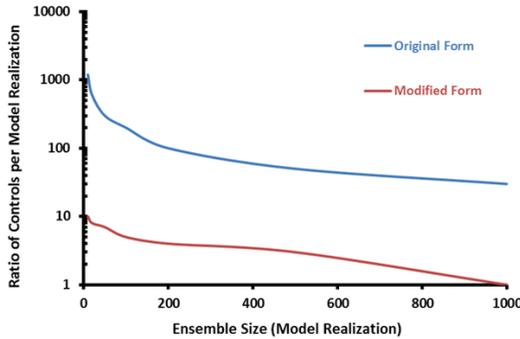


Fig. 3.8: Hypothesis testing results using the traditional methodology for the original (blue) and modified (red) form using the highest uncertainty case with the HPG formulation. A lower ratio is needed for higher ensemble sizes of model realizations to satisfy the hypothesis and achieve the desired confidence interval.

3.3.2.4 Mean of Individual Gradients (MIG)

The results presented in Fig. 3.6- Fig. 3.8, which are obtained using the HPG formulation, illustrate that for small ensemble sizes we would require ratios as high as 1:1100 to satisfy the hypothesis and achieve the desired confidence interval. For a two-dimensional control problem these results are completely counter-intuitive. Thus we test the MIG formulation to estimate a ‘robust’ gradient when using a ratio other than 1:1. This approach was followed by Van Essen et al. (2009) albeit using the adjoint formulation and also by Raniolo et al. (2013) using EnOpt gradients. Since we are dealing with a two dimensional problem we should need 3 function evaluations to estimate a gradient, for a sufficiently small perturbation size, as illustrated in Fig. 3.2. Using the MIG formulation we observe in **Fig. 3.9(a)** that for a perturbation size equal to 0.01 we would at best need a ratio of 1:5 to estimate a mean angle less than 10 degrees. Increasing the ratio will improve the gradient estimate, and the difference between the original and modified formulation is minor at best. These results are a marked improvement in terms of the ratio needed compared to the results with the HPG formulation. The results are based on the highest uncertainty case while the same trend is observed for the low uncertainty case.

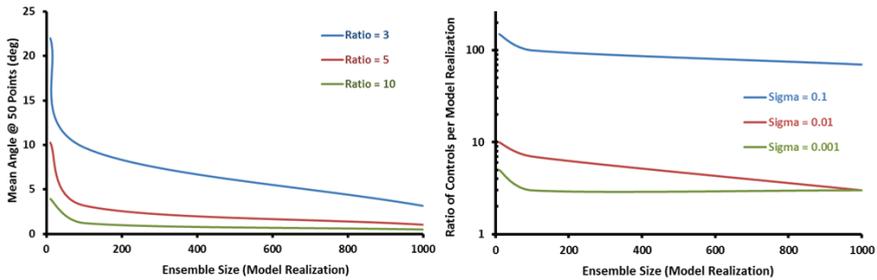


Fig. 3.9: (a) Illustration of the mean angle for increasing ensemble sizes and ratios. We observe a significant reduction in the ratio required even for small ensemble sizes compared to Fig. 3.8 for the highest uncertainty case. (b) Hypothesis testing results using the traditional methodology for different perturbation sizes (σ) and ratios. Lower σ values would require significantly smaller ratio compared to the HPG formulation.

The effect of perturbation size estimates from the MIG formulation is shown in **Fig. 3.9(b)**. We observe, ‘akin’ to the deterministic case, that a decreasing perturbation size results in an increase in the quality of the gradient estimated. For a sufficiently small perturbation size ($\sigma = 0.001$) a ratio of 1:3 is sufficient to satisfy our null hypothesis and achieve the 95% confidence interval even for small ensemble sizes of model realizations. We observe that for an increasing perturbation size the ratio increases, although, it is significantly lower than the ratio required when using the HPG formulation. In summary, for this case using the MIG formulation gives much better

angles for smaller ratios especially when working with small ensembles of model realizations.

3.4 3D Reservoir Model: “Egg Model”

The ‘Egg Model’, first introduced by Van Essen et al. (2009), is a channelized reservoir with seven vertical layers. **Fig. 3.10** is an illustration of the permeability field of a single model realization with the locations of the eight (mainly peripheral) injection wells (blue) and four production wells (red) completed in all the layers. The model represents a channelized depositional system in the form of discrete permeability fields modeled with $60 \times 60 \times 7 = 25.200$ grid cells of which 18.553 cells are active. A detailed description of a standardized version of this Egg Model along with reservoir and fluid properties is given in Jansen et al. (2013). No capillary pressures are included and the reservoir rock is assumed to be incompressible. The bottom hole pressures of the producers are constrained between 385 and 400 bar, while the injectors are rate-controlled between 1 and $79.5 \text{ m}^3/\text{day}$. The initial reservoir pressure is at 400 bars. Production of the field is simulated for a period of 3600 days or slightly less than 10 years. There are 40 control time steps of 90 days, thus using injection rates as controls we have of $40 \times 8 = 320$ controls, i.e. a 320 dimensional problem. The objective function used in this work is NPV given by equation (2.1). We have used an oil price $r_o = 126 \text{ \$/m}^3$, $r_{wp} = 19 \text{ \$/m}^3$ and $r_{wi} = 6 \text{ \$/m}^3$ and a discount rate $b = 0\%$, i.e. undiscounted NPV.

We use a commercial fully implicit black oil simulator (Eclipse, 2011) for the reservoir simulations to estimate the (robust) EnOpt gradient of J with respect to the controls \mathbf{u} . An in-house simulator is used to calculate the adjoint gradient, see Kraaijevanger et al. (2007) for details. The in-house reservoir simulator is used to compute the adjoint-based optimization results while Eclipse (2011) is used to calculate the EnOpt gradient, thus there is no coupling between the different simulators. The model has been benchmarked, i.e. almost identical results in terms of phase rates in the producers, for the different simulators used, details of which can be found in Jansen et al. (2013).

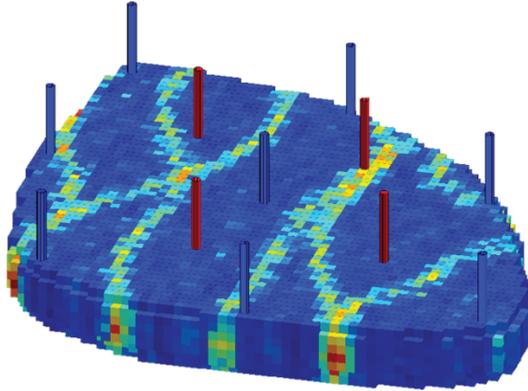


Fig. 3.10: Permeability field of egg model displaying the position of the injectors (blue) and producers (red)

3.4.1 Deterministic case

The methodology presented above is first tested on a single realization of the Egg Model and thus we are dealing with a deterministic case. When working with a 320-control problem it is, of course, impossible to visualize the objective function space. Randomly creating points to evaluate our methodology is also risky since we will not be sure that our points have covered all features of the objective function space, due to the ‘emptiness’ of a 320-dimensional space. Thus, in order to obtain ‘relevant’ points which capture the nature of the objective function space we perform an adjoint- based optimization from two different initial strategies. **Fig. 3.11** shows the optimization process for 40 iterations for an initial strategy of constant rates equal to $79.5 \text{ m}^3/\text{day}$. Thus we have 40 points to test the methodology, i.e., instead of using a large number of randomly distributed control points, as was done for the Rosenbrock function, we now test the gradient quality for 40 points along a pre-defined control trajectory. **Fig. 3.11** also shows two dashed regions, a steep region, indicated with a red ellipse, in which there exists significant scope for optimization, and a relatively flatter region, indicated with the black ellipse, which is indicative of a peak, ridge or plateau in the objective function space. While using the same hypothesis, we investigate the two regions separately, because different parameters are required to achieve the necessary confidence intervals. The two regions are defined arbitrarily based on visual inspection

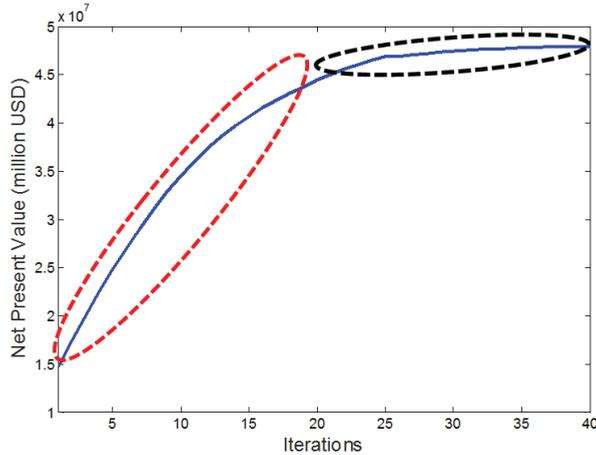


Fig. 3.11: Illustration of the optimization process (blue curve) using adjoint gradients. The curve is divided into two parts, a red ellipse (indicative of the steeper region of the curve) and a black ellipse (indicative of the flatter region of the curve)

3.4.1.1 The Steep Region

Fig. 3.12 displays the results for the points encapsulated by the red ellipse in Fig. 3.11. It depicts the decrease in the mean angle for an increasing ensemble size with a constant perturbation size, $\sigma = 0.01$. The results are the mean of 5 experiments with different random seeds. The left subplot of Fig. 3.12 illustrates the mean angle of the different points in the red ellipse. We observe that we need an ensemble size of approximately 150 to satisfy the null hypothesis. We also observe that increasing the ensemble size leads to higher-quality gradients. Thus for a 320-dimensional problem with approximately 150 samples we can estimate a high-quality gradient. We also tested (results not shown) that for increasing perturbation sizes the quality of the gradient estimate decreases, while for smaller perturbation sizes we also obtain inferior gradient estimates, most likely due to numerical round-off errors. E.g., with $\sigma = 0.1$ (i.e. 10 times larger), and $\sigma = 0.0001$, (i.e. 100 times smaller), we would need 400 and 300 samples respectively to satisfy the null hypothesis and achieve the desired confidence interval. To obtain these results we have used a time-correlated covariance matrix with a correlation length equal to 20 which was arbitrarily chosen. There is no method to determine this value a-priori. A correlation length of 20 means that a control is correlated over 20 control time steps. For the steeper region a correlation length of 20 performs almost equally well as a correlation of 10, thus in the steeper region the impact of correlation length is not significant.

The results presented in Fig. 3.12(a) are obtained with equation (2.14) and the modified formulation suggested by Do and Reynolds (2013) [equations (2.17) and (2.18)]. This performs better than using equation (2.14) with the original formulation

[equations (2.5) and (2.6)]. Irrespective of the gradient formulation used, an increasing ensemble size leads to higher-quality gradients and the modified and original formulation gradients converge for larger ensemble sizes. Fig. 3.12(b) shows the effect of smoothing on the different gradient estimates. We observe that smoothing of the gradient, i.e. pre-multiplication of the gradient by the covariance matrix, has a marginally negative impact on the gradient quality for this set of points when using $\sigma = 0.01$. We also observe that single smoothing of the straight gradient obtained from equation (2.14), i.e. equation (2.15), performs better than using the smoothed gradient given by equation (2.11). We also note that a double smoothing of the straight gradient, i.e. equation (2.16), gives better results for smaller ensemble sizes and inferior results for larger ensemble sizes. Pre-multiplication of the gradient given by the cross-covariance vector, i.e. equation (2.12), achieves better results than those obtained with equation (2.11).

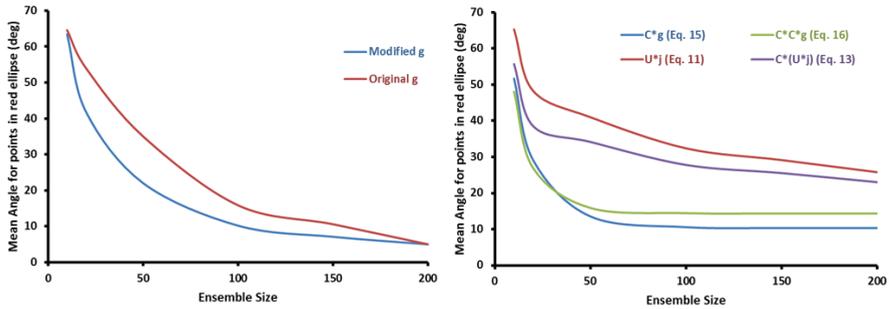


Fig. 3.12: (a) illustrates the performance of the two different formulations used to estimate the 'straight gradient' g given by equation (2.14), while (b) displays the effect of using different versions of the 'smoothed' gradient and their relative behavior.

3.4.1.2 The Flatter Region

The general trend in the results observed for the steeper region is also observed for the flatter region. However, estimating the straight gradient which gives good estimates for the steeper region does not achieve good gradients in the flatter part. Smoothing of the gradients, on the other hand, achieves much-higher-quality gradients in this flatter region, in both regions the modified formulation of Do and Reynolds (2013) performs best. A major difference in gradient estimation for the flatter region is the sensitivity of the gradient estimate quality to the choice of the correlation length used. Although we do not achieve the 95% confidence interval for this region we observe that we need smaller perturbation sizes, $\sigma = 0.001$, and correlation lengths varying from 8 to 12 control time steps in conjunction with an ensemble size equal to 300 to satisfy the hypothesis at most of the points. This shows that correlation length and the covariance matrix have more impact in the flatter region. It is virtually impossible to know a-priori the correct correlation length, and the lengths used here were obtained through an

exhaustive trial and error approach. Similar to the results for the steeper region, we observe that a smoothed gradient given by equation (2.15) achieves a better angle compared to directly estimating a smoothed gradient using equation (2.11). **Fig. 3.13** shows the quality of the different smoothed gradient formulations. When using equation (2.11), then a further single smoothing, i.e. equation (2.13), which is akin to a double smoothing of equation (2.14), i.e. equation (2.16), achieves better quality gradients for all ensemble sizes. Irrespective of the different gradient estimates used, the gradient quality improves with an increase in ensemble size. We also observe that for this region with ensemble sizes higher than 300 we do not observe a significant improvement in gradient quality. The angles obtained are never greater than 90 degrees, i.e. within the first quadrant with respect to the adjoint gradient, irrespective of the ensemble size used. This is particularly important, because it implies that even for smaller ensemble sizes we are able to approximate, in general, a roughly ‘correct’ up-hill direction.

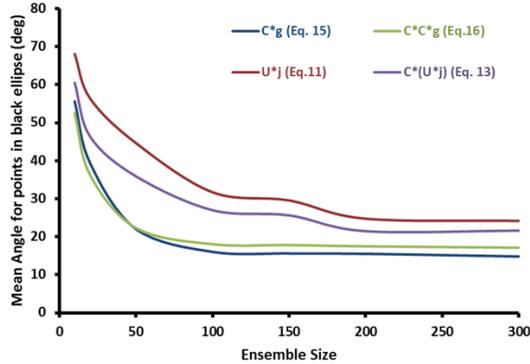


Fig. 3.13: Comparison of the effect of smoothing and different gradient estimates for the flatter region (black ellipse) of Fig. 3.11.

3.4.1.3 Effect of Gradient Quality on Optimization

In order to quantify the impact of gradient quality on the optimization process and to justify the choice of 10 degrees as a definition of the null hypothesis we have performed a set of optimization experiments with varying degrees of gradient quality. The methodology used to quantify the impact of gradient quality using approximate gradient with varying angle errors is as follows

1. Compute the adjoint gradient using the in-house reservoir simulator at each optimization iteration.
2. Stochastically perturb the adjoint gradient with a chosen perturbation size to mimic a desired angle away from the adjoint gradient.

3. Check if the perturbed gradient lies between a 1 degree +/- range from the desired angle. If not either increase or decrease perturbation size used to find the desired angle.
4. Use the perturbed gradient in the optimization loop.
5. Repeat for different angle errors to mimic either high or low quality gradient estimates i.e. varying the angles during the optimization.

Fig. 3.14 is an illustration of the impact of using different quality angles during the optimization process. For the comparison the number of iterations is fixed to 40. We observe that for an angle of 10 degrees throughout the optimization process we obtain very similar values of the objective function for the final optimized solution compared to the optimum achieved with the adjoint method though initially the optimization path is marginally inferior. For angles greater than 30 degrees we observe that though the optimization has not yet converged, for the same number of iterations we achieve an inferior objective function solution. For the results in this sub-section adjoint based gradients computed from the in-house reservoir simulator, for details see Kraaijevanger et al. (2007), has been used.

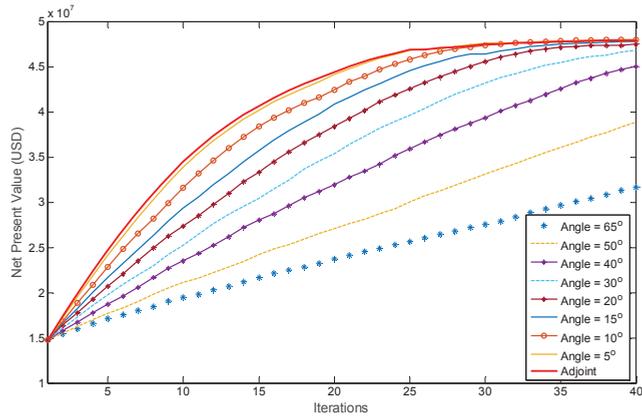


Fig. 3.14: Illustration of the impact of the use of inferior quality gradients during the optimization process.

3.4.2 Egg Model with Uncertainty

The main purpose of this study is to quantify the quality of the ‘robust’ ensemble gradient and determine the optimal ratio required to estimate a high-quality gradient for a realistic reservoir test case. Thus, to test our methodology we use an ensemble of 100 equi-probable geological models, six of which, obtained from Van Essen et al. (2009) are displayed in **Fig. 3.15**. The uncertainty is captured through the different permeability fields and the different directions and orientations of the channels, see

Jansen et al. (2013) for details. The hypothesis tests are performed to estimate a ‘robust’ ensemble gradient using all hundred realizations of this model. The controls and the fluid model as well as the other properties are exactly the same as for the deterministic case. To generate the points to test our methodology, the same approach is used as described above. That is, we first perform robust optimization using the adjoint method as described in Van Essen et al. (2009), and then assess the quality of the gradient along this control trajectory. The result of the robust optimization is displayed in **Fig. 3.16**. The optimization was limited to 25 iterations due to the computational complexity involved. Again, for the analysis, we divide the optimization process into two regions using the same reasoning as explained above.

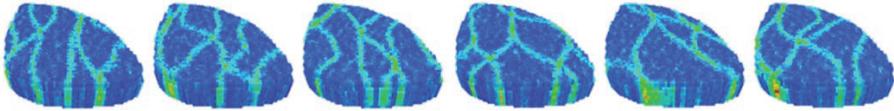


Fig. 3.15: Six randomly chosen realizations displaying the uncertainty in the geological models obtained from Van Essen et al (2009); see Jansen et al. (2013) for details about the models

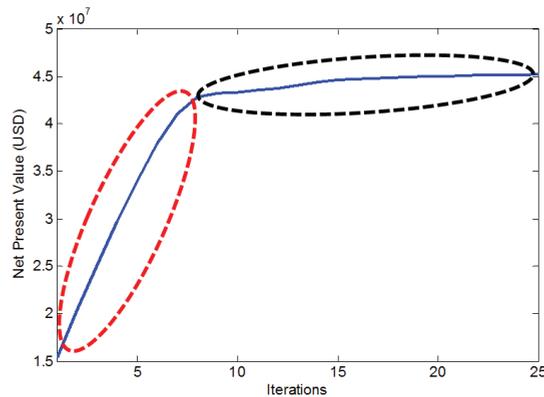


Fig. 3.16: Illustration of the expected objective function value over 100 realizations with the robust optimization process (blue curve) using adjoint gradients. The curve is divided into two parts, a red ellipse (indicative of the steeper region of the curve) and a black ellipse (indicative of the flatter region of the curve).

3.4.2.1 The Steep Region

Following the analysis provided for the deterministic case we first consider the steeper part of the optimization curve encapsulated by the red ellipse as shown in Fig. 3.16. We have eight points (control strategies) which lie in this region for which the original and modified formulation using the computationally attractive 1:1 ratio have been compared. We observe that for all the eight points we are able to satisfy our null hypothesis, i.e. the angle is less than 10 degrees, when using the modified robust formulation [based on equations (2.22) and (2.23)]. However if we use the 1:1 ratio

with the original formulation [based equations (2.19) and (2.20)] we never satisfy our hypothesis at any of the points. With the original formulation we observe angles between 80 and 90 degrees for all the points when using equation (2.14). (As for the deterministic case, all results are the mean of 5 experiments with different random seeds.) If we use the ‘smoothed’ versions of the gradient as given by equations (2.11), (2.13), (2.15) and (2.16), we observe much better angles as shown in **Fig. 3.17**, though the hypothesis is still not satisfied. Thus smoothing of the gradient leads to better estimates of the ensemble gradient, and we observe a similar trend as for the deterministic case in that the smoothed version of equation (2.14), i.e. equation (2.15), achieves better results than equation (2.11). We also observe that using a relatively larger perturbation size ($\sigma = 1$) for this region leads to the best results. These results highlight the advantage of using the modified formulation with the 1:1 ratio because the maximum contribution (largest increase in objective function) is from within this region so achieving good angles at acceptable computational costs is more important for this region. Though the 1:1 ratio satisfies our hypothesis, using higher ratios leads to higher-quality gradients for both gradient formulations with the HPG formulation as depicted in Fig. 3.17.

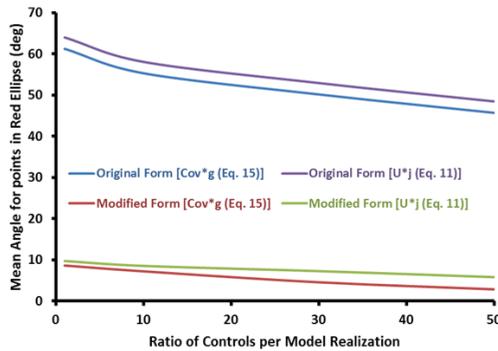


Fig. 3.17: Illustrates the difference in gradient quality between the original and modified formulations for varying ratios of controls per model realization using the HPG ratio formulation.

For the Rosenbrock function we observed significant differences in gradient quality when using the HPG and MIG ratio formulations. For this case we observe a similar trend in the results. **Fig. 3.18** consists of two plots which show the difference in gradient quality for the two ratio formulations. Fig. 3.18(a) depicts the results obtained when using the original formulation, where, similar to the Rosenbrock results, we always achieve significantly better gradient estimates with the MIG formulation. We also observe, as expected, that as the ratio increases the gradient quality improves. Fig. 3.18(b) depicts the results when using the modified formulation. The results indicate that the modified formulation is less sensitive to the different ratio formulations,

although the MIG formulation still performs better. Irrespective of the ratio formulation the modified form, satisfies the hypothesis at all the points, even for the 1:1 ratio.

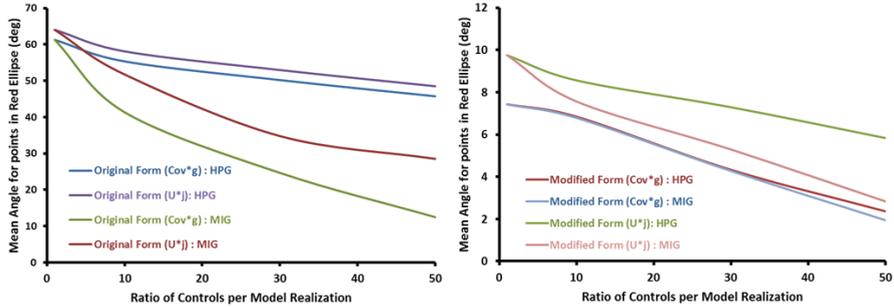


Fig. 3.18: (a) Difference in gradient quality for the two ratio formulations (HPG and MIG) for the original gradient formulation while (b) highlights the gradient quality when using the modified gradient formulation. Note the different vertical scales

3.4.2.2 The Flatter Region

The 1:1 ratio with the modified formulation [based on equations (2.22) and (2.23)] leads to very high quality gradient estimates for the steeper part of the optimization curve. However, for the flatter part, though still better than the original formulation [based on equations (2.19) and (2.20)], the gradient quality is not as high. An increase in the ratios significantly improves the gradient quality as shown in **Fig. 3.19**. We also observe, slightly different to our previous observations that the different versions of the ‘smoothed’ gradient, given by equations (2.11), (2.13), (2.15) and (2.16), give very similar results for both the modified and original formulations. Similar to the deterministic case we use a smaller perturbation size ($\sigma = 0.01$) and a correlation length that varies between 8 and 17. It requires a substantial number of simulations to perform an extensive search for the ‘correct’ correlation length which has a strong impact on the gradient quality.

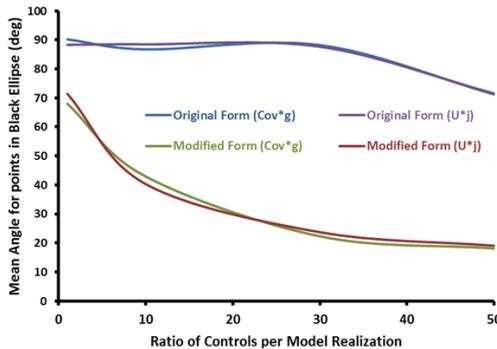


Fig. 3.19: Illustrates the difference in gradient quality between the original and modified formulations for varying ratios of controls per model realization using the HPG formulation.

Similar to the results shown above for the steeper region, the different ratio formulations show a similar trend for the points in flatter region. **Fig. 3.20** consists of two plots which show the difference in gradient quality for the different ratio formulations. Fig. 3.20(a) depicts the results obtained when using the original formulation and Fig. 3.20(b) depicts the results when using the modified formulation. We observe that the same trend in the results exists, although on the flatter part of the curve we have satisfied our hypothesis for only a few points for this robust case. We could argue that for the robust case and, considering our problem dimension, the difference between 10 and 20 degrees is marginal. Thus if we were to have a hypothesis which had an allowable error margin of 20 degrees then, with the MIG ratio formulation and both the original or modified gradient formulation, we would satisfy our hypothesis at most of the points for a ratio of 1:50, i.e. we would need to apply 50 control samples per geological realization to estimate a ‘good’ gradient. This would imply that, to estimate a gradient at one point within the optimization process we would need to perform 5000 reservoir simulations since the ensemble size is 100. This is computationally challenging and therefore, since the increase in objective function value for the flatter region is less than 1% we suggest the use of the modified formulation with a 1:1 ratio.

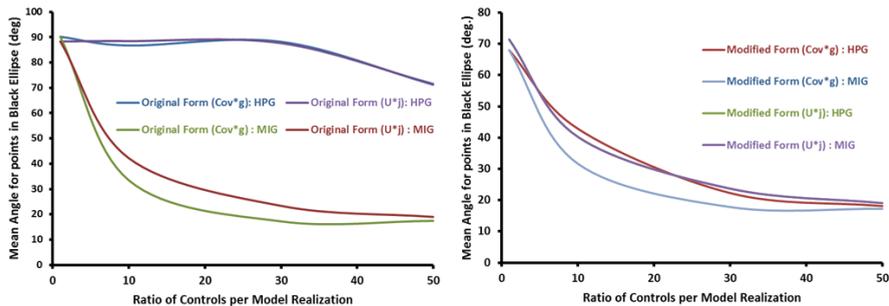


Fig. 3.20: (a) shows the difference in gradient quality for the two ratio formulations (HPG and MIG) for the original gradient formulation while (b) highlights the results when using the modified gradient formulation.

3.4.2.3 Angle Effect on Optimization

Fig. 3.21 is a comparison of the performance of a robust optimization process between the ‘adjoint-based’ formulation (red line), the modified EnOpt formulation (blue dashed line) and the original EnOpt formulation (black line), both using the single smoothed 1:1 ratio which is computationally efficient. This provides a quantitative analysis for the results presented above along with a justification for the use of better quality gradients during an optimization experiment. We observe that in the ‘steeper’ region we are able to effectively “track” the path of the adjoint gradient with the modified formulation which is accordance with the gradient quality achieved

while for the original formulation although we increase the objective function we get inferior results compared to the adjoint method. In the flatter region we observe, in accordance with the gradient quality results, we get a marginally inferior solution in terms of NPV with the modified formulation compared to the adjoint method thus highlighting the impact of inferior gradient quality. As the gradient quality in the steeper region is high this benefits the final optimized achieved NPV with the modified formulation compared to the original formulation.

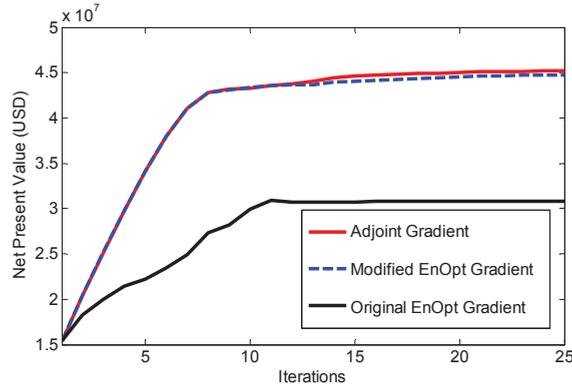


Fig. 3.21: Comparison of the effect of gradient quality on the optimization process for three different robust gradient formulations. In the steeper region the modified EnOpt formulation computes very good quality gradients while in the flatter region the gradient quality decreases.

3.5 Conclusions

- ✓ For the relatively simple Rosenbrock function we need an ensemble size equal to 900, 5 or 3 to satisfy the null hypothesis and achieve the desired confidence interval depending on the perturbation size used ($\sigma = 0.1$, $\sigma = 0.01$ and $\sigma = 0.001$ respectively) to generate the ensemble of controls.
- ✓ Including uncertainty within the Rosenbrock function we find that the modified formulation [based on equations (2.22) and (2.23)] with the computationally attractive 1:1 ratio outperforms the original formulation [based equations (2.19) and (2.20)], irrespective of the degree of uncertainty. However the degree of uncertainty does play a role in the quality of the gradient estimate. Higher-quality gradients can be obtained with an increasing ensemble size of model realizations i.e. using higher ratios.
- ✓ When working with higher ratios significantly better results are obtained through the use of the ‘Mean of Individual Gradients’ (MIG) formulation compared to the HPG formulation. This result is not surprising as this would be expected from the central limit theorem. Additionally the ratios required

are smaller; for the Rosenbrock case, depending on the perturbation size, a ratio of 1:3 is sufficient to satisfy the hypothesis and meet the desired confidence interval using the MIG formulation.

- ✓ The analysis for the “Egg Model” case was divided into two regions: 1) a ‘steeper’ region, where with higher perturbation size ($\sigma = 0.01$) and lower ensemble sizes (150 samples) the hypothesis was satisfied at all points, and 2) a ‘flatter’ region where a smaller perturbation size ($\sigma = 0.001$) and a higher ensemble size (300 samples) were needed.
- ✓ In the flatter region, in addition to the perturbation size, a correlated covariance is needed where the choice of correlation length has a significant impact on the gradient quality. Through a trial and error procedure we observe that correlation lengths between 8 and 13 give the best results and lead to satisfying the hypothesis at most points.
- ✓ For the robust “Egg Model” case, we observe a similar trend in the results as for the uncertain Rosenbrock function. The modified formulation using the 1:1 ratio achieves significantly better results compared to the original formulation at any point along the optimization curve
- ✓ In the flatter part of the optimization curve the modified formulation with the 1:1 ratio, although performing better than the original formulation, never satisfies the hypothesis. Using higher ratios in the flatter region is necessary to achieve a good quality gradient. However, in the steeper part, the hypothesis is satisfied for all the points.
- ✓ As the results have shown, the developed methodology can, in theory, be used to quantify the ensemble size required to achieve a high-quality gradient. However, all the angles obtained in this work are always less than 90 degrees which suggests that with ensemble methods, irrespective of the ensemble size (in our case), we estimate the ‘correct’ uphill direction.
- ✓ Use of a gradient of inferior quality in an optimization experiment strongly influences the final achievable NPV for a finite number of optimization iterations.
- ✓ We recommend to use, out of the 30 potential robust gradient formulations identified in our paper, the single ‘smoothed’ modified formulation, i.e. equation (2.15) based on equations (2.22) and (2.23), using the 1:1 ratio for recovery optimization under uncertainty. In an optimization context, it is vital to use good quality

4

ENOPT WITH COVARIANCE MATRIX ADAPTATION

Current implementations of EnOpt, a rapidly emerging method for model-based optimization, use a Gaussian ensemble of control perturbations with a constant covariance matrix, and thus a constant perturbation size, during the entire optimization process. The Covariance Matrix Adaptation Evolutionary Strategy (CMA-ES) is a gradient-free optimization method developed in the ‘machine learning’ community, which also uses an ensemble of controls, but with a covariance matrix that is continually updated during the optimization process. It has shown to be an efficient method for several difficult but small-dimensional optimization problems and has recently been applied in the petroleum industry for well location and production optimization. In this study we investigate the scope to improve the computational efficiency of EnOpt through the use of covariance matrix adaptation (CMA-EnOpt).³ The resulting method is applied to the water flooding optimization of a small multi-layer test model and a modified version of the Brugge benchmark model. The controls used are inflow control valve settings at pre-defined time intervals for injectors and producers with undiscounted net present value as the objective function. We compare EnOpt and CMA-EnOpt starting from identical covariance matrices. For the small model we achieve only slightly higher (0.7%-1.8%) objective function values and modest speed-ups with CMA-EnOpt compared to EnOpt. Significantly higher objective function values (10%) are obtained for the modified Brugge model. The

³ This chapter is taken from: Fonseca, R.M, Leeuwenburgh O., Van den Hof, P.M.J. and Jansen, J.D., (2015): Improving the Ensemble Optimization Method through Covariance Matrix Adaptation (CMA-EnOpt), *SPE Journal* 20(1).

possibility to adapt the covariance matrix, and thus the perturbation size, during the optimization allows for the use of relatively large perturbations initially, for fast exploration of the control space, and small perturbations later-on, for more precise gradients near the optimum. Moreover, the results demonstrate that a major benefit of CMA-EnOpt is its robustness with respect to the initial choice of the covariance matrix. A poor choice of the initial matrix can be detrimental to EnOpt, whereas the CMA-EnOpt performance is near-independent of the initial choice and produces higher objective function values at no additional computational cost.

4.1 Introduction

Derivative-free methods such as genetic algorithms or evolutionary strategies developed and widely used in the ‘machine learning’ community can be viewed as an alternative to approximate gradient techniques such as EnOpt. Such derivative-free methods are being used with increasing frequency for model-based optimization of oil recovery. One of the more popular derivative free methods used for petroleum engineering applications is the Covariance Matrix Adapted-Evolutionary Strategy (CMA-ES), which was developed by Hansen and co-workers (Hansen and Ostermeier 1996, 2001, Hansen 2006). CMA-ES has recently been used for well placement optimization by Ding (2008) and Bouzarkouna et al. (2011), in a flooding optimization problem by Schulze-Riegert et al. (2011), and for a smart well optimization problem by Pajonk et al. (2011) amongst others.

CMA-ES aims to ‘learn’ about the objective function search space to adaptively update the covariance matrix used to generate the ensemble of controls. Current implementations of EnOpt use a user defined constant covariance matrix to generate the ensemble of controls. It is impossible for a user to have a-priori knowledge about the optimal covariance to be used. Additionally as illustrated in chapter 3 different regions in the objective function space require covariance matrices with different structures and properties to achieve high quality gradient estimates. Thus in this chapter we propose a hybrid scheme, Covariance Matrix Adapted EnOpt (CMA-EnOpt), which combines the explicit use of the gradient information (EnOpt) with the continuous adaptation of the covariance matrix (CMA). The CMA algorithm leads to an improved sampling mechanism which uses “local” knowledge of the nature of the objective function search space. In this chapter we will first provide an introduction to CMA-EnOpt, followed by its application to a small synthetic 3D reservoir model and a modified version of the Brugge benchmark model. A comparison of the results to those obtained with EnOpt will illustrate the advantages of CMA-EnOpt for relatively large-scale model-based production optimization.

4.2 Update rules

The approximate gradient \mathbf{g} from equation (2.14) can be used in any gradient-based optimization algorithm. In this thesis we have used a simple steepest ascent scheme according to

$$\mathbf{u}^{\ell+1} = \mathbf{u}^{\ell} + \alpha^{\ell} \mathbf{g}^{\ell}, \quad (4.1)$$

where the superscript ℓ is the iteration counter, and α^{ℓ} is a step length in the direction of the gradient. Note that, to ensure dimensional consistency, α^{ℓ} must have dimensions, and its value will therefore be dependent on the units system applied. If \mathbf{u} and \mathbf{g} both contain elements with different dimensions (e.g. when the elements of \mathbf{u} are pressures and rates) an additional scaling of the gradient elements may be required. Following Oliviera and Reynolds (2014), we scaled the gradient by its infinity norm and then, for each iteration, used an initial step size $\alpha^{\ell} = 0.1$. Thereafter, we allowed for a maximum of three back-tracking steps, each time reducing the step size with a factor of one half. In more sophisticated optimization algorithms an improved update direction (i.e. one different from \mathbf{g}) is determined by employing optimization methods that make use of the second derivatives of J with respect to \mathbf{u} , i.e. of the Hessian matrix, or, more commonly, of approximations to the Hessian. In particular, so called quasi-Newton methods use gradient information of subsequent iterates to construct an approximate Hessian \mathbf{H}^{ℓ} . The corresponding update rule then becomes

$$\mathbf{u}^{\ell+1} = \mathbf{u}^{\ell} + \alpha^{\ell} (\mathbf{H}^{\ell})^{-1} \mathbf{g}^{\ell}, \quad (4.2)$$

where the definition of \mathbf{H}^{ℓ} depends on the particular type of quasi-Newton method applied; see e.g. Nocedal and Wright (2006) or Luenberger and Ye (2010) for further details. Unlike equation (4.1), equation (4.2) is dimensionally consistent if α is taken dimensionless, because \mathbf{H}^{-1} acts as a natural scaling matrix for the gradient vector \mathbf{g} . Note that, as usual, in an actual implementation computing the inverse is avoided, and a system of equations is solved instead. The gradient is the direction of a tangent (hyper) plane in a point touching the objective function, while the Hessian gives curvature information in that point, i.e. it defines a convex quadratic function. The basic idea underlying the various quasi-Newton methods is that the curvature information contained in the approximate Hessian is gradually increased by subsequent inclusion of gradient information from previous iterations. Although we do not use a quasi-Newton algorithm in the optimization examples in our study, the concept of using information from subsequent iterates to improve the estimate of the curvature of the objective function is an important aspect of CMA-ES, and thus also of CMA-

EnOpt. Moreover, we note that the use of the preconditioners \mathbf{C}_{uu} and $\tilde{\mathbf{C}}$ in equations (2.15) and (2.16) seem to play a similar role as the preconditioner \mathbf{H}^{-1} in equation (4.2). However, \mathbf{C}_{uu} and $\tilde{\mathbf{C}}$, unlike \mathbf{H}^{-1} , do not restore dimensional consistency.

4.3 Covariance Matrix Adaptation

CMA-ES is a stochastic iterative optimization method in which the covariance matrix is updated at every iteration such that its largest principal direction, i.e. the eigenvector corresponding to its largest eigenvalue, is (approximately) re-aligned in the direction of the maximal increase of the objective function. CMA-ES uses two types of updates for the covariance matrix as briefly explained below. For a detailed overview we refer to Hansen (2011).

4.3.1 Rank- μ Update

The motivation behind a rank- μ update is to use information obtained within one single iteration (i.e. one ensemble of random control vectors $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_M\}^\ell$ and their corresponding objective function values $\{J_1, J_2, \dots, J_M\}^\ell$ through selecting the ‘best’ μ members (i.e. those corresponding to the μ highest objective function values) out of the M ensemble members:

$$\tilde{\mathbf{C}}_{uu}^{\ell+1} = (1 - c_\mu) \tilde{\mathbf{C}}_{uu}^\ell + c_\mu \frac{1}{\mu} \tilde{\mathbf{U}} \tilde{\mathbf{U}}^T, \quad \tilde{\mathbf{U}} = [\mathbf{u}_1 - \mathbf{u}^\ell \quad \mathbf{u}_2 - \mathbf{u}^\ell \quad \dots \quad \mathbf{u}_\mu - \mathbf{u}^\ell], \quad (4.3)$$

where $0 < c_\mu < 1$ is a learning rate, and where the control vectors $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_M$ have been ranked such that for their corresponding objective function values it holds that $J_1 \geq J_2 \geq \dots \geq J_\mu > J_{\mu+1} \geq \dots \geq J_M$. Equation (4.3) is called a rank- μ update because the matrix product $\tilde{\mathbf{U}} \tilde{\mathbf{U}}^T$ is, at most, of rank μ . Note that we use the distribution mean \mathbf{u}^ℓ instead of the ensemble mean $\bar{\mathbf{u}}^\ell$. It can be shown that for $\mu = M$ and $c_\mu = 1$, $\tilde{\mathbf{C}}_{uu}^{\ell+1}$ would be an unbiased estimator of the distribution covariance $\tilde{\mathbf{C}}^{\ell+1}$ (Hansen 2006). However, because we typically choose $\mu < M$, the entries of the covariance matrix will be selectively influenced by the ensemble members corresponding to the μ highest objective function values. The choice of the learning rate c_μ turns out to be crucial to the success of the optimization as will be demonstrated later. Hansen (2011) discusses strategies to determine optimal values for c_μ (and the learning rate c_1 introduced below) based on the dimension of the problem. However their test cases are of a relatively small dimension, whereas production optimization problems typically have hundreds to thousands of control variables. The choice of μ is up to the user; in this study we used $\mu = M/4$.

4.3.2 Rank-One Update

In equation (4.3) the covariance matrix update is determined using the best objective function values within one single iteration. It is also possible to update the covariance matrix by using information from previous iterates, in a similar fashion as updating the Hessian in quasi-Newton methods. (For remarks about the relationship between the Hessian and the covariance matrix, see Hansen (2011).) The expression for such an update, derived in Hansen (2006), is given by

$$\tilde{\mathbf{C}}_{uu}^{\ell+1} = (1-c_1)\tilde{\mathbf{C}}_{uu}^{\ell} + c_1\mathbf{p}^{\ell+1}(\mathbf{p}^{\ell+1})^T, \quad (4.4)$$

where c_1 is again a learning rate and \mathbf{p} is the ‘evolution path’, which is a function of iterates \mathbf{u}^{ℓ} in earlier steps. Roughly speaking, \mathbf{p} is obtained as a summation of previous iterates emphasizing the most recent iterates while gradually ‘forgetting’ the earlier ones. For the exact definition of \mathbf{p} , see Hansen (2006). Because the outer product of two vectors results in a matrix of rank one, equation (4.4) is referred to as a ‘rank-one update’. The rank-one update has been shown to be particularly powerful when using small ensemble sizes with CMA-ES (Hansen, 2006).

4.3.3 Combined Rank Update

Combining equations (4.3) and (4.4) one obtains the update rule

$$\tilde{\mathbf{C}}_{uu}^{\ell+1} = (1-c_{\mu}-c_1)\tilde{\mathbf{C}}_{uu}^{\ell} + c_{\mu}\underbrace{\frac{1}{\mu}\tilde{\mathbf{U}}\tilde{\mathbf{U}}^T}_{\text{rank-}\mu \text{ update}} + c_1\underbrace{\mathbf{p}^{\ell+1}(\mathbf{p}^{\ell+1})^T}_{\text{rank-one update}}. \quad (4.5)$$

Equation (4.5) utilizes information within one iteration as well as information from previous iterations. Hansen (2011) suggests that the former is more important when using a larger ensemble and that the latter is more important when using smaller ensembles. Several variations to equation (4.5) have been proposed; see e.g. Ros and Hansen (2008) and Arnold and Hansen (2010) wherein the off-diagonal elements are set to zero and only the diagonal elements i.e. the variances of the covariance matrix are updated:

$$\tilde{C}_{uu,ii}^{\ell+1} = (1-c_{\mu}-c_1)\tilde{C}_{uu,ii}^{\ell} + c_{\mu}\underbrace{\frac{1}{\mu}\tilde{U}_{ii}\tilde{U}_{ii}^T}_{\text{rank-}\mu \text{ update}} + c_1\underbrace{p_i^{\ell+1}(p_i^{\ell+1})^T}_{\text{rank-one update}}, i=1,\dots,M. \quad (4.6)$$

4.4 Covariance Matrix Adapted-EnOpt (CMA-EnOpt) Algorithm

In EnOpt a constant distribution covariance matrix $\tilde{\mathbf{C}}$ is used which, for uncorrelated controls of the same type, is typically chosen as a diagonal covariance

matrix with equal diagonal elements σ^2 . Often, a time correlation is imposed on the controls resulting in a block-diagonal matrix with each block corresponding to the control time steps of a single well or inflow control valve (ICV). The value of the standard deviation σ can have a strong influence on the quality of the approximate gradient and therefore on the performance of the optimization algorithm. However, there is no well-defined method to choose this value. (The same is true for the choice of the time correlation length.) Thus, we propose to use the covariance adaption strategy described above to gradually improve the distribution covariance matrix in EnOpt, leading to the following algorithm for CMA-EnOpt.

1. Set $\ell = 0$. Choose an initial control vector \mathbf{u}^0 and evaluate the corresponding objective function value.
2. Generate an ensemble of randomly perturbed controls around \mathbf{u}^0 from an initial, user defined, diagonal covariance matrix $\tilde{\mathbf{C}} = \sigma^2 \mathbf{I}$ or block-diagonal covariance matrix $\tilde{\mathbf{C}}$.
3. Run a reservoir simulation for every member of the perturbed control ensemble and calculate the corresponding objective function values using equation (2.1).
4. Compute the EnOpt gradient \mathbf{g}^ℓ using equation (2.14).
5. If the optimization stopping criterion, or the maximum allowed number of iterations is achieved, stop. Else, set $\ell = \ell + 1$.
6. Determine a step size α^ℓ and compute an updated control vector $\mathbf{u}^{\ell+1}$ using equation (4.1) and the corresponding objective function value.
7. Compute the updated covariance matrix $\tilde{\mathbf{C}}_{uu}^{\ell+1}$ using either equation (4.5) or equation (4.6)
8. Regenerate an ensemble of randomly perturbed controls around $\mathbf{u}^{\ell+1}$ from the updated covariance matrix $\tilde{\mathbf{C}}_{uu}^{\ell+1}$.
9. Go to step 3.

4.5 Results: 5-Spot Synthetic Reservoir Model

Advances in technology have led to an increase in the application of ICVs to regulate flow rates and maintain pressure in the reservoir. We consider a control problem where ICV settings of injection and production wells in a 3D synthetic reservoir model, taken from JOA 2007, are manipulated to optimize waterflooding over the producing life of the reservoir. The model, illustrated in **Fig. 4.1(a)**, consists of $25 \times 32 \times 5 = 4000$ grid blocks. The approximate size of each grid block is $110 \times 90 \times 20$ m, so that the reservoir volume is $2.5 \times 3.5 \times 0.1$ km³. The geological structure consists of uplifted blocks, separated by faults. The reservoir is produced

using an inverted five-spot well pattern, i.e. four producers at the corners of the grid with an injector in the center. The reservoir is divided into five layers with different horizontal permeabilities. **Fig. 4.1(b)** is the top view of the transmissibility multipliers used for this model, the white cells are grid blocks which are inactive. We observe that there is a sealing fault on the North-Western side of the block, close to producer 1. **Table 4.1** lists the reservoir and fluid properties of the model. A Corey model with exponents equal to 2 for both oil and water is used for the relative permeabilities where the connate water saturation is 0.2, the residual oil saturation is 0.3 and the end point relative permeabilities to oil and water are 0.8 and 0.4 respectively. Capillary pressure effects are not included.

<u>Property</u>	<u>Values</u>	<u>Units</u>
Porosity	0.2	--
Permeability (layer 1 – layer 5)	100-300-50-600-100	mD
Reservoir pressure @ 1950 m	200	bar
Density of oil	800	kg/m ³
Density of water	1000	kg/m ³
Temperature	77	°C
Oil compressibility @ 200 bar	4e-5	1/bar
Water compressibility @ 200 bar	4e-5	1/bar
Rock compressibility	0	1/bar
Viscosity of oil @ 1 bar	2	cP
Viscosity of water @ 1 bar	0.5	cP

The wells penetrate all five layers with one ICV in each layer. The producing life of the reservoir is divided into 15 optimization control steps with each of one year (365 days) in duration and there are 25 controls per control step which results in a total of $15 \times 25 = 375$ controls to be optimized. Water is injected at a constant pressure of 300 bars and the production wells are operated at a minimum pressure of 15 bar. We used an oil price $r_o = 130$ \$/m³, water production costs $r_{wp} = 25$ \$/ m³, and water injection costs $r_{wi} = 6$ \$/m³. Well index multipliers were used to model the ICVs in the simulator with bounds of 1×10^{-4} and 1. For the simulation of the model we used a commercial fully implicit finite difference black oil simulator (Eclipse, 2011).

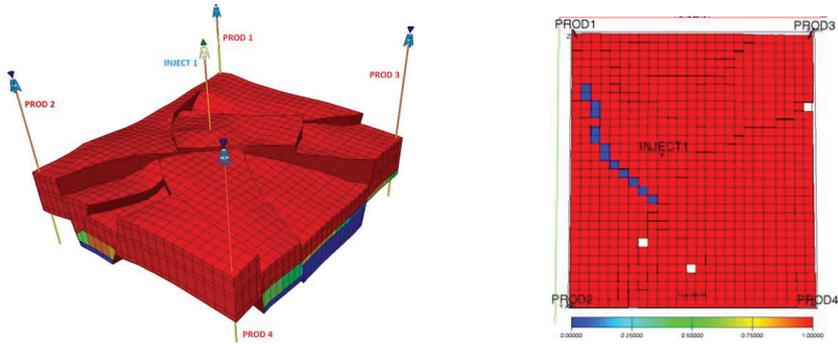


Fig. 4.1: (a) Five-spot reservoir model. The colors indicate the initial oil saturation. (b) Transmissibility multiplier values for the model. One fault has a zero transmissibility (blue) thus is sealing.

4.5.1 Comparison between EnOpt and CMA-EnOpt

We performed several comparisons between EnOpt and CMA-EnOpt to optimize the ICV settings with the aim to maximize NPV as defined in equation (2.1). The starting point for the optimization was an initial control vector having values equal to 1. Thus all the ICVs were fully open as a starting strategy. The initial value of σ , for use in a diagonal covariance matrix, was chosen equal to 0.1 and we used a fixed ensemble size of 50 samples. Random control values outside the range $[1 \times 10^{-4}, 1]$ were simply reset to their bounds. We first compare the results using only a diagonal update for the covariance matrix; see equation (4.6). The optimization was allowed to run for 50 iterations which usually resulted in a near-horizontal (i.e. nearly converged) objective function graph for the EnOpt method; see **Fig. 4.2**. When using approximate gradients and inexact line search techniques, it is very well possible that the curves do not always increase monotonically. We have chosen to allow for this to happen. Although monotonicity could be enforced, this would induce a higher computational cost and an increased risk of becoming trapped in a local optimum. We used various settings of the initial distribution covariance matrix (i.e. of the standard deviation σ) to generate the initial ensemble, leading to different optimization results, see **Fig. 4.2**. The best EnOpt run resulted in an objective function value of 9.1×10^9 \$ while the CMA-EnOpt run for the same initial covariance distribution achieved a slightly higher value of 9.15×10^9 \$, i.e. 0.7% higher. An illustration of the corresponding ICV settings for one of the wells is presented in **Fig. 4.3**, which shows the differences in the control strategies between EnOpt (blue line) and CMA-EnOpt (red line). The major difference is in the nature of the controls: CMA-EnOpt obtains controls that switch almost completely between the upper and lower bounds. We note that such (near-) bang-bang controls are, in some optimization problems, the optimal solutions, and are often easier to implement in practice. As shown by Zandvliet et al. (2007) a locally optimal solution of an optimal

control problem in which the system equations and the integral term in the objective function are both linear in the controls, and that has only simple bound constraints or linear constraints on the controls, is necessarily a bang-bang solution (possibly in combination with singular arcs, i.e. areas where the solution is not bang-bang). The full analysis of necessary and sufficient conditions for bang-bang control would require rewriting the optimization problem in terms of an optimal switching time with the aid of an adjoint solution, which we therefore did not pursue for this example. We refer to Zandvliet et al. (2007) for detailed information on the conditions for bang-bang control.

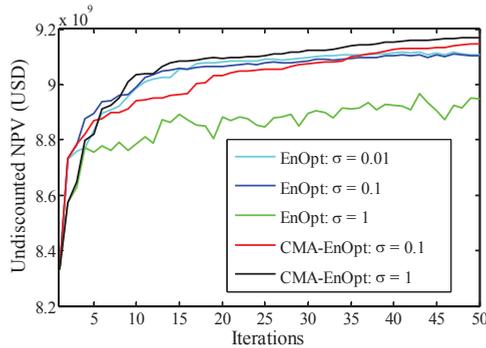


Fig. 4.2: Comparison of optimization performance for different initial covariance matrices.

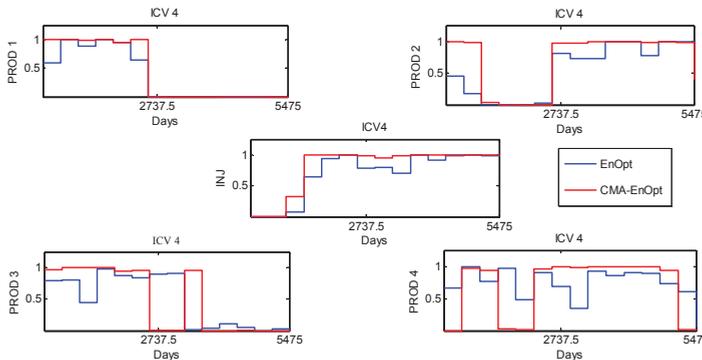


Fig. 4.3: Optimal control settings for ICV 4 as computed by EnOpt (blue) and CMA-EnOpt (red) with initial $\sigma = 0.1$ for both cases

4.5.2 Improved Robustness

Exact gradient-based methods (in combination with an exact line search) are inherently local methods because they always result in uphill directions, unless forced to take steps in other directions to scout for other optima. Most gradient-free methods have been shown to possess more global search characteristics. In the EnOpt method the

initial distribution covariance matrix may be chosen to mimic the desired behaviour: large variances results in a more global search strategy and small variances a more local one. We tested EnOpt and CMA-EnOpt with different initial starting values for the standard deviation σ . Fig. 4.2 illustrates that the choice of the initial covariance matrix has a significant impact on the performance of EnOpt, because the matrix remains constant throughout the optimization. When using $\sigma=1$ for EnOpt (green line), the algorithm achieves poor results, probably because the sampling takes place in a too large area which leads to poor approximations of the gradient. For $\sigma=0.1$ (blue line) and $\sigma=0.01$ (light blue line) better results are achieved. For CMA-EnOpt (red and black lines) the initial starting covariance matrices have relatively little impact on the final objective function values and both runs achieve better results than EnOpt. When starting with $\sigma=1$, CMA-EnOpt achieves a slightly higher objective function value than when starting with $\sigma=0.1$. However if the algorithm is continued for more iterations we observe (not shown here) that both the curves (red and black) achieve very similar objective function values. For this example, we therefore find that CMA-EnOpt performs better than EnOpt with an optimized distribution covariance matrix, and much better than EnOpt with a poor initial guess for the covariance. The main benefit of CMA-EnOpt therefore appears to be its robustness with respect to initial choices of the covariance matrix.

4.5.3 Learning Rates

CMA-EnOpt contains several parameters that require user-defined values, in particular the learning rates c_μ and c_1 . **Fig. 4.4** illustrates that if we choose the learning rates too low, the advantage of CMA-EnOpt over EnOpt is negligible, if any. Higher learning rates are seen to achieve significantly better results. Here, a high learning rate (fast learning) corresponds to a 75-25% update rule with $c_\mu=0.20$ and $c_1=0.05$. In this case $(c_\mu + c_1) \times 100\% = 25\%$ new information is incorporated into the covariance matrix every iteration. A low learning rate (slow learning) corresponds to a 99.5-0.5% update rule with $c_\mu=0.004$ and $c_1=0.001$, in which case $(c_\mu + c_1) \times 100\% = 0.05\%$ new information is incorporated. These results were obtained when we only updated the diagonal elements of the covariance matrix. Hansen (2011) reports, for CMA-ES, that if the full covariance matrix is updated, high learning rates can have a detrimental impact on the optimization because they may lead to covariance matrix degeneration, i.e. to a situation where the elements of the covariance matrix become so small that sampling takes place in a too small area.

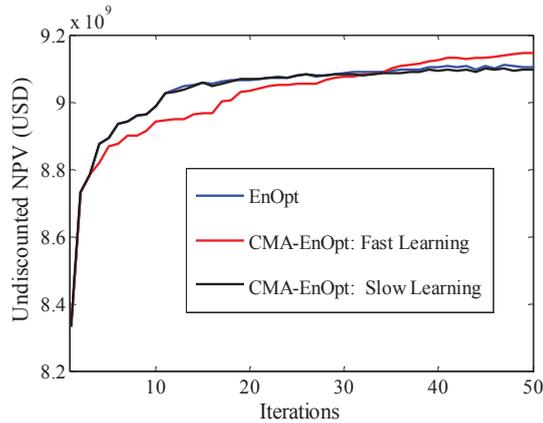


Fig. 4.4: Impact of optimization method and learning rates: EnOpt (blue curve), CMA-EnOpt with low learning rates (0.5%, black curve) and high learning rates (25%, red curve).

Fig. 4.5 illustrates estimated standard deviations σ for two control variables (i.e. the square root of the corresponding diagonal values of the distribution covariance matrix) for different learning rates. We observe from Fig. 4.5 that different control variables have different optimal standard deviations at each iteration. The CMA-EnOpt algorithm generally results in a gradual decrease of the standard deviation of a control when approaching the optimum. For control nr. 19 in Fig. 4.5 the magnitude of the corresponding gradient is an order higher than that of control nr. 299, indicating that either control nr. 19 is still much further from the optimum than control nr. 299, or that the objective function has a much higher curvature in the direction of control nr. 19. We note that of the 375 controls, only five controls have standard deviations like control 19, i.e. standard deviations higher than the initial standard deviation, with control 19 being the highest, while the remaining 370 controls have much lower standard deviations. The low learning rates, based on the recommendations for small-size problems as described in Hansen (2011) and Hansen (2006), lead to negligible changes in the standard deviation (black line, overlapped by the blue line in Fig. 4.5). In this case CMA-EnOpt behaves just like EnOpt with a fixed standard deviation (blue line) for all controls. Thus, learning rates have a significant impact on the performance of the optimization algorithm; if chosen too conservatively, the advantage of CMA-EnOpt over EnOpt will be negligible.

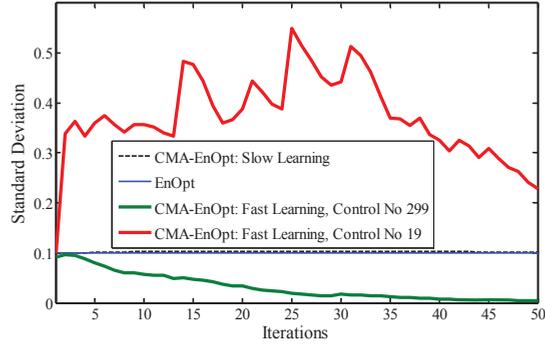


Fig. 4.5: Standard deviations for two control variables for different optimization methods and learning rates: CMA-EnOpt (high learning rates: red and green curves), CMA-EnOpt (low learning rates: black curve) and EnOpt (blue curve, nearly overlapping the black curve).

4.5.4 Correlations and Block-Diagonal Update

Imposing correlations over control times (effectively imposing smoothness on the control solution) may lead to an improved efficiency of the EnOpt algorithm when using many control time steps; see e.g. Chen et al. (2009), Leeuwenburgh et al. (2010), and Oliveira and Reynolds (2014). The CMA-EnOpt results presented thus far were obtained with a diagonal update of \tilde{C}_{uu} , using equation (4.6), without imposing smoothness on the controls. Fig. 4.6 illustrates the impact of introducing non-zero correlations on the control perturbations over time using the spherical correlation function as defined in Zhao et al. (2013) where the correlation length is set equal to the total number of control time steps. This results in the red curve i.e. a diagonal CMA-EnOpt update with an additional correlation (smoothing). The black curve represents the case without correlation. Although, in this case, the impact of imposing a correlation over time only marginally increases the objective function value, the red curve is nearly always above the black curve which suggests an improved computational efficiency.

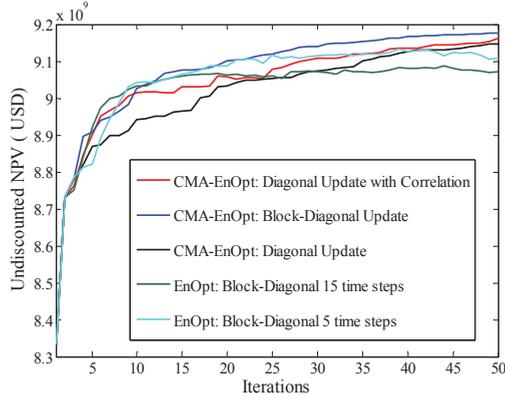


Fig. 4.6: Impact of imposing time correlations on the optimization process.

To obtain those results an arbitrary correlation length was chosen. The correlation length can have a significant impact on the results in different problems but unfortunately there is no pre-defined way of knowing the ideal correlation length, as shown in Oliveira and Reynolds (2014). The efficiency of using such a correlation function in conjunction with the diagonal update based CMA-EnOpt algorithm raises the question as to whether it is possible to estimate the optimal correlation between the controls using the CMA-EnOpt method. In order to investigate this we introduce a block-diagonal update of $\tilde{\mathbf{C}}_{uu}$ for the CMA-EnOpt method, using equation (4.5) where also off-diagonal covariance updates are allowed, but only for cross-covariances between controls belonging to the same ICV. Thus instead of using an artificial or pre-defined correlation function and length to impose smoothness on the controls, we allow the CMA-EnOpt algorithm to estimate the optimal cross-correlation between the controls, i.e. each ICV is associated with its own 15x15 covariance matrix (where 15 is the total number of control time steps), starting from a diagonal covariance matrix. The blue curve in Fig. 4.6 represents this ‘block-diagonal’ update, which performs better than both the diagonal update (black curve) and the diagonal update with smoothing (red curve). As a comparison, we also introduced time-correlations (smoothing) in EnOpt with the aid of a (constant) block-diagonal covariance matrix $\tilde{\mathbf{C}}$. Using different correlation lengths (green and yellow lines), we achieve higher objective function solutions in the earlier iterations but solutions with lower objective function values at the final iteration compared to CMA-EnOpt. Also in Fig. 4.4 EnOpt seems to perform better early-on in the optimization, which suggests to use EnOpt early-on and only switch to CMA-EnOpt later during an optimization run. However, it is unknown in advance what is the correct point during the optimization to switch between the two methods, and we did not pursue this possibility. Finally, we checked the influence of regularization by (double) pre-multiplication of the gradient \mathbf{g} with $\tilde{\mathbf{C}}$ (constant, for

EnOpt) or with \tilde{C}_{uu} (updated after every iteration, for CMA-EnOpt) similar to the premultiplication in equations (2.15) and (2.16). **Table 4.3** depicts the results, and it follows that, for this particular example, regularization generally does not have a positive effect and, in nearly all cases, results in somewhat lower objective function values.

Table 4.2: Effect of Regularization . All results are expressed in \$.

Gradient formulation for EnOpt	Gradient formulation for CMA-EnOpt	EnOpt: Block-diagonal, 5 time-step correlation	EnOpt: Block-diagonal, 15 time-step correlation	CMA-EnOpt Block-diagonal	CMA-EnOpt Diagonal, 5-time step correlation	CMA-EnOpt Diagonal, 15 time-step correlation
\mathbf{g}	\mathbf{g}	9.102e9	9.073e9	9.175e9	9.165e9	9.14e9
$\tilde{C}\mathbf{g}$	$\tilde{C}_{uu}\mathbf{g}$	9.103e9	9.035e9	9.001e9	9.132e9	9.05e9
$\tilde{C}\tilde{C}\mathbf{g}$	$\tilde{C}_{uu}\tilde{C}_{uu}\mathbf{g}$	9.088e9	9.018e9	8.983e9	9.112e9	8.925e9

4.5.5 Full Matrix Update

In the theory section we discussed the similarity between CMA-EnOpt and quasi-Newton methods, with the adapted covariance matrix being approximately similar to the inverse of the Hessian matrix. So far, however, we did not use the full covariance matrix for two reasons.

- Bouzarkouna et al. (2011) and Ros and Hansen (2008) show that for certain types of objective functions, like NPV, which can be decomposed into a sum of the individual NPVs from each well, an uncorrelated (diagonal) covariance matrix achieves better solutions in comparison to using a full covariance matrix.
- Because the updated covariance matrix is used to sample a new ensemble of controls for a gradient estimate of the next iteration, either an Eigen or Cholesky decomposition of the covariance matrix is needed to generate the ensemble of controls, which can be computationally demanding especially in problems of a high dimension.

In view of these points we, so far, used diagonal or block-diagonal updates. This approach is supported by the results shown in **Fig. 4.7** which compares the various covariance matrix update types available with CMA-EnOpt for this model.

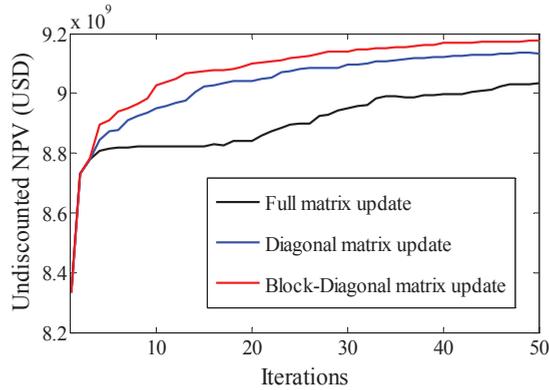


Fig. 4.7: Comparison of the impact of different matrix updates.

We observe that with the full matrix update (black curve) the achieved solutions are inferior to those resulting from the diagonal update (blue curve), which we believe is a result of the introduction of ‘unnecessary’ correlations between controls that should be, logically, uncorrelated (e.g. wells on opposite corners of the five-spot pattern). The block-diagonal update (red curve) of the covariance matrix achieves the best results. There are a number of advantages associated with the use of a block-diagonal update in this example:

- the obtained objective function values are higher than for other update types,
- the solutions are somewhat smoother, as shown in **Fig. 4.8**,
- there is no need to define a correlation length, and
- the computational burden of decomposing a large matrix is avoided.

In addition, some of the rotational aspects of the covariance matrix are retained, which is known to be especially useful when dealing with objective function search spaces that contain ridges, in which case the elongated sampling area may align itself with a ridge.

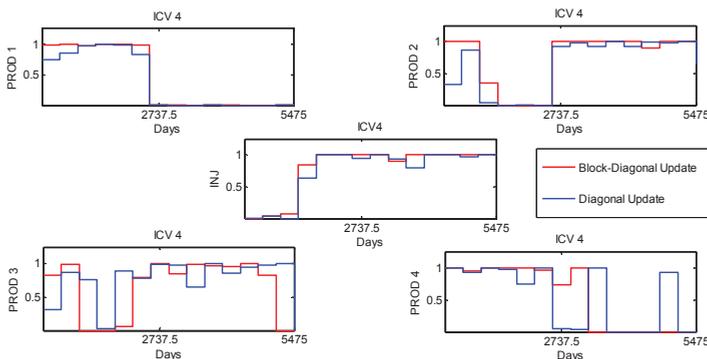


Fig. 4.8: Optimal control settings for ICV 4 corresponding to different CMA-EnOpt update types.

4.5.6 Comparison of Update types

In the CMA-EnOpt equations written above there are two different sources of information used for the matrix update, namely rank one and rank μ updates. While the literature on CMA-ES suggests that the former is more important in smaller ensembles and the latter in larger ensembles, we investigated which update is more useful for our case. **Fig. 4.9** illustrates the effect of the different matrix update types, and shows that for our problem the rank μ update (blue curve) performs significantly better than the rank one update (green curve), in terms of objective function value and computational efficiency. In fact, for this model and learning rates the rank one update has a detrimental effect on the optimization as can be seen in the combined rank update (black curve), the result of which lies between the different rank updates. We note that further testing is required to obtain full insight into the relationships between various aspects of the CMA-EnOpt algorithm, such as optimal learning rates and update strategies.

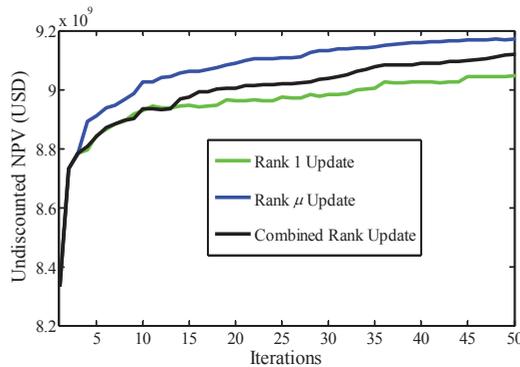


Fig. 4.9: Comparison of the impact of different rank updates. Rank 1 update: $c_\mu = 0$ and $c_1 = 0.05$. Rank μ update: $c_\mu = 0.20$ and $c_1 = 0$. Combined update: $c_\mu = 0.20$ and $c_1 = 0.05$.

4.5.7 Comparison with CMA-ES

The motivation for CMA-EnOpt was derived from the CMA-ES algorithm given in Hansen (2006). Therefore, an obvious next step would be the comparison of EnOpt and CMA-EnOpt with CMA-ES, i.e. with the evolutionary strategy that formed the basis for the covariance matrix adaptation strategy in CMA-EnOpt. Initial comparisons, illustrated in **Fig. 4.10**, indicate that the CMA-ES results are inferior to those of the two EnOpt varieties. CMA-ES achieves results which are approximately 2.5% lower than CMA-EnOpt for the diagonal update and 1.4% lower for the full matrix update, while also EnOpt achieves better results than CMA-ES for the same computational burden, i.e. 50 optimization iterations. Therefore, explicitly utilizing (approximate) gradient information seems to pay off for this problem (Note: tuning parameters such

as ensemble size, initial covariance matrix, learning rates etc. are exactly the same for all the results in Fig. 4.10.). However, further numerical comparisons are required to evaluate the strength and weaknesses of the three methods.

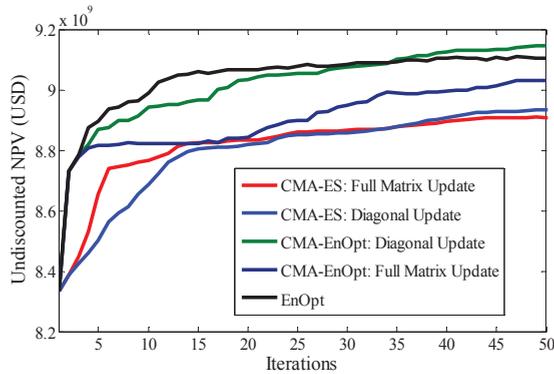


Fig. 4.10: Performance comparison of CMA-ES with CMA-EnOpt on the five-spot model with a full matrix update (red, purple) and a diagonal matrix update (blue, green).

4.5.8 Effect of Different Random Number Sequences

The results shown in this paper are based on single runs starting from the same initial random number seed. A comparison of stochastic methods requires multiple sets of comparisons with different random number sequences. Chen and Oliver (2012) have shown that sizeable variability in the results for different random seeds exists for cases where smaller ensembles are used to estimate the gradient. When they use a larger ensemble size, which is still smaller than the size used in our paper, the variability is reduced. We have run our experiments with five different random seeds, the results of which are displayed in Fig. 4.11. We observe the trends to be similar to the earlier results in our paper (i.e., CMA-EnOpt always outperforms EnOpt). Additionally there is a marginal reduction in variability when using CMA-EnOpt, but we may need more runs to conclude definitively in this regard. Note: the scale of the y-axis in Fig. 4.11. The range of NPV values which defines the variability is rather small, so even for EnOpt the variability is not very large, which is in accordance with the results shown in Chen and Oliver (2012).

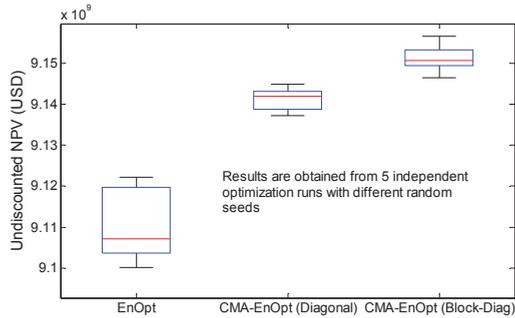


Fig. 4.11: Box plot to illustrate the effect of different random seeds on the stochastic ensemble optimization method. Note the scale of the y-axis which implies that the NPV values are quite close. Red lines: median. Whiskers (black): maximum and minimum values. Box edges (blue): 25 and 75 percentiles.

4.6 Modified Brugge Model

To test our results on a more complex and challenging reservoir model we selected a modified version of the Brugge benchmark model (Peters et al. 2010, 2013). The model, shown in **Fig. 4.12**, consists of 60,048 grid blocks. It represents a segment of an anticlinal structure with one major fault. The reservoir is produced using a peripheral well pattern, i.e. 10 injectors on the down-flank of the structure and 20 producers towards the top of the structure. The reservoir is divided into nine layers with varying permeabilities and porosities. All rock and fluid properties were chosen identical to those in the original Brugge model (Peters et al. 2010, 2013). However, the oil-water contact has been lowered to 1780 m (compared to an original contact depth of 1678 m), thus increasing the STOIP of the model, while the corresponding well locations are also adjusted as shown in Fig. 4.12. We consider a control problem where ICV settings of injection and production wells are manipulated to optimize waterflooding over the producing life of the reservoir which is set to 20 years, or 7200 days.

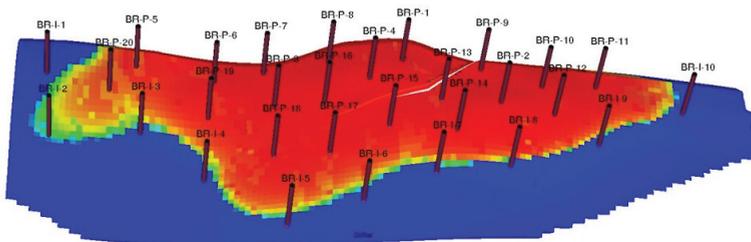


Fig. 4.12: Modified Brugge reservoir model. The colors indicate the initial water saturation.

The wells penetrate all nine layers with 3 ICVs in every well (except where bottom ICVs would be placed in the water leg) resulting in a total of 87 controls per time step.

The producing life of the reservoir is divided into 20 time intervals of one year (360 days) each, which results in a total of 1740 controls to be optimized. Water is injected at a constant rate of 3000 m³/day and the production wells are operated at a maximum liquid rate of 1200 m³/day. The bottom hole pressure limits were 190 bar for injectors and 40 bar for producers. All wells are assumed to be operational from the start of field production. We used an oil price $r_o = 140$ \$/m³, a water production cost $r_{wp} = 30$ \$/m³, and a water injection cost $r_{wi} = 10$ \$/m³. The discount rate b was set to 0. Well index multipliers were used to model the ICVs in the simulator with bounds of 1×10^{-4} and one. The starting point for the optimization was an initial control vector having values equal to one. The initial value of σ was equal to 0.1 and we used a fixed ensemble size of 50 samples. Perturbed control values outside the range $[1 \times 10^{-4}, 1]$ were simply reset to their bounds. We used a commercial fully implicit finite difference black oil simulator (Eclipse 2011).

4.6.1 Results

Because this model is significantly more complex and larger than the small model discussed before, we chose to perform only a limited number of optimization experiments. **Fig. 4.13** shows a comparison between EnOpt (black curve), CMA-EnOpt with only diagonal updates (blue curve) and CMA-EnOpt with block diagonal updates (red curve). The optimization was run for 50 iterations, i.e. 2500 reservoir simulations. User-defined parameters and initial strategies were the same for all algorithms. Although neither of the CMA-EnOpt methods had converged yet to an optimum after 50 iterations, CMA-EnOpt with block diagonal updates clearly outperformed EnOpt, with an increase of 10.5% in objective function value for the same number of reservoir simulations.

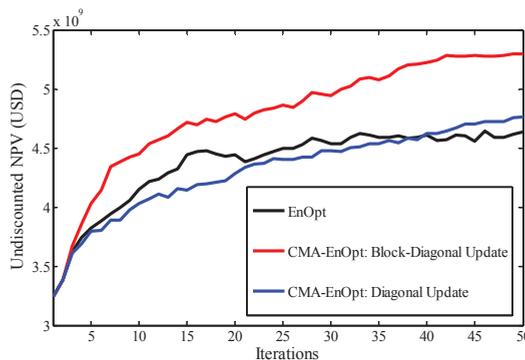


Fig. 4.13: Comparison of different methods for the modified Brugge model.

The results shown in Fig. 4.13 confirm the results obtained from the simple five-spot model. The increased robustness to the choice of the initial covariance matrix is seen to

similarly apply for this case. In addition we observe that the relative increase in objective function value for the complex Brugge model is significantly higher than for the simple five-spot model.

4.6.2 Robustness to the Initial Covariance Matrix

The choice of a good covariance matrix in EnOpt is a matter of trial and error, which for the modified Brugge case requires a significant computational effort. The main idea of CMA-EnOpt is to avoid the trial and error procedure and still achieve results of practical importance. We re-ran the simulations for the modified Brugge example for two additional initial choices of the covariance matrix, the results of which are summarized in **Table 4.2**. We observe that, irrespective of the initial choice of the covariance matrix, the trend in the results (i.e. CMA-EnOpt performing better than EnOpt) is similar to those in Fig. 4.13. We also observe that the initial choice of the covariance matrix still plays a role in the optimization, but that we achieve better results for CMA-EnOpt. The EnOpt (block-diagonal) result is from an experiment with a constant covariance matrix with a correlation length of five time steps for the correlation function described in Zhao et al. (2013). Varying the correlation length may lead to superior or inferior results, which is not known a-priori and hence would require trial and error experiments. This has not been pursued since the experiments are computationally expensive. For the three different choices of the initial covariance matrix using block-diagonal covariance matrices (i.e. time-correlated covariance matrices) CMA-EnOpt has done better than EnOpt for two choices of covariance matrices while for the other choice, it is marginally inferior. Note: due to the significant computational effort required for this model the effect of different random seeds on the results has not been tested. The stopping criteria for the experiments was the maximum number of iterations, in this case 40 iterations.

Initial Sigma	EnOpt	CMA-EnOpt (diagonal)	EnOpt (block-diagonal)	CMA-EnOpt (block-diagonal)
1	4.58e9	4.97e9	5.33e9	5.42e9
0.1	4.6e9	4.63e9	5.26e9	5.23e9
0.01	4.49e9	4.45e9	4.87e9	5.32e9

4.6.3 Update Comparison

In the simple five-spot example the rank- μ update for CMA-EnOpt with a block-diagonal update scheme performed best. We performed a similar experiment on the more complex modified Brugge case and observed different results as illustrated in

Fig. 4.14. For this model the rank one (green line) update performs better than the rank μ (blue line) update while after 40 iterations the combined rank (black line) update performs best. Note that the rank μ update has a higher weight (0.20) than the rank one update (0.05) similar to the values used in the previous example.

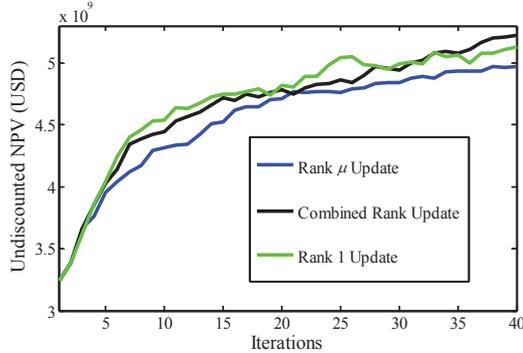


Fig. 4.14: Comparison of the rank updates for the modified Brugge model.

Thus for the Brugge model the combined rank update (black line in Fig. 4.14) eventually achieves the highest objective function value while the rank μ -only update achieves the worst solution for this case. More experience is needed to arrive at general recommendations on which type of update is most suited for a particular case.

4.7 Conclusions

- ✓ A comparison between CMA-EnOpt and EnOpt for a simple five-spot model showed consistently (somewhat) higher objective function values and modest speed-ups for CMA-EnOpt, depending on the choice of user-defined parameters in both algorithms.
- ✓ The major benefit of CMA-EnOpt is its robustness with respect to the initial choice of the covariance matrix. A poor choice of the initial matrix can be detrimental to EnOpt, whereas the CMA-EnOpt performance is near-independent of the initial choice.
- ✓ Learning rates are crucial for the success of CMA-EnOpt. For both the simple five-spot model and the modified Brugge model, a 75%-25% update rule proved to be successful.
- ✓ For the simple five-spot model, the methods that explicitly use gradient information (EnOpt and CMA-EnOpt), performed better than the method that doesn't do so (CMA-ES). (Not tested for the modified Brugge model).
- ✓ A comparison between CMA-EnOpt and EnOpt for the modified Brugge model revealed slightly lower to significantly higher (-1% - +9%) objective

function values depending on choice of user-defined parameters in both algorithms.

- ✓ Updating a block-diagonal (i.e. time-correlated) covariance matrix leads to significant improvements in the results as well as in the efficiency of the algorithm, compared to using a prescribed correlation (smoothing) and compared to updating either diagonal elements only, or updating the full matrix.
- ✓ The different rank updates play different roles in the success of the optimization; for the simple five-spot model the rank μ update performed much better than the rank one update, with a combined rank update ending up in-between. For the complex Brugge model, however, the rank μ update performed worst, while the combined-rank update performed best. Further experience is needed to arrive at general recommendations.
- ✓ Robustness to the choice of the initial covariance matrix, and higher objective function values are the main advantages of CMA-EnOpt over EnOpt.

5

HIERARCHICAL BI-OBJECTIVE OPTIMIZATION

Van Essen et al. (2011) proposed two hierarchical multi-objective methods to include short-term targets in life-cycle production optimization. However the work by Van Essen et al. (2011) had two limitations: 1) the adjoint formulation is used to obtain gradient information, requiring simulator source code access and an extensive implementation effort, and 2) one of the two proposed methods relies on the Hessian matrix which is obtained by a computationally expensive method. In order to overcome the first of these limitations, we used ensemble-based optimization (EnOpt). EnOpt does not require source code access and is relatively easy to implement. To address the second limitation, we used the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm to obtain an approximation of the Hessian matrix. We performed experiments in which a water flood was optimized in a geologically realistic multi-layer sector model. The controls were inflow control valve settings at pre-defined time intervals. Undiscounted Net Present Value (NPV) and highly discounted NPV were the long-term and short-term objective functions used. We obtained an increase of approximately 14% in the secondary objective for a decrease of only 0.2-0.5% in the primary objective. We also applied the same techniques to perform hierarchical optimization under geological uncertainty. The chapter demonstrates that ensemble-based hierarchical bi-objective optimization with and without geological uncertainty can achieve results of practical value in a computationally efficient manner.⁴

⁴ Based on Fonseca, R.M., Leeuwenburgh, O., Van den Hof, P.M.J. and Jansen, J.D. 2014. Ensemble based Hierarchical Multi-Objective Production Optimization of Smart Wells. *Computational Geosciences* **18**(3-4).

5.1 Introduction

Operational decisions in reality are generally based on short-term objectives of a project in terms of operational criteria, production contracts etc., and strategies to optimize such objectives are often in conflict with optimal long-term strategies. Jansen et al. (2009) observed that significantly different optimized long-term water flooding strategies result in nearly equal values of the objective function, defined as net present value (NPV). They concluded that the life cycle optimization problem is ill-posed and contains redundant degrees of freedom (DOFs). Thus, there exist multiple solutions to the optimization problem, and different initial starting points may lead to different solutions in an optimal subset of the decision variable space. Oliver et al. (2008) illustrated a similar problem for optimization problems that appear in computer-assisted history matching which are well known to be very ill-posed. The presence of redundant DOFs formed the basis of the multi-objective optimization approach of Van Essen et al. (2011). They suggested a hierarchical optimization scheme to include secondary objectives into the life cycle optimization using the adjoint formulation. They observed a significant increase in short-term objectives with minimal change to the primary objective function, similar results were obtained by Chen et al. (2012) and Suwartadi et al. (2012). In this chapter we investigate the applicability of the EnOpt method instead of the adjoint method for multi-objective optimization. We also propose a modification of the hierarchical optimization scheme which improves the computational efficiency of the algorithm. This chapter provides a practical and relatively easy to implement alternative to adjoint based multi-objective optimization for cases with and without geological uncertainty being incorporated into the optimization framework.

5.2 Life Cycle Optimization

Life cycle optimization of hydrocarbon recovery requires at least one decision variable as well as a model that provides relatively accurate long-term predictions. The most widely implemented secondary recovery mechanism in the petroleum industry is water flooding. Water flooding is our choice of recovery mechanism for the following reasons.

- There are many decision variables involved in a water flooding strategy.
- The process is well understood and can be modeled accurately over long time intervals.

Recent improvements in technology have led to an increase in the application of downhole chokes or inflow control valves to regulate flow rates and maintain pressure in the reservoir. Smart wells, i.e. wells with Inflow Control Valves (ICVs), are

important tools to increase oil recovery and delay water production in multi-zone reservoirs. ICVs can also be used to manipulate the stream lines in a single heterogeneous reservoir thus sweeping the reservoir more efficiently (Jansen, 2011). Thus it is important to find the optimum value for the settings of the ICVs so as to maximize their potential. In order to improve economic life cycle performance dynamic optimization has to be performed over the producing life of the reservoir due to the slowly changing nature of the saturation distribution. We consider a control problem where ICV settings can be manipulated to achieve the best possible objective function.

5.3 Hierarchical Bi-objective Optimization

The process of optimizing systematically and simultaneously a collection of objective functions is called multi-objective optimization. There are various methods of multi-objective optimization such as the weighted sum method, Pareto optimality, goal programming etc. Van Essen et al. (2011) introduced a hierarchical optimization scheme to solve the multi-objective production optimization problem. They argued that the weighted sum method (in which each objective function is assigned a weight factor) suffers from an arbitrariness in choosing the weight factors and therefore preferred a hierarchical method (in which the second objective is optimized while keeping the optimized value of the first objective (almost) fixed). They proposed two hierarchical methods in combination with the adjoint formulation, which are explained below. We have, in this work, used those hierarchical methods to investigate their applicability in combination with the EnOpt method. .

5.3.1 Null-space based optimization

Van Essen et al. (2011) introduced a hierarchical optimization scheme to achieve multi-objective production optimization, which prioritizes the objective functions. The optimization of the secondary objective function J_2 is constrained by a maximum allowable change in the primary objective function. Thus the primary objective function J_1 will remain close to its optimal value. A general formulation for hierarchical optimization is as follows

$$\begin{aligned}
 & \max_{\mathbf{u}_{1:K}} J_2(\mathbf{u}_{1:K}), \\
 & \text{s.t. } \mathbf{f}_{k+1}(\mathbf{u}_k, \mathbf{x}_k, \mathbf{x}_{k+1}) = \mathbf{0}, \quad k = 0, \dots, K-1, \\
 & \mathbf{c}_{k+1}(\mathbf{u}_{k+1}, \mathbf{x}_{k+1}) \leq \mathbf{0}, \quad k = 0, \dots, K-1, \\
 & J_1^* - J_1(\mathbf{u}_{1:K}) \leq \varepsilon,
 \end{aligned} \tag{5.1}$$

where \mathbf{u} is the control vector or input vector (ICV settings), \mathbf{x} is the state vector (grid block pressures and saturations), \mathbf{f} is a vector-valued function that represents the system equations, \mathbf{x}_0 is the state vector representing the initial state of the reservoir, the subscript k indicates discrete times and K is the total number of time steps. The vector of inequality constraints \mathbf{c} concerns production system capacity limitations or operational constraints. The parameter $\varepsilon > 0$ has an appropriately small value compared to J_1^* . Solving the above equations requires the knowledge of J_1^* which is the optimized value of J_1 obtained from the primary objective optimization. Thus the hierarchical optimization constrains the optimization of the secondary objective with respect to the primary objective function. The ordering of the different objective functions is not unique, thus secondary objectives can be implemented as primary objectives and vice versa. This hierarchical scheme is attractive when there is a presence of redundant degrees of freedom in the primary objective function. To exploit these degrees of freedom we require the Hessian of the primary objective function. Some concepts detailing the need for this Hessian and methods to exploit the redundant degrees of freedom are explained in the appendix.

5.3.1.1 Approximate Hessian

Van Essen et al. (2011) proposed the use of a finite difference scheme in combination with the adjoint formulation to approximate the second order derivatives of the objective function. Without the adjoint formulation, to estimate a finite difference based Hessian we require $n \times (n+1)$ function evaluations where n is the number of controls. Thus this method is computationally infeasible for realistic reservoir models and large numbers of controls, due to the high number of function evaluations needed. To alleviate this short-coming we propose to use the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm which approximates and updates the Hessian during the optimization of the primary objective function. This leads to a significant reduction in overall computational costs incurred during optimization of the secondary objective function. We note that Dehdari and Oliver (2012) use the BFGS Hessian in the context of constrained production optimization (as an alternative to steepest ascent), while we use the Hessian only to get information about the null space of the primary objective.

5.3.1.2 BFGS algorithm

From an initial guess \mathbf{u}^0 and an approximate Hessian matrix, \mathbf{H}^0 the following steps are repeated until \mathbf{u} converges:

- ✓ Calculate the gradient \mathbf{g} at \mathbf{u}^ℓ and update the set $\mathbf{u}^\ell = \mathbf{u}^{\ell-1} + \alpha^\ell \mathbf{g}^\ell$.
- ✓ Set $\mathbf{s}^\ell = \mathbf{u}^\ell - \mathbf{u}^{\ell-1}$.

- ✓ Compute $\mathbf{y}^\ell = \mathbf{g}^\ell - \mathbf{g}^{\ell-1}$.
- ✓ Compute the updated Hessian as

$$\mathbf{H}^\ell = \mathbf{H}^{\ell-1} + [\mathbf{y}^\ell (\mathbf{y}^\ell)^T] / [(\mathbf{y}^\ell)^T \mathbf{s}^{\ell-1}] - [(\mathbf{s}^{\ell-1} \mathbf{H}^{\ell-1})^T \mathbf{s}^{\ell-1} \mathbf{H}^{\ell-1}] / [(\mathbf{s}^{\ell-1})^T \mathbf{H}^{\ell-1} \mathbf{s}^{\ell-1}]$$

$J(\mathbf{u})$ denotes the objective function to be minimized. Practically, \mathbf{H}^0 can be initialized with $\mathbf{H}^0 = \mathbf{I}$, so that the first step will be equivalent to a gradient descent, but further steps are more and more refined by \mathbf{H}^ℓ , the approximation to the Hessian. Note that because we typically take fewer iteration steps than there are DOFs, the BFGS approach will only lead to an approximate Hessian.

5.3.2 Null-space-based optimization algorithm

The algorithm is a modification of the algorithm proposed by Van Essen et al. (2011). The modification is the implementation of the BFGS algorithm to approximate the Hessian matrix.

- ✓ Find an optimal strategy \mathbf{u}^* for the primary objective function J_1 and set $\mathbf{u} = \mathbf{u}^*$ with $n = 0$ as a starting point for the secondary optimization problem where n is the iteration index.
- ✓ Use the approximated Hessian \mathbf{H} at \mathbf{u}^* and perform a singular value decomposition to obtain the orthonormal basis \mathbf{B} for the null-space of \mathbf{H} .
- ✓ Form the projection operator \mathbf{P} according to

$$\mathbf{P} = \mathbf{B}\mathbf{B}^T.$$

- ✓ Find the gradient \mathbf{s} for the secondary objective function J_2 .
- ✓ Project this improving direction \mathbf{s} onto the orthonormal basis \mathbf{B} to obtain the projected direction \mathbf{d} , such that \mathbf{d} is an improving direction for J_2 and does not affect J_1 . Thus \mathbf{d} is

$$\mathbf{d} = \mathbf{P}\mathbf{s}.$$

- ✓ Update the control vector \mathbf{u} using the projected direction \mathbf{d} in the steepest-ascent method.

$$\mathbf{u}^{n+1} = \mathbf{u}^n + \alpha^n \mathbf{d}^n$$

where α is an appropriately small step size.

- ✓ Update \mathbf{H} using the BFGS algorithm for the new set of controls.
- ✓ Perform steps 2 to 6 until convergence of J_2

The EnOpt algorithm has been used to approximate the gradient of the secondary objective function \mathbf{s} . Further details on the algorithm sketched above are provided in the appendix. Note that an infinitesimally small update of J_2 , projected on the null-space of J_1 , would result in no change in the value of J_1 (by definition). However, because we use an approximate Hessian, and because we take finite updates of J_2 , small changes in the value of J_1 may occur.

5.3.3 Switching algorithm

The null-space based hierarchical algorithm presented above is computationally cumbersome and not feasible for realistic reservoir models having a large number of input parameters when implemented using a finite difference based Hessian. To overcome this short-coming, Van Essen et al. (2011) presented a practical alternative method to the null-space based hierarchical algorithm with the use of a switching function according to

$$J_{switch} = \Omega_1 J_1 + \Omega_2 J_2, \quad (5.2)$$

where Ω_1 and Ω_2 are switching functions for J_1 and J_2 that take on values of 1 and 0 or vice versa:

$$\begin{aligned} \Omega_1(J_1) &= \begin{cases} 1 & \text{if } J_1^* - J_1 > \varepsilon, \\ 0 & \text{if } J_1^* - J_1 \leq \varepsilon, \end{cases} \\ \Omega_2(J_1) &= \begin{cases} 0 & \text{if } J_1^* - J_1 > \varepsilon, \\ 1 & \text{if } J_1^* - J_1 \leq \varepsilon. \end{cases} \end{aligned} \quad (5.3)$$

Here ε is the threshold value as defined in the inequality constraint in equation (5.1) and J_1^* is the value of the primary objective at the optimal solution achieved during life cycle optimization. The (transposed) gradient of J_{switch} with respect to the input parameters is then given by

$$\frac{dJ_{switch}^n}{d\mathbf{u}} = \Omega_1(J_1^n) \frac{dJ_1^n}{d\mathbf{u}} + \Omega_2(J_1^n) \frac{dJ_2^n}{d\mathbf{u}}. \quad (5.4)$$

The use of a balanced objective function in the optimization will give improving directions for either J_1 or J_2 , thus switching between feasible and infeasible solutions. However, the convergence towards an optimal solution may be rather slow due to the switching between the different solutions. In order to improve convergence speed Van Essen et al. (2010) suggested the following adaptation in which gradients of the secondary objective function are projected onto the null space of the optimal primary objective function:

$$\mathbf{P} = \frac{dJ_1}{d\mathbf{u}} \left(\frac{dJ_1^T}{d\mathbf{u}} \frac{dJ_1}{d\mathbf{u}} \right)^{-1} \frac{dJ_1^T}{d\mathbf{u}}. \quad (5.5)$$

In the neighborhood of the optimum, the complement of the gradient with respect to the primary objective function can be used as a first-order approximation to the null space of the Hessian of this function. Thus the improved direction for the secondary objective is given by

$$\mathbf{a} = (\mathbf{I} - \mathbf{P}) \frac{dJ_2^T}{d\mathbf{u}}, \quad (5.6)$$

Hence the alternative switching search direction \mathbf{d} for solving the hierarchical problem is

$$\mathbf{d}_{n+1} = \Omega_1(J_1^n) \left(\frac{dJ_1^n}{d\mathbf{u}} \right)^T + \Omega_2(J_1^n) (\mathbf{I} - \mathbf{P}) \left(\frac{dJ_2^n}{d\mathbf{u}} \right)^T. \quad (5.7)$$

The EnOpt method is used to approximate the gradients for both the primary and the secondary objective functions. The two hierarchical methods presented above are tested on a geologically realistic sector model explained below.

5.4 Results : Bi-Objective (nominal) Optimization

In this section we use the 5-spot synthetic reservoir model described in **section 4.5** for the experiments. An optimal life-cycle strategy of ICV settings for the individual layers is obtained by optimizing the NPV as described in equation (2.1), with $r_o = 130$ $\$/\text{m}^3$, $r_{wp} = 25$ $\$/\text{m}^3$, $r_{wi} = 6$ $\$/\text{m}^3$. The discount rate b was set to 0. **Fig. 5.1** is an illustration of the optimization with undiscounted NPV as the objective function, which is equivalent to cumulative cash flow over the producing life of the reservoir. The optimal solution \mathbf{u}^* was obtained using an ensemble size of 50. Well productivity index (PI) multipliers are used to model ICVs in the simulator with bounds of 10^{-4} and 1. The starting point for the optimization is an initial control vector having values equal to 1. Thus all the ICVs are open as a starting strategy. The optimization was allowed to run for 80 iterations although there was no significant improvement in objective function value after 65 iterations as indicated in Fig. 5.1. The optimized value of the objective function is 8.902×10^9 $\$$. Additional iterations were performed to allow the BFGS algorithm to estimate a Hessian matrix that is as close to the true Hessian at the optimum as possible.

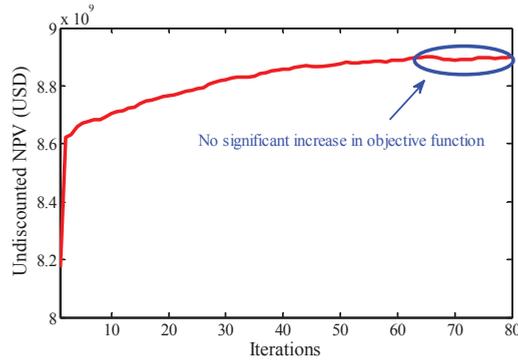


Fig. 5.1: Life cycle optimization with undiscounted NPV as the objective function.

5.4.1 Objective functions

We use undiscounted NPV as the primary objective function. However every project also aspires to recover the initial investments as soon as possible. In an ideal case we would like to additionally meet such short-term economic objectives while still maintaining the life cycle objectives. Thus we choose a secondary objective function which highlights the importance of maximizing short-term production. The secondary objective function has the same cost structure as the primary objective function but with a very high discount rate b of 25%.

5.4.2 Unconstrained optimization

First, the secondary objective function J_2 is optimized without being constrained by the primary objective function J_1 . This case serves as a comparison to the hierarchical structure explained in the theory. The optimization was performed with an ensemble size of 75 members and was allowed to run for 65 iterations. The results are illustrated in **Fig. 5.2**. The starting point of the optimization is the optimal solution achieved during optimization of the primary objective. A decrease of 1.65 % is seen in the primary objective function to achieve an increase of 14.2% in the secondary objective function. Note that we did not optimize J_2 until convergence. However, the increase of J_2 with 14.2 % is, by design, just equal to the increase obtained with the hierarchical optimizations discussed below.

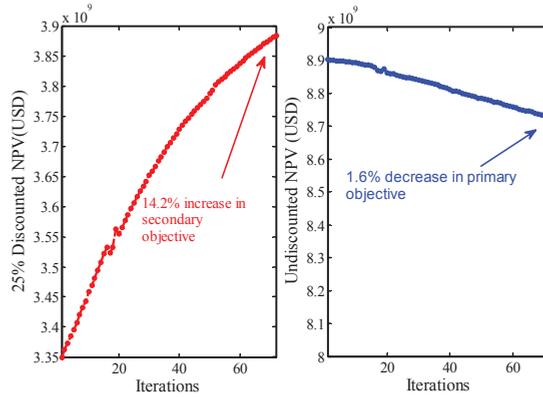


Fig. 5.2: Illustration of an unconstrained optimization of secondary objective (red) starting from the optimal strategy u^* obtained from life cycle optimization.

5.4.3 Null-space based optimization

Fig. 5.3 illustrates the optimization of the secondary objective function within the null space of the primary objective function. Since, typically, none of the singular values of the Hessian are exactly equal to zero, a cut-off criterion must be defined to estimate the null space of the Hessian matrix. We have used a cut-off criterion of $\sigma_i / \sigma_1 < 2 \times 10^{-9}$, where σ_i are the singular values of the Hessian matrix, with the values arranged from largest to smallest. The resulting null space consists of 187 vectors and its dimension remains almost constant throughout the optimization. The value of the secondary objective function at the optimum of the primary objective is 3.35×10^9 \$, which is the starting point for the null-space based optimization. We achieve a value of 3.823×10^9 \$ after completing the null-space based optimization equivalent to a 14.2% increase in the secondary objective function at the price of a 0.52% decrease in the primary objective function. This decrease is much smaller than that obtained with the unconstrained optimization approach and clearly illustrates the advantage of using the hierarchical multi-objective optimization approach. Another illustration of the impact of the null-space-based optimization is provided in **Fig. 5.4** which shows a comparison of the cumulative cash flow over time resulting from the optimal life-cycle strategy (green), the strategy resulting from null-space-based optimization (red) and from unconstrained optimization of the secondary objective (blue). A considerable increase in the short to medium term cumulative cash flow can be observed in the results of the null-space-based hierarchical optimization compared to those of the life-cycle optimization.

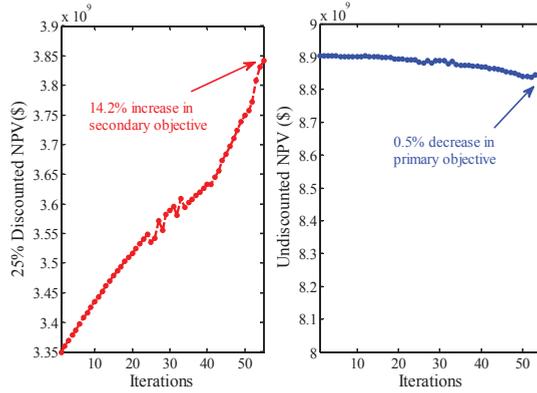


Fig. 5.3: Comparison of the primary (blue) and secondary (red) objective functions obtained by the null-space based optimization algorithm

The multi-objective optimization is seen to be useful to increase cash flow in the initial stages of the project whilst maintaining the life cycle goals. The unconstrained optimization, as expected, has the best cumulative cash flow in the early years of the economic life. However, as shown in the inset figure, compromises are made to the life-cycle target. The inset plot in Fig. 5.4 also shows that the null-space based optimization achieves a solution which performs better in the long-term compared to the unconstrained optimization.

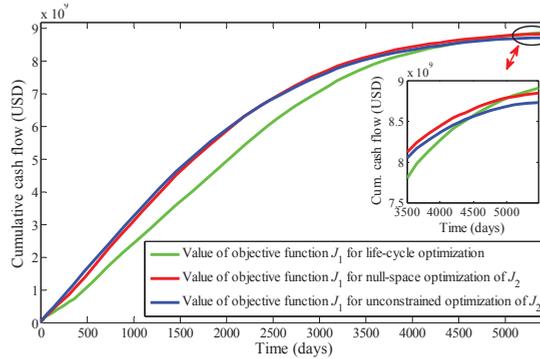


Fig. 5.4: Comparison of cumulative cash flows over time for the different optimization strategies: life-cycle optimized strategy (green), null-space based optimized strategy (red) and unconstrained secondary optimized strategy (blue).

5.4.4 Switching algorithm

The mathematical formulation of the switching algorithm described in the theory requires the definition of a criterion ε . The criterion used is $\varepsilon = 0.003 J^*$, i.e. we allow a maximum decrease of 0.3 % in the primary objective function value. An ensemble size of 75 samples was used and the optimization was allowed to run for 50 iterations; although after 35 iterations the improvements in the secondary objective became

minimal. The results are shown in **Fig. 5.5**. The algorithm achieves an increase of 14.17% in the secondary objective for a corresponding 0.21% decrease in the primary objective. The switching algorithm thus performs very well when applied to this model. Since the performance is dependent on parameters such as ensemble size used for gradient evaluation, step length, and line search parameters, fine-tuning of these parameters may lead to further improved performance of the switching algorithm.

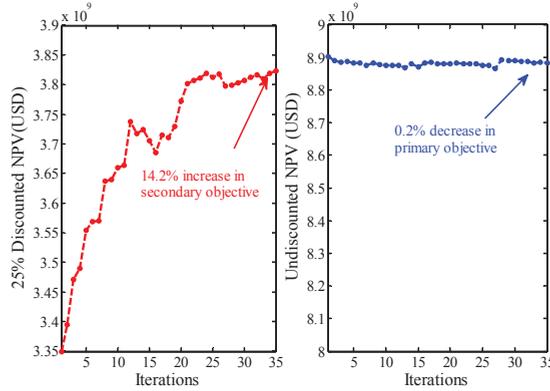


Fig. 5.5: Switching optimization showing a 14.2% increase in secondary objective function (red, left) for a corresponding 0.2% decrease in the primary objective (blue, right).

Similar to Fig. 5.4 for the null-space based optimization, **Fig. 5.6** compares the cumulative cash flow over time for the switching algorithm with the unconstrained and life-cycle only optimization results. It is observed that after 500 days the cumulative cash flow with life cycle optimization is approximately 1×10^9 \$. However the control strategy obtained with switching optimization achieves a cash flow of 1.5×10^9 \$. This increase of 0.5×10^9 \$ over 500 days will enable the project to achieve the break-even point faster. Similar to the results obtained in Van Essen et al. (2011) and the results shown above, the NPV at the end time of the unconstrained optimization (blue curve) is decreased by 1.6%. In comparison, the NPV obtained by the switching algorithm (red curve) decreases by only 0.2%. Finally, in **Fig. 5.7** the set of controls obtained by the optimization using the switching algorithm (red) is compared to the optimal set of controls obtained after life cycle optimization (black). The control sets are fairly different, so it can be concluded that rather different control sets may achieve very similar results for the primary optimization while drastically improving the secondary optimization.

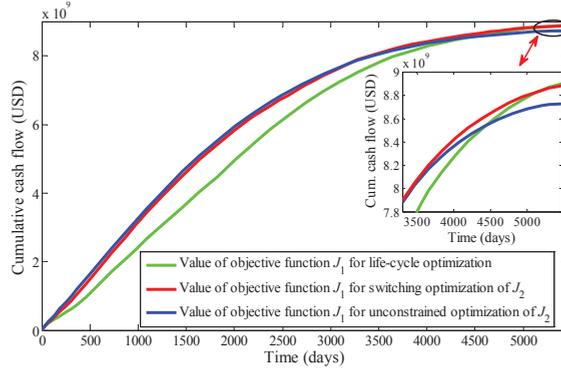


Fig. 5.6: Comparison of cumulative cash flow over time for the switching (red), unconstrained (blue) and optimal life-cycle (green) strategies.

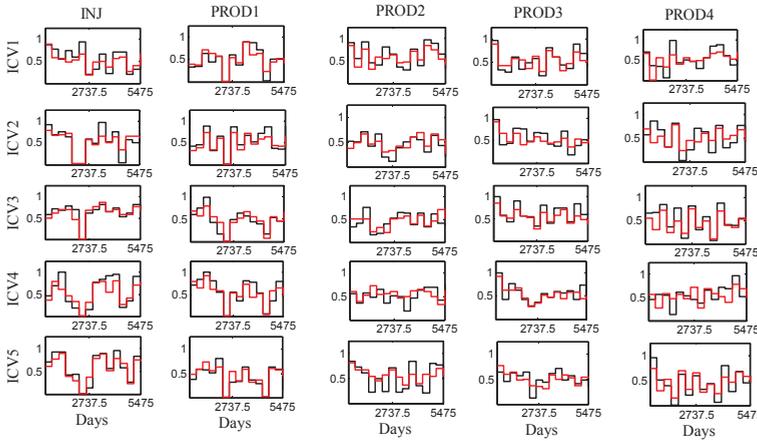


Fig. 5.7: Comparison of the control strategy for the switching algorithm (red) with the control strategy obtained during life cycle optimization (black), for the individual ICVs.

5.4.5 Computational aspects

In our example we use an ensemble size of 100 members. The primary objective function J_1 was maximized in 80 iterations, which therefore would require at least $(100 + 1) \times 80 = 8080$ forward simulations. The use of back-tracking when the value of J_1 did not increase during a specific iteration increased the actual number of forward simulations to 8127. For the optimization of the secondary objective J_2 we used 75 ensemble members and 55 iterations for the null-space method, corresponding to a theoretical minimum number of $(75 + 1) \times 55 = 4180$ forward simulations, while we actually used 4215 simulations because of back-tracking. For the switching method we only used 35 iterations, corresponding to $(75 + 1) \times 35 = 2660$ (theoretical minimum) and 2698 (actual) forward simulations.

5.4.6 Discussion

The results for the model used here show that the switching method achieves better results compared to the null-space-based method in that the same improvement in secondary objective was obtained with fewer iterations and with a smaller decrease in the primary objective. It is not possible at this point, however, to draw any definitive general conclusion regarding the comparative performance of these methods for more complex models. The choice for the different objective functions may have a large impact and significantly affect the scope for multi-objective optimization. The Hessian approximation with the BFGS algorithm has shown to achieve good results for this case but may perform differently in another case. The cut-off criterion used to define the dimension of the null space, as discussed in **section 5.4.3**, was chosen through trial and error in this work. However this criterion was found to be very important for the success of the null-space optimization (not shown). Another parameter that influences the results is ε , i.e. the maximum allowed decrease in the primary objective. Note that here we used ε only for the switching method, but that it could also be used for the null-space method.

5.4.7 Reactive control

For the given oil price of 130 \$/bbl and water production costs of 25 \$/bbl, the economic feasibility threshold will be achieved at a water cut (WCT) of 83%. Thus a reactive control strategy can be defined with a well shut-off limit of 83% WCT. In such a reactive strategy, the wells are initially operated with all the ICVs fully open and whenever the WCT limit is reached in a well it is shut-in. As expected the reactive control strategy has a much lower NPV at the end of the life of the reservoir. **Fig. 5.8** is a comparison of the cumulative cash flow for the reactive control strategy to the life-cycle and the switching based optimized strategies. The strategy obtained with the switching method (red) achieves an improved short-term performance compared to optimized life-cycle strategy (green line), but it is not as good as the reactive control strategy (blue), which gives the best short-term performance. Thus the two multi-objective optimization methods presented improve the short-term/secondary objectives but do not truly recover the best possible short-term/secondary objectives as obtained with the reactive strategy. (Note that the unconstrained optimization of J_2 also leads to short-term results worse than those obtained by the reactive strategy, which is probably because we did not iterate to convergence, and/or because we were heading towards a local optimum different from the optimum found in the reactive strategy. However, Fig. 5.8 does confirm the advantage of optimal life-cycle strategies in comparison to a reactive control based operational strategy when long-term objectives are important.

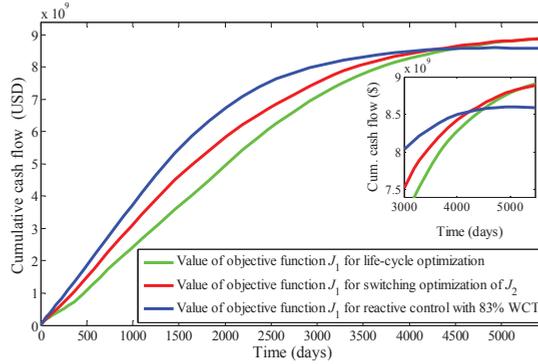


Fig. 5.8: Comparison of the cash flow over time for the switching (red), reactive control (blue) and optimal life-cycle (green) strategies.

5.5 Results: Bi-Objective Optimization under Uncertainty

In this section we use the Egg model and the accompanying ensemble of equi-probable permeability realizations described in [section 3.4](#) to test the applicability of EnOpt for bi-objective optimization incorporating geological uncertainty.⁵

5.5.1 Life-cycle optimization

Water injection rates are the controls to be optimized with a maximum allowable injection rate per well fixed at $60 \text{ m}^3/\text{day}$ and a minimum rate of $0 \text{ m}^3/\text{day}$. The producers are operated at a minimum bottom hole pressure of 385 bars without rate constraints. The producing life of the reservoir is divided into 40 optimization time intervals of 90 days each, and the control vector \mathbf{u} has therefore $N = 8 \times 40 = 320$ elements. An optimal life-cycle strategy of injection rates for the individual wells is obtained by optimizing the NPV as described in equation (2.1), with $r_o = 126 \text{ \$/m}^3$, $r_{wp} = 19 \text{ \$/m}^3$, and $r_{wi} = 5 \text{ \$/m}^3$. The discount rate b is set to 0. The initial strategy (starting point) of the life-cycle optimization is a control vector with maximum injection flow rates at all control times. **Fig. 5.9** illustrates the optimization process where the blue lines represent the evolution of the objective function values for the 100 different geological realizations during the iterations while the red line is the expected value of the ensemble. The optimized expected objective function value is approximately 42.4 million \$. Due to a lack of significant change in the objective function value the optimization process was terminated after approximately 70 iterations.

⁵ Taken from Fonseca R.M., Stordal, A.S., Leeuwenburgh, O., Van den Hof, P.M.J. and Jansen, J.D. 2014. Robust ensemble-based multi-objective optimization. *Proc.in ECMOR XIV*.

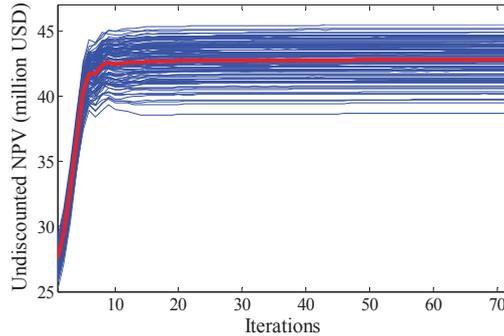


Fig. 5.9: Objective function values during the optimization procedure for the original robust EnOpt formulation. Red: average value of the ensemble. Blue: individual ensemble members.

5.5.2 Hierarchical switching optimization

The hierarchical switching optimization method is used to achieve multi-objective optimization under uncertainty as illustrated in **Fig. 5.10**. We observe a mean increase of approximately 15.2% in the secondary objective function (highly discounted NPV) compared to a marginal allowable mean decrease of 0.5% in the primary objective function. The switching optimization begins from the optimized solution achieved by the modified robust gradient formulation for life-cycle optimization. The modified formulation, i.e. equations (2.22) and (2.23) in combination with (2.14), is used for the hierarchical optimization. The results illustrates the use of ensemble-based multi-objective optimization under geological uncertainty to achieve results of practical importance. Note: we have terminated the optimization for the secondary objective function after 100 iterations which translates to 10,000 reservoir simulations for the gradient estimate and 10,107 simulations for evaluation of the updated control set, i.e. a total of 20,107 reservoir simulations.

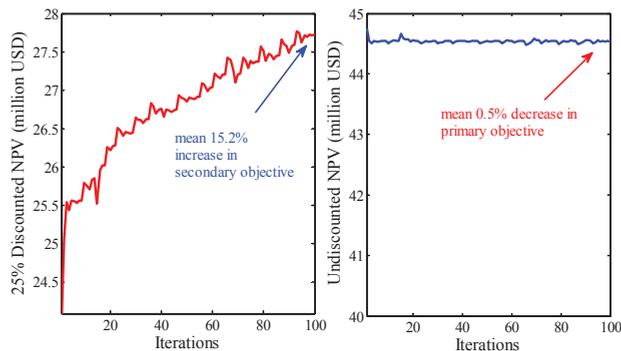


Fig. 5.10: Illustration of hierarchical switching optimization of the secondary objective function (red) and the corresponding decrease in the primary objective function (blue) starting from the optimized solution achieved with the modified robust formulation.

Fig. 5.11 depicts a comparison of the mean cumulative cash flow over time for the optimized solutions achieved by the switching algorithm (blue), life-cycle optimization (green) and reactive control (red). It is observed that after 500 days the cumulative cash flow with life cycle optimization is approximately 10.3 million \$ compared to the control strategy obtained with switching optimization which achieves a cash flow of 15.7 million \$. This 52% increase of 5 million \$ over 500 days will enable the project to achieve the break-even point faster. Similar to the results obtained in Fonseca et al. (2014) the reactive control strategy gives the best short-term performance while the switching algorithm, although inferior to reactive control, leads to an improved short-term performance compared to the optimized life-cycle strategy.

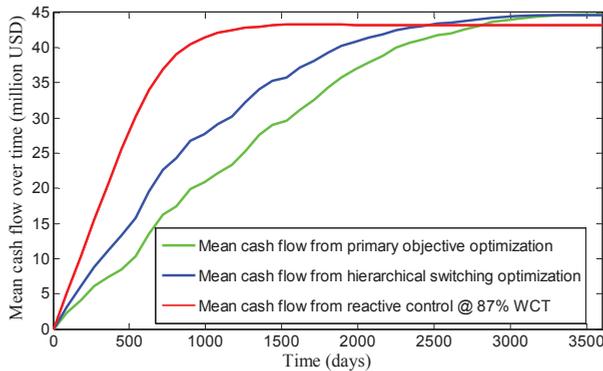


Fig. 5.11: Comparison of the mean cash flow over time for the entire ensemble of geological realization for the different optimization strategies.

5.6 Conclusions

- ✓ Compromises made to short-term targets during life cycle optimization can be partly corrected for with an ensemble-based hierarchical multi-objective optimization method.
- ✓ The EnOpt method is a good alternative to achieve practical results when the adjoint formulation is not available for hierarchical multi-objective optimization.
- ✓ In our numerical simulation examples, two hierarchical multi-objective methods showed a 14.2% improvement in the secondary objective function (NPV @ 25% discount rate) approximately constrained to the primary objective function (NPV @ 0% discount rate). The results obtained with the null-space-based optimization algorithm are similar to those resulting from the switching algorithm, although for the case investigated here, the switching algorithm was found to be computationally somewhat more efficient.

- ✓ The BFGS algorithm, used to estimate the Hessian for the null-space method, is computationally attractive compared to a finite difference method especially when dealing with large control sets, and led to good results for the case reported here.
- ✓ Hierarchical multi-objective optimization of ICV settings shows significant scope for improvement in short to medium term goals approximately constrained to life cycle targets.

6

PARETO FRONTS FOR BI-OBJECTIVE OPTIMIZATION

Conflicting objectives are frequently encountered in most real-world problems. When dealing with conflicting objectives, decision makers prefer to obtain a range of possible optimal solutions from which to choose. In theory, methods exist which can produce a range of possible solutions, some of which are “Pareto Optimal”. The application of these methods to solve bi-objective production optimization problems is increasing. Liu and Reynolds (2015) used the Normal Boundary Intersection (NBI) method to find points on the boundary of the objective function space by solving a series of constrained optimization problems using adjoint gradients. In this work, we investigate the applicability of using approximate ensemble gradients to solve a constrained optimization problem to generate a range of solutions. We compare the performance of this method to a traditional weighted sum technique for bi-objective water flooding optimization of two different synthetic reservoir models. The two objectives used in this work are, undiscounted (0%) net present value (NPV), representing long-term targets and highly discounted (25%) NPV, representing short-term operational targets. The controls are inflow control valve (ICV) settings over time for one model and water injection rate controls for the other. The effect of different starting points and the computational efficiency of the constrained optimization method are also investigated.⁶

⁶ Taken from Fonseca, R.M., Reynolds, A.C. and Jansen, J.D. 2015. Generation of a Pareto front for bi-objective water flooding Optimization using approximate ensemble gradients. Submitted to *Computational Geosciences*

6.1 Introduction

A majority of studies and applications of life-cycle water flooding optimization using a model-based approach have focused on a single objective optimization with emphasis being placed on the theoretical understanding and practical application of the optimization methodology. Life-cycle optimization essentially aims to find a strategy which optimizes long-term reservoir management targets, but life-cycle optimization is often at the expense of operationally significant short-term targets. Thus, there is a need to solve a bi-objective problem to obtain a strategy which accounts for the two objectives as the long-term perspective is usually in conflict with the short-term targets, which are decided by operational constraints, contractual obligations etc. Van Essen et al. (2011), introduced a hierarchical optimization framework to solve such a multi-objective optimization problem which provides a single optimal strategy that incorporates multiple objectives. However, decision makers usually prefer to have multiple strategies to choose from, especially when dealing with conflicting objectives. Isebor and Durlofsky (2014) applied an evolutionary algorithm to generate points along a “Pareto” front for a bi-objective water flooding problem. A major drawback of this approach was the computational effort required to obtain the points on a Pareto front. Also they did not compare the front generated with any other method used to generate Pareto fronts to check if the front obtained was Pareto optimal. Liu and Reynolds (2014) applied the normal boundary intersection method (NBI) first introduced in Das and Dennis (1998) to a bi-objective water flooding problem with and without geological uncertainty. Liu and Reynolds (2014) showed that the NBI method is computationally more efficient than the method of Isebor and Durlofsky (2014) and produces better solutions than the traditional weighted sum method. The NBI method involves solving a series of constrained optimization sub-problems. In Liu and Reynolds (2014), these constrained optimization problems were solved using an augmented Lagrangian method using an adjoint formulation to compute the gradients. Recently many studies have used EnOpt for life-cycle production optimization problems. Fonseca et al. (2014) applied EnOpt to solve a bi-objective optimization problem using the hierarchical structure proposed by Van Essen et al. (2011). Additionally there has been an increase in the number of applications of different evolutionary algorithms to solve either a bi-objective production optimization problem, Isebor and Durlofsky (2014) etc. or for history matching applications as detailed in Liu and Reynolds (2014).

In this chapter we investigate the applicability of an approximate gradient technique, EnOpt to generate points along a “Pareto” front within acceptable computational effort. A secondary objective is the application of an approximate

gradient technique to solve constrained optimization problems using the augmented Lagrangian method.

6.2 Theory

This section investigates the applicability of the use of approximate ensemble gradients to calculate points on a Pareto front for bi-objective production optimization problems. In this chapter we also use an economic objective function, a generalized formulation of which is given by (2.1). In this chapter the two objective functions are:

1. Undiscounted NPV, i.e. $b = 0.0$ (0%) in equation (2.1), representing long-term objectives
2. Highly discounted NPV, $b = 0.25$ (25%) in equation (2.1), representing short-term objectives

6.2.1 Update Rules

In this chapter we have used the steepest ascent scheme given by equation (4.1) to find an updated set of controls. The gradient is scaled by its infinity norm and we have chosen a step length to be 10% of the difference between the maximum and minimum values of the controls. We allowed for a maximum of five back-tracking steps, each time reducing the step size with a factor of one half if the objective function J decreases from one iteration to the next. If after the five back-tracking steps we still do not find an increase in J we accept the current control strategy and continue with the optimization until a convergence criteria, for e.g., maximum number of optimization iterations or total number of reservoir simulations is satisfied.

6.3 Multi-Objective Optimization

Most real world problems have multiple objectives that need to be satisfied. Usually these objectives are in conflict with each other, i.e. one must accept decreases in one objective to achieve increases in another objective. The process of optimizing systematically and simultaneously a collection of objective functions is called multi-objective optimization. In theory, there exist many methods to solve a multi-objective problem and recently there has been an increased focus on finding methods to solve multi-objective problems in the reservoir simulation community. These objectives are usually defined as long-term (life-cycle) objectives from a reservoir engineering viewpoint and short-term objectives from a production engineering/operational constraints viewpoint. Van Essen et al. (2011) showed that these two objectives may be in conflict with each other and suggested the use of a hierarchical framework for multi-

objective optimization. An alternative to hierarchical bi-objective optimization (in which the primary objective is considered more important than the secondary objective), is regular bi-objective optimization in which there is no predefined preference for one of the objectives. Isebor and Durlofsky (2014), and Liu and Reynolds (2014) have introduced methodologies to generate the ‘Pareto front’ i.e. a range of possible solutions for a decision maker for a regular bi-objective reservoir optimization problem. Isebor and Durlofsky (2014) presented their methodology using a hybrid evolutionary algorithm, PSO-MADS and reported results which were obtained with a significant computational effort. Liu and Reynolds (2014) presented a method using adjoint gradients which was shown to be computationally much more efficient. We have, in this work, used the methods introduced in Liu and Reynolds (2014) to investigate their applicability in combination with the EnOpt method.

A point in general is defined as Pareto optimal if at that point the value of one objective function cannot be increased unless the value of a second objective function is decreased or in other words a control set is Pareto optimal if there does not exist any other control set which achieves better objective function solutions. Liu and Reynolds (2014) provide details of the commonly used theoretical definitions to determine whether points are non-dominated i.e. Pareto optimal and lie on a Pareto front.

6.4 Weighted Sum Method

The life-cycle waterflooding problem is inherently a long-term optimization problem as shown in, Van Essen et al. (2011) and short-term goals are sacrificed to achieve the optimal long-term targets. A traditional technique to balance two conflicting objectives is the weighted sum method, see Marler and Arora (2004) which aims to optimize a weighted objective function that combines both objectives in a single function, according to

$$J_{ws} = w_1 J_1 + w_2 J_2, \quad (6.1)$$

where J_{ws} is the weighted sum objective function constructed from the long-term and short-term objective functions J_1 and J_2 with w_1 and w_2 as weighting factors. Liu and Reynolds (2014) among others showed that the biggest drawback of this method in finding solutions on a Pareto curve is that the solutions tend to be concentrated on one part of the curve, i.e., the solutions generated are not evenly distributed along the Pareto front. Another disadvantage is that the weighted sum method cannot obtain points on the concave part of the Pareto front, see, for example, Figure 1 in Liu and Reynolds (2014).

6.5 Adjusted Weighted Sum Method

To overcome the difficulties of the weighted sum method, Liu and Reynolds (2014) proposed an adjusted weighted sum formulation where the weights w_1 and w_2 are now replaced by equations (6.2) and (6.3)

$$\tilde{w}_1 = \frac{\frac{w_1}{J_1(\mathbf{u}_1^*) - J_1(\mathbf{u}_2^*)}}{\frac{w_1}{J_1(\mathbf{u}_1^*) - J_1(\mathbf{u}_2^*)} + \frac{w_2}{J_2(\mathbf{u}_2^*) - J_2(\mathbf{u}_1^*)}}, \quad (6.2)$$

$$\tilde{w}_2 = 1 - \tilde{w}_1, \quad (6.3)$$

and equation (6.1) is replaced by

$$\tilde{J}_{ws} = \tilde{w}_1 J_1 + \tilde{w}_2 J_2. \quad (6.4)$$

Note that $w_1 = 1$ implies $\tilde{w}_1 = 1$ and $\tilde{J}_{ws} = J_1$ so maximizing \tilde{J}_{ws} with $\tilde{w}_1 = 1$ corresponds to maximizing J_1 . Similarly, $w_1 = 0$ implies $\tilde{w}_1 = 0$ and in this case maximizing \tilde{J}_{ws} corresponds to maximizing J_2 . Liu and Reynolds (2014) found that choosing decreasing w_1 from 1 to 0.1 in increments of 0.1, computing the corresponding values of \tilde{w}_1 and \tilde{w}_2 and maximizing \tilde{J}_{ws} for each of these \tilde{w}_1, \tilde{w}_2 values tended to result in points that were well distributed along the Pareto front when maximizing \tilde{J}_{ws} and equation (6.1) with the same set of w_1 values did not generate a well-distributed Pareto front.

6.6 Normal Boundary Intersection (NBI) Method

In order to overcome the disadvantages of the weighted sum method, Das and Dennis (1998) proposed a technique, the Normal Boundary Intersection (NBI) method, to find points on the boundary of a feasible set starting from points along the “utopia line” (a line in the objective function space which is connected by the optimum solutions for the individual objective functions) by optimizing the magnitude of a unit normal to the utopia line in the objective function space. A detailed description of the Normal boundary Intersection method can be found in Das and Dennis (1998) and, for petroleum engineering applications, in Liu and Reynolds (2014). The NBI method is motivated by the fact that the Pareto front must coincide with a part of the boundary of the feasible region. The disadvantage of NBI is that boundary points may or may not be Pareto optimal, i.e., may or may not lie on the Pareto front. However once optimal design vectors $\mathbf{u}_1^*, \dots, \mathbf{u}_n^*$ are generated it is easy to check if each point is non-dominated by any other, which must be the case if $(J_1(\mathbf{u}_n^*), J_2(\mathbf{u}_n^*))$ is a point on the Pareto front, see definitions in Liu and Reynolds (2014, 2015). The following is a brief

description of the method as described in Liu and Reynolds (2014) for bi-objective water flooding optimization problems. For the two objective functions denoted by J_1 and J_2 the NBI procedure is repeated for different points along the utopia line. The general formulation for NBI is given by

$$\begin{aligned} & \max_{\mathbf{u}, t} t, \\ \text{s.t. } & \mathbf{e}(\mathbf{u}, t) = \Phi\boldsymbol{\beta} + t\mathbf{n} - \mathbf{j}(\mathbf{u}) = 0; \\ & \boldsymbol{\beta} = (\beta_1, \beta_2)^T, \beta_1 > 0, \beta_2 > 0, \beta_1 + \beta_2 = 1. \end{aligned} \quad (6.5)$$

$$\text{where } \Phi = \begin{pmatrix} J_1(\mathbf{u}_1^*) & J_1(\mathbf{u}_2^*) \\ J_2(\mathbf{u}_1^*) & J_2(\mathbf{u}_2^*) \end{pmatrix} \text{ and } \mathbf{j}(\mathbf{u}) = [J_1(\mathbf{u}), J_2(\mathbf{u})]^T \quad (6.6)$$

and where \mathbf{u}_1^* and \mathbf{u}_2^* are the optimal strategies obtained for the individual optimizations of J_1 and J_2 . The line segment which connects $\mathbf{j}(\mathbf{u}_1^*)$ and $\mathbf{j}(\mathbf{u}_2^*)$ in the objective space as $[\mathbf{j}(\mathbf{u}_1^*) \ \mathbf{j}(\mathbf{u}_2^*)][\beta_1 \ \beta_2]^T = \Phi\boldsymbol{\beta}$ is defined as the utopia line. To solve the equality constrained optimization problem as described in equation (16), Liu and Reynolds (2014) used the augmented Lagrangian method. While there exists several techniques to solve constrained optimization problems, we too have applied the augmented Lagrangian method as the main purpose of this work is to investigate the applicability and ability of approximate gradient techniques like EnOpt to generate solutions along a Pareto front. A by-product of this work is the demonstration of the applicability of an approximate gradient technique to solve constrained optimization problems. The augmented Lagrangian method (Nocedal and Wright, 2006) used to solve the different NBI sub-problems is based on the augmented Lagrangian function which is defined by

$$J_{nbi}(\mathbf{u}, t, \boldsymbol{\lambda}, \mu) = t - \boldsymbol{\lambda}^T \mathbf{e} - \frac{1}{2\mu} \mathbf{e}^T \mathbf{e}, \quad (6.7)$$

where \mathbf{e} is defined in equation (6.5), $\boldsymbol{\lambda}$ is a vector of Lagrange multipliers and μ the penalty parameter. The constraint violation is given by $\sigma_{cv} = \sqrt{\mathbf{e}^T \mathbf{e} / 2}$. Liu and Reynolds (2014), since they were using adjoint gradients, calculated the gradient of the Lagrangian function with respect to \mathbf{u} in terms of the gradients of objective functions J_1 and J_2 with respect to \mathbf{u} . In this work, since we use approximate ensemble gradients, we calculate the gradient of the Lagrangian function directly using equation (2.14). The following is a brief algorithmic description of the NBI method as implemented in our case.

Calculate the initial optimization parameters: $\boldsymbol{\beta}$, \mathbf{n} , \mathbf{u} , t , $\mathbf{e}(\mathbf{u})$, $\boldsymbol{\lambda}$, μ where $\mathbf{u} = \beta_1 * \mathbf{u}_1^* + \beta_2 * \mathbf{u}_2^*$ and t , following Liu and Reynolds (2014), is initialized as

$$t = \frac{\mathbf{n}^T (\mathbf{J}(\mathbf{u}) - \Phi\boldsymbol{\beta})}{\mathbf{n}^T \mathbf{n}}.$$

The initial penalty parameter is then given by $\mu = \mathbf{e}^T \mathbf{e} / (0.1 \cdot t)$ and the Lagrange multipliers by $\boldsymbol{\lambda} = \mathbf{e} / \mu$. Note that we scale each constraint \mathbf{e} by its own absolute value.

- While $\sigma_{cv} > 0.01$ (outer loop)
- Until stopping criteria is satisfied (inner loop)
- Maximize the Lagrangian function given by equation (6.7) until convergence is achieved. Note: $\boldsymbol{\lambda}, \mu$ are constant within the inner loop and can only change in the outer loop. Gradients are approximated using equation (2.14) in conjunction with equations (2.17) and (2.18).
End Inner Loop.
- Check criteria to update $\boldsymbol{\lambda}$ and μ using the formula given in Liu and Reynolds (2014)
- Repeat until convergence of outer loop.

6.6.1 Tracking the Pareto front using NBI

The NBI method as implemented by Liu and Reynolds (2014) chooses as a starting point a combination of the optimal control sets \mathbf{u}_1^* and \mathbf{u}_2^* depending on the weight factors chosen. Due to the non-linearity of the problem these initial points usually have objective function values that do not lie exactly on the utopia line. The NBI problem does not necessarily require the starting points to be on or close to the utopia line, so we propose to generate points on the Pareto front by starting from a point on the front which has already been obtained with different values of β_1 and β_2 . This is akin to “tracking” a front. In the results section, we discuss the advantages/disadvantages of using this method of generating solutions on a front.

6.6.2 Hierarchical Switching Method

We will compare the results obtained from the hierarchical switching optimization, described in Chapter 5 to the other methods presented above. The advantage of using a hierarchical switching method is that a user can decide the maximum allowable decrease in the primary objective value which is practically impossible to know when using the weighted sum method. However, with this hierarchical method, only a single control set is generated which may or may not be Pareto optimal since no other information is available for comparison.

6.7 Results : 5-Spot Reservoir Model

In this section, for the numerical experiments, we use the synthetic 5-spot reservoir model described in **section 4.5**. In this work, as in Liu and Reynolds (2014), the normal vector \mathbf{n} is obtained by setting the second component of \mathbf{n} to 1 and solving the following equation

$$\mathbf{n}^T [\mathbf{j}(\mathbf{u}_1^*) - \mathbf{j}(\mathbf{u}_2^*)] = 0 \quad (6.8)$$

In equation (6.8), $\mathbf{j}(\mathbf{u}_1^*) = [9.1060 \times 10^9 ; 3.3522 \times 10^9]^T$ and $\mathbf{j}(\mathbf{u}_2^*) = [8.7086 \times 10^9, 4.4759 \times 10^9]^T$. The solution of this equation gives $\mathbf{n} = [2.822 ; 1]^T$, which is the same for all the different starting points used in this work. The optimization is not dependent on the choice of \mathbf{n} . Solving multiple NBI sub-problems for different choices of weight combinations, we obtain the solutions shown in **Fig. 6.1**. The black circles are obtained for starting points based on the first step of the NBI algorithm presented previously. The objective of this initialization is to obtain a starting point on or close to the utopia line. Due to the non-linearity of the problem the objective function values achieved for this initial guess were never on the utopia line, but always slightly above the line. Using the solutions already obtained we also test the applicability of finding solutions which satisfy the constraints starting from points (control sets) that have previously satisfied the constraints. This is akin to “tracking” points along a front. The red circles in Fig. 6.1 are the points achieved when the tracking process begins from $\beta_1 = 0.1$ (point A in Fig. 6.1). We observe that for most of the weight combinations, the tracking procedure achieve solutions which dominate the solutions represented by the black circles. Since there is no preference to choose from which end the tracking begins, we also began the tracking from $\beta_1 = 0.9$ (point B in Fig. 6.1), to obtain the solutions shown by blue circles in Fig. 6.1. We observe that in this case, the tracking procedure achieves solutions that dominate the solutions from the other two initialization procedures for all the points. Additionally this tracking procedure is computationally more efficient as is discussed later. Thus, different initial guesses for a given value of β_1 can have a significant impact on the solutions achieved with the bi-objective optimization algorithm. Besides the different starting points, all other algorithmic details are exactly the same for the three different sets of points generated. The gradients are estimated with an ensemble size equal to 30 (for computational purposes) with a perturbation size equal to 0.001.

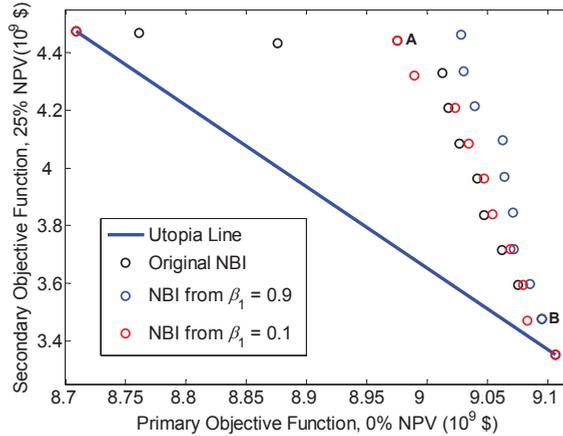


Fig. 6.1: Boundary points (blue stars) achieved using the NBI method which could constitute a Pareto front

Table 6.1 provides the objective function values for 11 different optimum points (black circles) along a boundary front. We observe that for the $\beta_1 = 0.4$ case, we obtain a 0.8% decrease in the primary, long-term objective function from its optimal value (9.1060×10^9 \$) and an approximately 22% increase in the secondary, short-term objective function. For the solution obtained with initial guesses based on front tracking procedure for $\beta_1 = 0.1$ we observe a 0.8% decrease in the primary objective to achieve a 33% increase in the secondary objective. There is only a 5% difference in the primary objective function values between the optimal strategies for the two objective functions J_1 and J_2 i.e. the first and last points in Table 6.1. Thus, for the objective functions chosen in this study, we do not expect to observe major increases in primary (long-term) objective for minor decreases in the secondary (short-term) objective, indicating that there may exist fewer redundant degrees of freedom in the short-term objective function.

Table 6.1: Objective function values of the black circles in Fig. 6.1.

β_1	β_2	Long-Term Objective ($\times 10^9$) \$	Short-Term Objective ($\times 10^9$) \$
1	0	9.1060	3.3522
0.9	0.1	9.0952	3.4757
0.8	0.2	9.0749	3.5938
0.7	0.3	9.0619	3.7161
0.6	0.4	9.0472	3.8363
0.5	0.5	9.0414	3.9645
0.4	0.6	9.0266	4.0854
0.3	0.7	9.0174	4.2075
0.2	0.8	9.0123	4.3293
0.1	0.9	8.9750	4.4430
0.08	0.92	8.8760	4.4336
0.02	0.98	8.7609	4.4692
0	1	8.7086	4.4759

Independent of the method used to generate the initial guess of a given β_1 , the approximate Pareto front generated with NBI (Fig. 6.1) shows that one can obtain a sharp increase in the secondary objective function for a very minimal decrease in the primary objective. **Fig. 6.2** is a plot of the optimization path for $\beta_1 = 0.05$ starting $\beta_1 = 0.1$ (point A in Fig. 6.1) solution as the initial guess. We observe a flat line i.e. a very minimal increase in the secondary objective function value for a relatively larger decrease in the primary objective function. The increase in the long-term objective is approximately 4% for a 0.05% decrease in the short-term objective. Thus it seems for this case that the Pareto front consists of two branches; a near horizontal one near the optimal secondary objective and a near vertical one near the optimal primary objective.

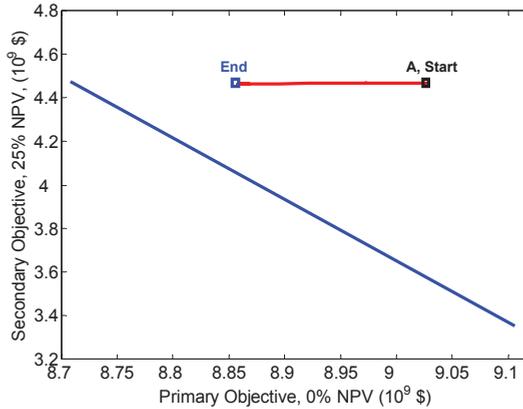


Fig. 6.2: Optimization path obtained by tracking the front using NBI for $\beta_1 = 0.05$ starting from $\beta_1 = 0.1$. Black and blue rectangles indicate the start and end points of the optimization respectively.

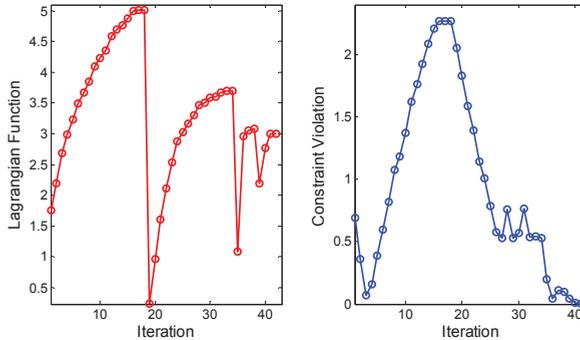


Fig. 6.3: (a) Illustration of the evolution of the Lagrangian function and (b) constraint violation at all the iterations

Fig. 6.3(a) is an illustration of the evolution of the Lagrangian function through the iteration process. The sharp drop in the value of the Lagrangian function corresponds to an update (decrease) of the penalty parameter μ in the augmented Lagrangian

method. In most of the cases we observe that we generally perform 5 outer loop iterations in which we update the penalty parameter for 3 iterations and in the other two iterations we update the vector of Lagrange multipliers λ for the remaining two iterations. **Fig. 6.3(b)**, right-side plot shows the constraint violation throughout the optimization process. Note that the constraint violation must be less than the given tolerance specified in the optimization algorithm to obtain convergence for the outer loop of the augmented Lagrangian algorithm. When the inner loop converges and the constraint violation is sufficiently small the algorithm converges. Thus it is possible that there are multiple points at which the constraint violation is satisfied, however there is only one point at which both the inner and outer loop's stopping criteria are satisfied.

6.7.1 Control Comparison

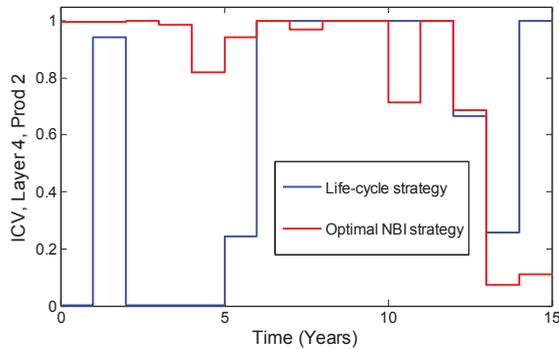


Fig. 6.4: Comparison of the controls (ICV settings) for layer 4 (highest permeability) for producer 2 over time for the optimal life-cycle strategy (blue) and the optimal NBI strategy (red) for $\beta_1 = 0.1$ using the front tracking procedure

Fig. 6.4 is a comparison of the optimal control set for two different strategies for the highest permeability layer in producer 2. The blue line is the life-cycle strategy i.e. the end point of the utopia line while the red line is the NBI strategy (blue circles in Fig. 6.1) obtained for weight combinations of $\beta_1 = 0.1$ and $\beta_2 = 0.9$ i.e. the strategy that achieved a 33% increase in the short-term objective for a 0.8% decrease in the long-term objective. From Fig. 6.4 we observe that significantly different strategies can be achieved by performing bi-objective optimization. For the red curve, the ICV setting is almost fully open for the first 10 years with lower setting values towards the end of the producing time period, which is in line with the emphasis on increasing the short-term increase in NPV. For the optimal life-cycle strategy, the same ICV is almost closed for four or the first five years and then is fully open through most of the remaining producing life, in order to virtually maintain the goal of life-cycle optimization. The optimized control settings for other valves is similar to the trend shown in Fig. 6.4. **Fig.**

6.5 shows the saturation distribution in layer 4 after 4 years of production for the different optimal strategies whose controls are compared in Fig. 6.4. We see that the optimal life-cycle strategy, being less aggressive, sweeps a much smaller area with less water being injected, while the optimal NBI strategy i.e. the one for $\beta_1 = 0.1$ using the front tracking procedure, is more aggressive, i.e. more water is injected and more oil is displaced and produced.

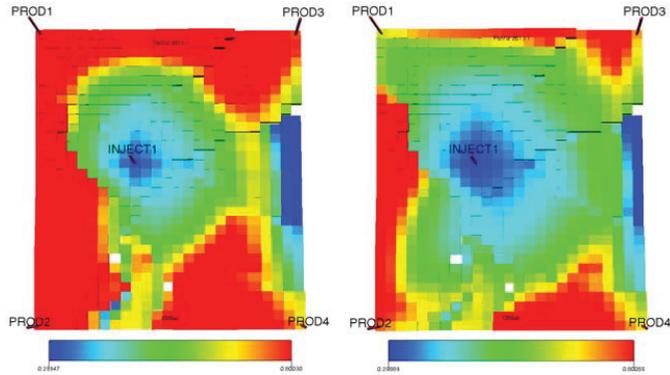


Fig. 6.5: Saturation distribution in layer 4 after 4 years of the production for a) (left), optimal life-cycle strategy and b) (right), optimal NBI strategy for the controls shown in Fig. 6.4.

6.7.2 Comparison of Weighted Sum Techniques

Liu and Reynolds (2014) showed cases where the adjusted weighted sum method produces a significantly better spread of solutions compared to the traditional weighted sum technique. Table 6.2 and Table 6.3 provides the solutions for the various weight combinations used where we observe, as reported in Liu and Reynolds (2014), that the adjusted weighted sum technique provides a better spread of solutions, and in particular gives a better representation of the front near the optimal long-term NPV which is the most important part of the front.

Table 6.2 : Solutions for different weight combination using the weighted sum method

w_1	w_2	Long-Term Objective ($\times 10^9$) \$	Short-Term Objective ($\times 10^9$) \$
1	0	9.1060	3.3522
0.9	0.1	9.0851	3.6872
0.8	0.2	9.0419	3.9719
0.7	0.3	8.9720	4.1763
0.6	0.4	8.9156	4.3591
0.5	0.5	8.8696	4.4106
0.4	0.6	8.7873	4.4313
0.3	0.7	8.7538	4.4476
0.2	0.8	8.7403	4.4448
0.1	0.9	8.7243	4.4492
0	1	8.7086	4.4759

Fig. 6.6 provides a visual comparison of the solutions obtained with the two different methods. Note: The stopping criterion used to achieve this set of points is exactly the same as the stopping criterion used for the inner loop in the augmented Lagrangian-based NBI method albeit the objective functions are different.

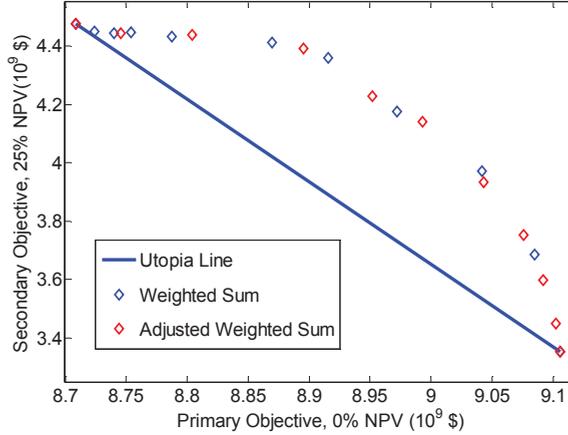


Fig. 6.6: Comparison of the spread in points along the Pareto front for the two variants of the weighted sum method.

Table 6.3: Solutions for different weight combination using the adjusted weighted sum method

w_1	w_2	Long-Term Objective ($\times 10^9$) \$	Short-Term Objective ($\times 10^9$) \$
1	0	9.1060	3.3522
0.9	0.1	9.1020	3.4502
0.8	0.2	9.0918	3.5996
0.7	0.3	9.0758	3.7522
0.6	0.4	9.0433	3.9326
0.5	0.5	8.9927	4.1400
0.4	0.6	8.9515	4.2268
0.3	0.7	8.8956	4.3905
0.2	0.8	8.8040	4.4384
0.1	0.9	8.7454	4.4423
0	1	8.7086	4.4759

6.7.3 Comparison of Weighted Sum and NBI

Fig. 6.7 is a comparison of the solutions achieved from the adjusted weighted sum method and the best results achieved with the NBI method. The results here are very interesting, for $w_1=0.9$ the solutions obtained with either method do not dominate each other while for $w_1=0.8$ and $w_1=0.7$ we observe that the adjusted weighted sum method achieved solutions which slightly dominate the solutions obtained with NBI, however for the other weight combinations the solutions obtained with NBI dominate. It is difficult to know why this behavior is observed and could be either case dependent or gradient quality dependent. However, Liu and Reynolds (2015) also find that the NBI

method generally gives a better representation of the front than is obtained with the weighted sum method. **Fig. 6.8** is a comparison of the optimization path for the different methods with the weight combination, $w_1 = 0.7$ and $w_2 = 0.3$. The original NBI and the weighted sum have the same starting point however they have very different paths. The adjusting of the weights in the adjusted weighted sum method leads to a significantly different starting point for the optimization. All the optimization results shown here are influenced not only by the gradient quality, but for the NBI method, also by the choices of the initial penalty parameter μ and Lagrange multipliers λ . Using a larger ensemble size for the gradient estimate (Fonseca et al. 2015) could lead to smoother optimization paths to solve the individual sub-problems and possibly better solution points, however for computational reasons this has not been investigated.

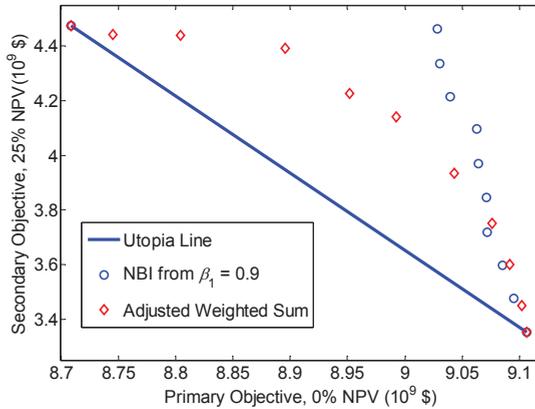


Fig. 6.7: Comparison of the points achieved from the adjusted weighted sum method and the NBI-based tracking method.

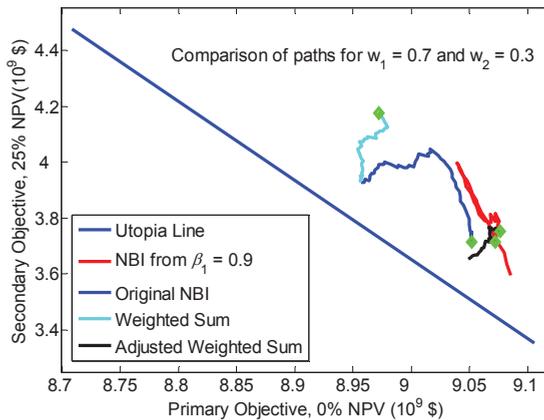


Fig. 6.8: Comparison of the optimization paths for the different methods described above. The green diamonds are the end points i.e. optimum solutions obtained by each of the methods.

Fig. 6.9 is an illustration of the total number of simulations taken for each of the methods including the two initial optimization runs to obtain the utopia line. The original NBI method was computationally most expensive with approx. 28000 total simulations while for the NBI tracking method about half the number of simulations required for the original NBI method were needed to achieve better solutions. Both the weighted sum variants were computationally much more efficient similar to the results shown in Liu and Reynolds (2014).

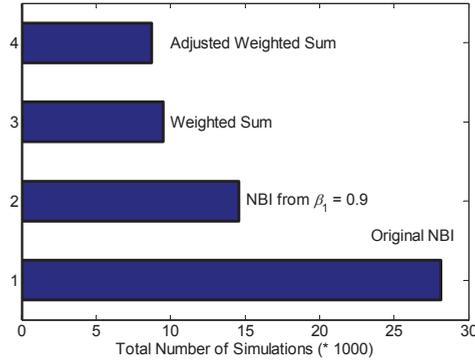


Fig. 6.9 : Comparison of the total number of simulations taken to generate the Pareto curve for the different methods.

6.7.4 Hierarchical Switching Optimization Method

The switching method optimizes the objectives alternately, while staying within a maximum allowable decrease ε in the primary objective. The choice of ε is user dependent. Thus, the user has to a-priori decide the maximum allowable acceptable decrease in the optimal primary objective function value. **Fig. 6.10** plots the optimization path where a maximum decrease of 0.3% in the primary objective is allowed (red curve). We see that we achieve approximately a 10% increase in the secondary objective. The values obtained are similar to using a weight combination of $w_1 = 0.7$ and $w_2 = 0.3$ for either the NBI method or the adjusted weighted sum method. However the solution is a non-dominated point when compared to all solutions obtained with the NBI and adjusted weighted sum methods. If the optimization is repeated for a 1% allowable decrease in the primary objective we observe that we achieve a 20 % increase in the secondary objective function (black dotted curve). This solution however is dominated by the solutions from the other two methods. The hierarchical method only provides a single strategy which may or not be Pareto optimal. A tracking procedure like implemented for the NBI method could be used with this method to generate a front. Alternatively we can use the primary objective

function as a constraint while optimizing a secondary objective, i.e., a lexicographic approach; for details see Liu and Reynolds (2015).

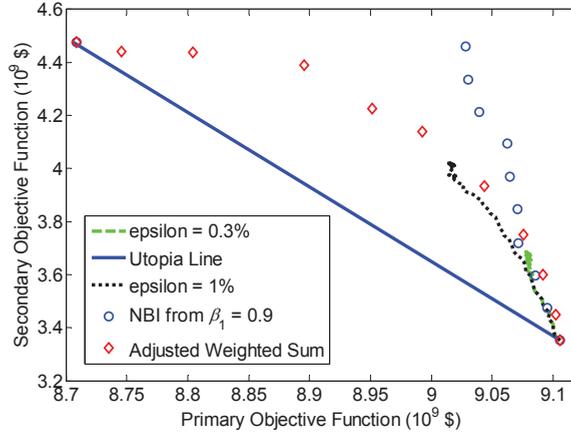


Fig. 6.10: Paths obtained from hierarchical switching optimization for different maximum allowable decreases in the primary objective function.

6.8 Results : Egg Model

In this section we use the egg model detailed in **section 3.4**. In this example we optimize controls in terms of injection rates for a relatively larger model. The normal vector \mathbf{n} is obtained, similar to the example reported above, by setting the second component of \mathbf{n} to 1 and solving equation (6.8) which gives $\mathbf{n} = [1.6329 ; 1]^T$. Where $\mathbf{j}(\mathbf{u}_1^*) = [4.7035 \times 10^7, 2.3004 \times 10^7]^T$ and $\mathbf{j}(\mathbf{u}_2^*) = [4.0269 \times 10^7, 3.4053 \times 10^7]^T$. For different linearly varying combinations of β we solve multiple NBI sub-problems to find solutions of the two objective functions, as shown in **Fig. 6.11**, which satisfy the stopping criteria of the augmented Lagrangian function and the constraint violation. The black circles are obtained for starting points which aim to start on the utopia line. Again due to the non-linearity of the problem, the objective function values of the starting points were never on the utopia line. The spread in the points found using NBI was more continuous compared to the solutions achieved in the previous example. Again instead of solving the sub-problems from a starting point close to or on the utopia line we aim to “track” points along a front. The blue circles in **Fig. 6.11** are the points achieved when the tracking process begins from $\beta_1 = 0$. The results illustrated in **Fig. 6.11** seem to suggest that there exists different fronts in the objective function space. We observe that till $\beta_1 = 0.7$ the points seem to lie on a line with a certain slope for both the original NBI method as well as the NBI tracking method. From $\beta_1 = 0.6$ onwards the points seem to align themselves along a line with a completely different slope for both the methods. The points obtained with NBI tracking seems to always

find solutions that dominate the solutions obtained by the original NBI method. Thus, similar to the previous example different starting points of the optimization have significant impact on the solutions achieved. Besides the different starting points, all other algorithmic details are exactly the same for the three different sets of points generated. The gradients are estimated with an ensemble size equal to 30 (for computational purposes) with a perturbation size equal to 0.01.

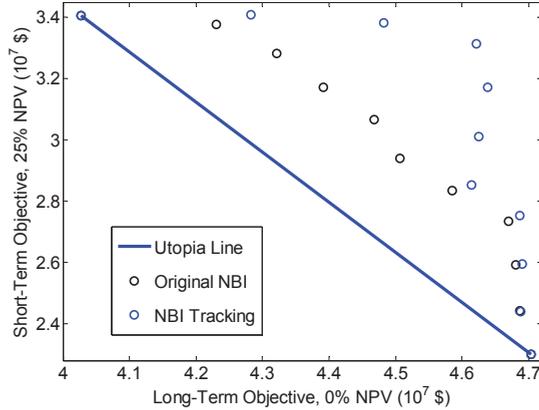


Fig. 6.11: Boundary points (blue circles) achieved using the NBI method with tracking which could constitute a (local) Pareto front.

With only a 0.7% decrease in the primary, long-term objective function, we have found a solution which achieves an approximately 19% increase in the secondary, short-term objective function for the black open circles i.e. the original NBI method for $\beta_1=0.7$. For the same weight combination the NBI tracking method finds a solution for which a 0.3% decrease in the long-term objective leads to a 20% increase in short-term gains. Additionally, for the red circles, i.e. NBI with tracking, we observe that for a 1.3% decrease in the primary objective, we can achieve an even more significant increase of 38% in the secondary objective, i.e. short-term gains. This last result corresponds to the $\beta_1=0.4$ solution.

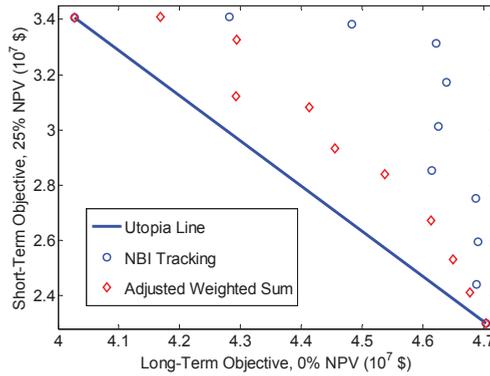


Fig. 6.12: Comparison of solutions achieved by the adjusted weighted sum method (red diamonds) and the NBI tracking method (blue circles)

The results from the previous example illustrated that the adjusted weighted sum method produced a much better spread in solution points compared to the weighted sum method. Thus for this example we compare the NBI solutions with solutions from the adjusted weighted sum technique. We observe, as shown in **Fig. 6.12**, that the solutions achieved by the NBI tracking method dominate the solutions from the adjusted weighted sum method (red diamonds). The solutions from the original NBI method also dominate the solutions from the adjusted weighted sum method. A comparison of the plot of remaining oil saturation for the top layer after 3 years of production illustrates the difference in the strategies; see **Fig. 6.13**. The optimal long-term strategy is less aggressive as significantly less area is swept by injected water while the NBI tracking solution for $\beta_1 = 0.4$, i.e. a 38% increase in short-term gains for a 1.3% decrease in long-term gains, is a more aggressive strategy as larger areas of the reservoir have been swept.

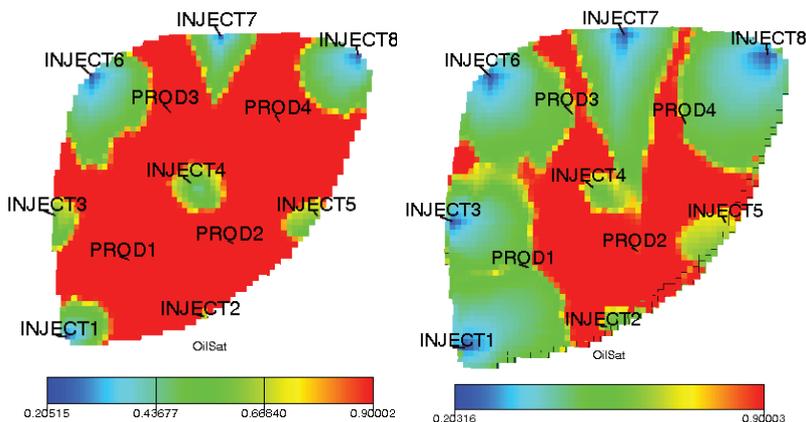


Fig. 6.13: Comparison of oil saturation distributions for the top layer after 3 years of production for a) optimal long-term solution (left side) and b) NBI tracking solution for $\beta_1 = 0.4$ (right side)

The control strategies which resulted in the saturation plots shown in Fig. 6.13 are illustrated in Fig. 6.14. The control settings have only been displayed for the first 3 years to highlight the differences between the strategies corresponding to the saturation plots shown above. We observe that the optimal NBI strategy injects much more water compared to the optimal long-term strategy. A similar trend in the control strategies is seen in the other wells.

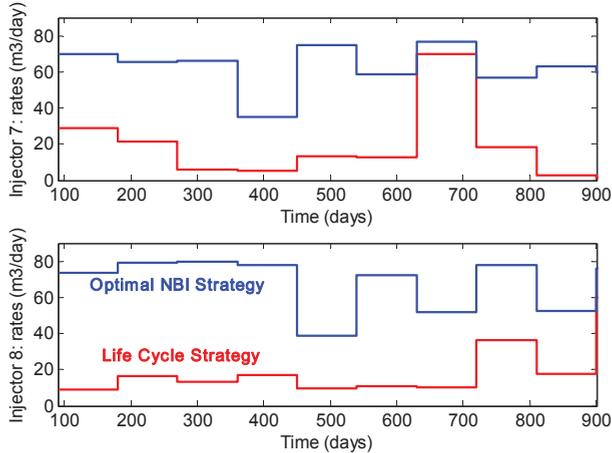


Fig. 6.14: Comparison of injection rate controls for injector 7 (top) and injector 8 (bottom) for the first 3 years of production resulting from the different strategies, life-cycle strategy (red line) and NBI strategy (blue line)

Fig. 6.15 is a comparison of the computational efficiency of the different methods. The original NBI method requires the highest computational effort similar to the results reported for the previous example. We needed about 21,000 simulations to achieve the 11 points for the original NBI method while we needed approximately 14,000 simulations when using the NBI to track the boundary front. The adjusted weighted sum method required less than 8,000 simulations to find the 11 points thus it is computationally most efficient though the solutions achieved are far from Pareto optimal compared to the solutions achieved by the NBI method.

6.8.1 Discussion

The differences between the results of the adjusted weighted sum and the traditional weighted sum method for this example are far from significant because the adjustment of the weights does not lead to very different weight combinations, i.e., scaling of the problem is not as important for this problem as it was for the 5 spot ICV problem and the problems investigated in Liu and Reynolds (2014). For this example, the difference between the optimal primary long-term objective function values is 14% which is

much higher than in the previous example and 48% for the short-term objective function values.

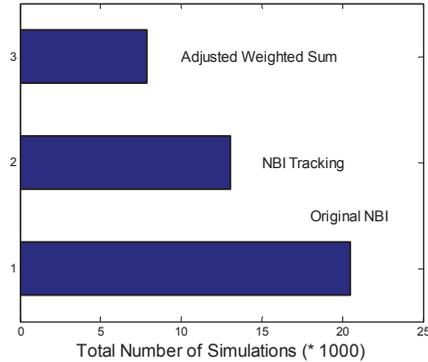


Fig. 6.15: Comparison of the computational effort required to achieve solutions using the different methods.

6.9 Conclusions

- ✓ Approximate gradient techniques like EnOpt can be used to generate solutions which may lie on a pareto front for a bi-objective optimization problem within acceptable computational effort.
- ✓ Tracking the Pareto front using NBI is a computationally more efficient method and produces better solutions for the decision maker to choose from compared to the original NBI form. Different starting point have a significant impact on the optimal solutions achieved.
- ✓ The adjusted weighted sum produces a more even distribution of solutions and is marginally computationally more efficient compared to the traditional weighted sum technique for this case.
- ✓ For some weight combinations the NBI method produces solutions which dominate solutions obtained by the weighted sum variants and vice versa.
- ✓ A hierarchical switching method provides a single solution which satisfies the maximum allowable decrease in objective function value however the solution is always dominated by the solutions obtained by the other methods.

7

ROBUST BI-OBJECTIVE OPTIMIZATION OF ON-OFF INFLOW CONTROL DEVICES: A REALISTIC MODEL

In this chapter we consider robust ensemble-based (EnOpt) multi-objective production optimization of on-off inflow control devices (ICDs) for a sector model inspired from a real-field case⁷. The use of on-off valves as optimization variables leads to a discrete control problem. We propose a re-parameterization of such discrete controls in terms of switching times, i.e. we optimize the time at which a particular valve is either open or closed. This transforms the discrete control problem into a continuous control problem which can be efficiently handled with the EnOpt method. Additionally this leads to a significant reduction in the number of controls which is expected to be beneficial for gradient quality when using approximate gradients. We consider an ensemble of sector models where the uncertainty is described by different permeability, porosity, net-to-gross and initial water saturation fields. The controls are the ICD settings over time in the three horizontal injection wells, with approximately 15 ICDs per well. Different optimized strategies resulting from different initial strategies were compared. We achieved a mean 4.2% increase in expected NPV at a 10% discount rate compared to a traditional pressure maintenance strategy. Next, we perform a sequential bi-objective optimization, and achieved an increase of 9.2% in the secondary objective (25% discounted NPV to emphasize short-term production gains) for a minimal decrease of 1% in the primary objective (0% discounted NPV to

⁷ Taken from Fonseca, R.M., Leeuwenburgh, O., Della Rossa, E., Van den Hof, P.M.J. and Jansen, J.D. 2015. Ensemble based multi-objective optimization of on-off control devices under geological uncertainty. Under review at *SPE Reservoir Engineering and Evaluation*.

emphasize long-term recovery gains), as averaged over the 100 geological realizations. The workflow is repeated for alternative numbers of ICDs showing that having fewer control options lowers the expected value for this particular case. The results demonstrate that ensemble-based optimization workflows are able to produce improved robust recovery strategies for realistic field sector models against acceptable computational cost.

7.1 Introduction

In the recent past there has been an increased focus on the application of different model-based optimization techniques for optimal control to achieve improved reservoir management strategies. Most of these studies have used relatively simple models, while a limited number of studies such as Bailey et al. (2005), Sarma et al. (2008), Alhutali et al. (2009), Chaudhri et al. (2010), Forouzanfar et al. (2013), Van Essen et al. (2010) and Raniolo et al. (2013) amongst others have used realistic field scale or sector models. Most of the studies that use realistic real field models concerned single-objective optimization on a single geological realization, with the exception of Alhutali et al. (2009) and Raniolo et al. (2013) who performed single-objective optimization using an ensemble of geological realizations.

Chen et al. (2009) reported a successful application of robust EnOpt using the 1:1 ratio for the SPE Brugge benchmark case. Raniolo et al. (2013) and Li et al. (2012) have investigated the applicability of approximate gradient techniques for life-cycle robust water flooding optimization. Yang et al. (2011) applied the robust optimization principle to a Steam-Assisted Gravity Drainage (SAGD) application.

Recently, Chen et al. (2012) presented a adjoint based robust multi-objective optimization scheme while Yasari et al. (2013), Pajonk et al. (2011), Schulze-Riegert et al. (2011) and Awotunde and Sibaweih (2011) have investigated the applicability of robust multi-objective optimization using evolutionary algorithms for well control and well placement optimization with objectives varying from economic criteria to Voidage replacement ratio and cumulative production volumes.

In this chapter we optimize the settings of on-off Inflow Control Devices [ICDs, sometimes also referred to as Inflow Control Valves (ICVs)] which have discrete settings, i.e. either 0 or 1. However, EnOpt, like other gradient-based techniques, cannot efficiently handle discrete control problems. Thus we use a re-parameterization of the controls into switching times as will be discussed in detail below. Using this technique, we investigate the applicability of an ensemble-based robust hierarchical multi-objective optimization framework to optimize an ensemble of sector models inspired from a real field case. The uncertainty in the ensemble of sector models is

characterized by differing permeability, porosity, net to gross and initial water saturation fields.

7.2 Theory

In this work we have chosen to use an approximate gradient method (EnOpt), detailed in chapter 2, for the optimization instead of a derivative-free technique because the computational costs for derivative free methods are usually higher. For our example, which incorporates geological uncertainty in the form of 100 realistic reservoir models, a derivative-free technique would be computationally extremely challenging. Additionally, in our limited experience, approximate gradient techniques usually have fewer tuning parameters compared to derivative-free methods. Most real world problems have multiple objectives that need to be satisfied. Usually these objectives are in conflict with each other, i.e. one must accept decreases in one objective to achieve increases in another objective. Thus for this field case inspired application we use the hierarchical switching method with EnOpt for the optimization, details on the implementation are provided in **Chapter 5**.

7.3 Control Parameterization

Recent advances in technology and the need for improved controllability of the oil recovery process have led to the use of inflow control devices (ICDs), also known as inflow control valves (ICVs). While there exists many variants of such devices, in our study we have used ICDs that can be individually activated using electric line or coiled tubing through relatively simple rigless intervention techniques. This type of ICD is currently commercially available with up to 16 ICDs per well, and is used in field applications. In the present study we optimize the settings of these ICDs over the producing life of a field. ICDs can have settings which vary continuously between 0 and 1 or could be restricted to settings of either 0 or 1, i.e. either fully open or fully closed. In our study the ICDs belong to the latter class, i.e. they have settings of either 0 or 1. Thus the optimization problem is now discrete in nature. To solve such discrete problems, integer programming techniques are usually used; see e.g. Isebor et al. (2014). Gradient-based methods like EnOpt, which have been successfully applied to problems with continuous variables, are not suitable for discrete control problems. In order to use EnOpt to solve this particular optimization problem, a parameterization of the controls into continuous variables is necessary. Sudaryanto and Yortsos (2000) suggested the use of switching times as control variables for a production optimization exercise to study the behavior of bang-bang (i.e. on-off) controls. Zandvliet et al. (2007) investigated the theoretical aspects of bang-bang control problems and provided

a list of references from different engineering applications where switching times have been used as controls. However, they did not use switching times as the controls; rather they used continuous control variables in combination with a cut-off criterion to mimic the discrete controls. We have also investigated this approach; however we did not achieve solutions which were better than a switching times approach. Hasan and Foss (2013) recently used switching times as controls for adjoint-based water flooding optimization with the halving-time-interval method to update the controls, and Namdar Zanganeh (2014) used a switching approach to optimize a surfactant (foam) injection application. In this paper we use the switching-time-interval-based parameterization of the controls based on the Switching Time Optimization (STO) method provided in Sudaryanto and Yortsos (2000). While Sudaryanto and Yortsos (2000) use explicit times to define a switch, we use time intervals as controls. There exists two distinct advantages of using such a parameterization

- a) It transforms a discrete control problem into a continuous control problem which can be efficiently handled by EnOpt.
- b) It leads to a possibly significant reduction in the number of control variables. This could be particularly important when using stochastic approximate gradient-based techniques like EnOpt because the gradient quality may deteriorate for increasing numbers of optimization variables.

For example, in this case, if we choose fixed control time steps of 1 year and use 48 ICD settings to be optimized at each control time step, using amplitude-based controls for a 20-year period would result in a control vector u containing $48 \times 20 = 960$ elements. In the switching times approach the user must pre-define the number of allowable switches during the simulation time for each ICD. In our case we use five switching times for each ICD over the producing life of the reservoir, thus leading to $48 \times 5 = 240$ controls, i.e. a factor 4 reduction in the number of controls. Note that we work with time intervals to define the times at which a particular ICD must be either switched on or off depending on its previous setting. We do not explicitly find the time a control is switched; rather we find the interval after which a control setting can be changed. Also note that five switching times is a maximum limit and it is possible to achieve an optimized strategy with fewer switches. In case the final switching time interval exceeds the end producing time, the interval is simulated only until the end time.

7.4 Results: Field Case Inspired Application

The reservoir model used in this study is a sector model inspired on a real field case. The reservoir formation is unconsolidated sandstone at approximately 4000 ft. depth

with a net thickness ranging from 30 to 45 ft. and is not significantly faulted. A total of 50,000 active grid blocks are used in this sector model. Model dimensions cannot be disclosed for confidentiality reasons. The reservoir rock is of good quality with porosities ranging from 20-35 % and net-to-gross ratios of approximately 50-90%. The permeability distribution is not very heterogeneous with values ranging from 100-700 mD, usually around 350 mD. Permeability modeling has been carried out following industry standard practices, based on facies and petro-physical modeling . Additionally, an integration of core data coming from some of the wells with computer-processed interpretation (CPI) logs has been incorporated. A porosity-permeability correlation has been identified and grids have been populated using a standard geostatistical algorithm. Following this, well test data has been integrated and the initial ensemble of permeabilities have been conditioned to historical data to get an updated ensemble which is used in this study. The reservoir is operated using a line drive strategy with horizontal wells having lengths of 4000-10000 ft with ICDs installed in some of the injectors. The field has unconventional reservoir and fluid properties compared to other fields in the vicinity. In particular, the field temperature is lower and the oil is heavier and more viscous, compared to neighboring fields.

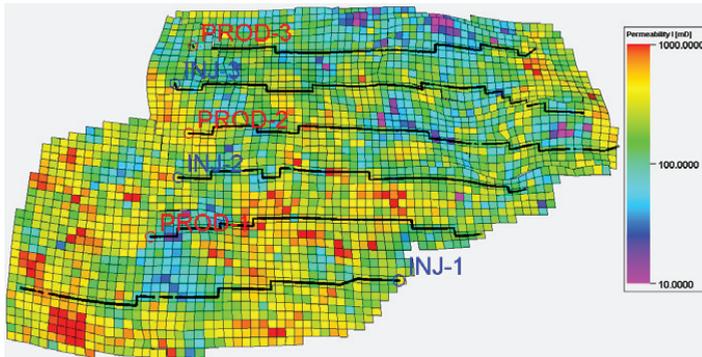


Fig. 7.1: Permeability field for layer 14 of realization number 65

Fig. 7.1 is an illustration of the permeability field of layer 14 from realization number 65 out of an ensemble of 100 realizations. The sector model shown below is produced using a line drive strategy with 3 injectors and 3 producers. The injectors are equipped with ICDs which can be mechanically operated using a coiled tubing unit and can only be either ‘open’ or ‘closed’, i.e. have values of either 0 or 1. Injector 1 has 15 ICDs along its length while injectors 2 and 3 have 16 and 17 ICDs respectively. The settings of these ICDs are the control variables for this optimization study. The wells are modeled as multi-segment wells in a commercial fully-implicit finite difference black oil simulator (Eclipse, 2011). Because the ICDs cannot be modeled to be fully closed, i.e. 0, due to numerical limitations, in this exercise the valve settings have a minimum

value of $1e-4$. To effectively capture the effect of geological uncertainty within the optimization framework we use a set of 100 different geological (sector) models. The realizations vary in terms of their permeability fields, porosity fields, varying net-to-gross ratios and initial water saturations.

The relative homogeneity in the permeability fields, coupled with the specific properties of the oil, suggest that for this model the scope for optimization lies in the reduction of the volumes of water injected and produced. **Fig. 7.2** illustrates the oil saturation of layer 14 from realization 65 after 20 years of production. The field water cut for a 20-year horizon is approximately 89%, this increase in the volumes of water produced and injected makes the problem interesting for optimization. Hence we have chosen an optimization time horizon of 20 years.

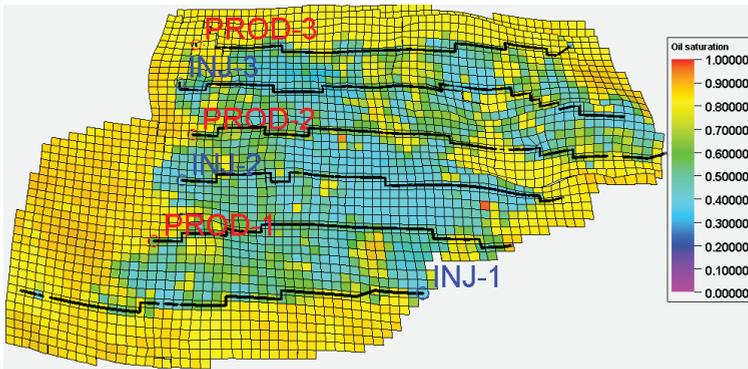


Fig. 7.2: Oil Saturation of layer 14 for realization 65 after a time of 20 years indicating a high water production.

7.4.1 Life Cycle Optimization

In this model the injectors and producers are operated on bottom hole pressure constraints while the ICD settings are allowed to vary over the producing life (20 years) of the reservoir. Based on engineering judgment we allow for only five switching times per ICD throughout the 20 years, and because we have 48 ICDs per control time the control vector \mathbf{u} has $N = 5 \times 48 = 240$ elements. An optimal life-cycle strategy of ICD settings in the injection wells is obtained by optimizing NPV, as described in equation (2.1) with $r_o = 90$ \$/bbl, $r_{wp} = 8$ \$/bbl, and $r_{wi} = 5$ \$/bbl. The discount rate b is chosen to be either of 0 or 10%. The initial strategy (starting point) of the life-cycle optimization is a control vector with all switching time intervals equal to 0; i.e. there are no switches allowed which implies that all the ICDs are fully open at all control times. **Fig. 7.3** illustrates the optimization process where the blue lines represent the evolution of the objective function values for the 100 different geological realizations during the iterations while the red line is the expected value of the

ensemble. The optimized expected objective function value is approximately 695 million \$. Due to a lack of significant further change in the objective function value the optimization process was terminated after 30 iterations

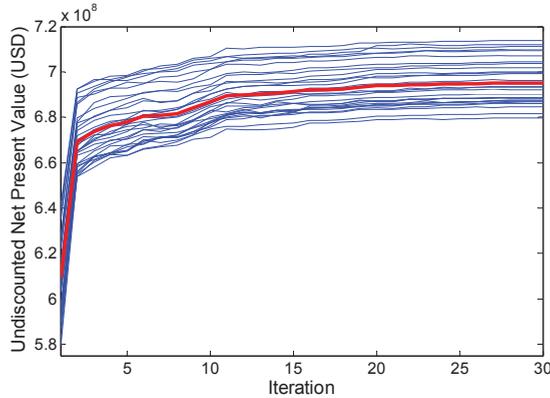


Fig. 7.3: Mean objective (undiscounted NPV) function value for the 20 year simulation time period. The red line is the mean value of the objective value while the blue lines are the objective function values for the individual ensemble members through the iteration process.

We observe, for an undiscounted NPV, an approximately 12% increase in mean objective function value at the end of the optimization compared to the starting point of the optimization (i.e. compared to a ‘do-nothing’ strategy). While this result is encouraging, an analysis of the cumulative oil and water rates shows an approximately 10% mean decrease in cumulative oil production for a corresponding 47.5% mean decrease in cumulative water injected and 70% mean decrease in cumulative water produced. While the volumes of water injected and produced have reduced significantly, this corresponds to a large decrease in cumulative oil production which may not be the most attractive strategy.

7.4.2 Effect of Discount Factor

In reality a discounted economic objective is traditionally used in the analysis of any project. Thus we also use discounted NPV as the objective function with a 10% discount factor. First, as in Fig. 7.3 we start the optimization from the “do-nothing” strategy, i.e. a strategy where all the ICDs are open throughout the life time of the field. **Fig. 7.4(a)** is an illustration of the optimization with 10% discounted objective function. A comparison of this result to a simulator-handled pressure maintenance strategy shows a 2.8% mean increase in NPV. The mean cumulative oil volumes obtained with this strategy are higher than with the pressure maintenance strategy. However, the volumes of water injected and produced are also higher. Since the

starting point of the optimization is rather aggressive (in terms of water injection volumes) we find an optimized strategy which reflects the same behavior. This strategy is hereafter referred to as ‘Opt. Strategy 1’.

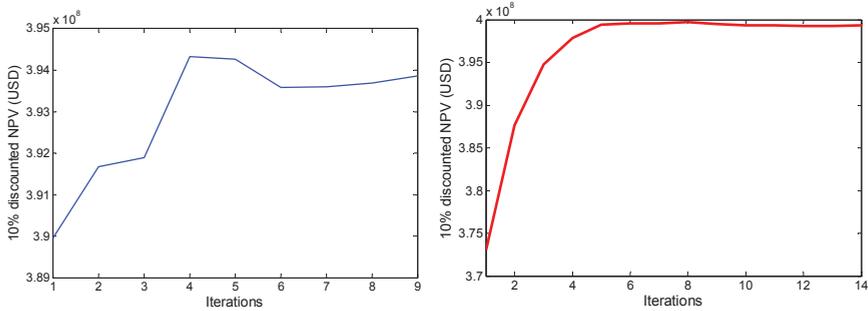


Fig. 7.4: (a) 10% discounted NPV, optimization started from ‘do-nothing’ strategy. (b) 10% discounted NPV, optimization started from end of Fig. 7.3, i.e. the optimized strategy. Notice the difference in the scale of the Y-axis between the two plots.

Following this result we perform an optimization experiment from a different initial strategy, i.e. a different point in the control space. The initial strategy is the optimized strategy from the undiscounted life-cycle result, shown in Fig. 7.3. We observe that starting from a different initial strategy has led us to a solution with a higher NPV, see **Fig. 7.4(b)**, compared to Fig. 7.4(a). In addition, compared to the pressure maintenance strategy, this optimized strategy (see Fig. 7.4(b)) hereafter referred to as ‘Opt. Strategy 2’, achieves a mean increase of 4.2%. Thus the optimization is fairly sensitive to the initial starting point of the optimization. Opt. Strategy 2 starts from a significantly less aggressive strategy (in terms of volumes of water injected) compared to Opt. Strategy 1 and thus achieves a lower mean cumulative oil production but also injects much lower volumes of water. However, irrespective of the starting point, the optimized strategy always achieves better solutions in terms of NPV compared to the simulator-handled pressure maintenance strategy. If we perform an optimization exercise while requiring the simulator to enforce pressure maintenance we observe that the optimization is not successful, because of interference of the simulator-handled control with the optimization, as can be seen in **Fig. 7.5**.

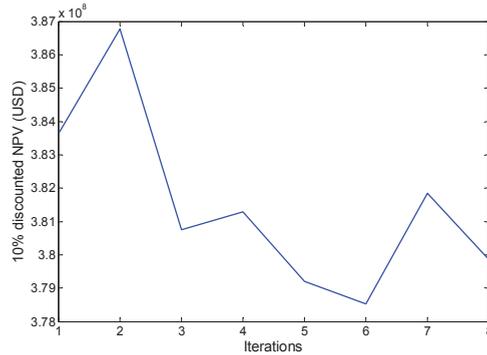


Fig. 7.5: 10% discounted NPV with pressure maintenance handled by the simulator which has a negative impact on the optimization procedure.

Property	Pressure Maintenance	Opt. Strategy 1	Opt. Strategy 2
NPV@ 10%	3.8363e8	3.9432e8 (+2.8%)	3.9971e8 (+4.2%)
Cum. Oil (STB)	1.3712e5	1.4130e5 (+3.0%)	1.3604e5 (-0.8%)
Cum. Water Prod (STB)	3.2245e5	3.3985e5 (+5.4%)	2.8454e5 (-13%)
Cum Water Inj. (STB)	4.6053e5	4.8332e5 (+5.0%)	4.2054e5 (-9.5%)

Table 7.1 highlights the key differences between the two different optimized strategies when using a discounted objective function. We observe that while Opt. Strategy 2 achieves a higher NPV compared to Opt. Strategy 1, the amount of cumulative oil produced is marginally lower than with the simulator-handled pressure maintenance. However, there is a significant reduction in cumulative volumes of water needed for injection as well as water produced. On the other hand, Opt. Strategy 1 achieves not only a higher NPV but also a 3% mean increase in cumulative oil produced at the cost of a higher cumulative water injection volume and thus a higher volume of water produced. The difference for the two strategies stems from the different initial starting points. While it is difficult to conclude in favour of either strategy, decision makers now have a choice based on their objectives.

7.4.3 Reactive Control

The economic water cut based on the economic parameters used to calculate the objective function is 91%, while the current average water cut is 89%. Thus we have not yet reached a reactive control strategy. (The reactive control strategy corresponds to the ‘do-nothing’ strategy, the NPV of which is the starting point of the optimization.) Thus, in general, for an undiscounted NPV the optimized strategy has an approximately 12% gain over the reactive strategy.

7.4.4 Optimized Control Strategy

Fig. 7.6 is an illustration of the comparison between an optimized strategy (red) of 16 ICDs installed in injector 2 with the reactive strategy. Recall that we defined a maximum of five allowable switches for each ICD throughout the 20-year optimization horizon. We observe that for the optimized strategy, for most cases, the ICDs do not need 5 switches; rather the optimized strategy consists of mainly one or two switches per ICD. Many life-cycle optimization studies have obtained optimal control sets that are non-smooth in nature, i.e. they display frequent adjustments to the control settings which is practically undesirable and probably not feasible to implement. However an optimized strategy as illustrated in Fig. 7.6 would be more appealing to implement as an operational strategy due to its smooth behavior. The settings for the ICDs in injectors 1 and 3 also showed very similar behavior to that observed in Fig. 7.6.

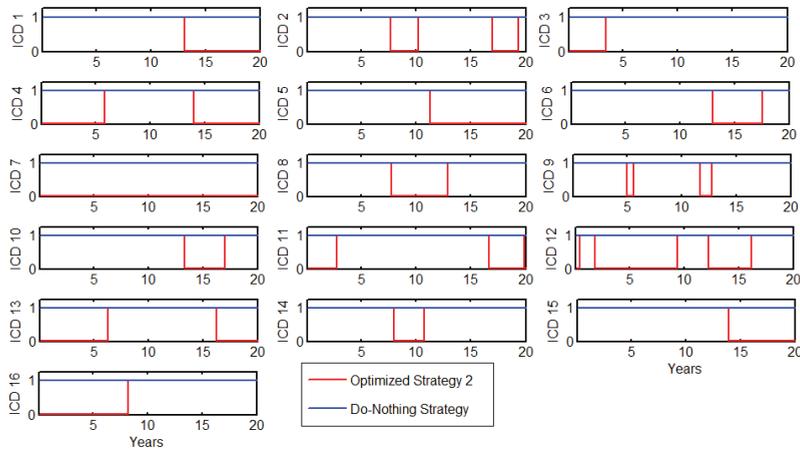


Fig. 7.6: Comparison of the optimized control strategy 2 (red) for injector 2 over the 20 year horizon to a do-nothing/reactive strategy (blue). In most cases the ICDs make one or two switches over the 20 year period.

7.4.5 Comparison of Different Gradient Formulations

Many alternative formulations to the standard robust ensemble gradient estimate have been proposed. Due to the computational complexity of the model, based on the findings of Fonseca et al. (2014, 2015) we investigate the application of the three robust gradient formulations discussed in the theory section. **Fig. 7.7** depicts the optimization process obtained by applying the different formulations. We observe that the original formulation (1:1 ratio) achieves a solution superior to the traditional pressure maintenance, “do-nothing” strategy; however, the “selected model formulation” (blue curve) leads to a solution which achieves an expected value 5% higher than the original formulation (green curve). Note: There is always ambiguity

with respect to how the models used in this approach are chosen and what impact a different set of chosen models would have on the optimization. The modified robust gradient formulation (red curve) achieves the highest expected objective function value which is approximately 5.5% higher than the value for the original formulation and 0.5% higher compared to the selected models formulation. The modified gradient formulation, like the original formulation, uses the entire ensemble of geological realizations, i.e. it accounts for all the uncertainty available to estimate the robust gradient, unlike the selected models approach.

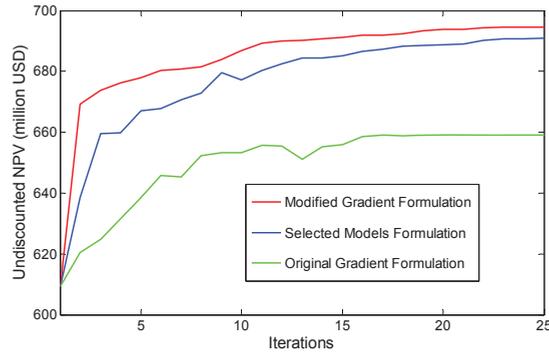


Fig. 7.7: Comparison of the optimization performance for different gradient formulations

7.4.6 Reduction in Number of ICDs

Since ICDs are expensive to install and operate, reducing the number of ICDs could be economically beneficial. However, the impact of having fewer ICDs could result in a reduction in controllability. The grouping of the ICDs is primarily based on the geological perspective and well path design. Van Essen et al. (2010) showed that a grouping of ICDs based on dynamic results instead of a geology-based grouping may lead to better results in terms of the objective function value. The well path design used in the present study is undulating. Compared to the example in Van Essen et al. (2010), in which the horizontal wells are completely horizontal, we expect that dynamic grouping will not lead to improved results. The dynamic grouping methodology proposed in Van Essen et al. (2010) is based on a visualization of the optimized control strategy. Visual inspection of the optimal control set illustrated Fig. 7.6 does not suggest any apparent dynamic grouping possibilities. **Fig. 7.8** is a comparison of the optimization procedure with a significant (approximately a factor four) reduction (blue curve) in the number of ICDs based on geological insight compared to the base case (red curve). We observe that having fewer ICDs results in a loss of controllability and thus an optimized strategy with a lower NPV. The difference in the objective function values between the two cases is approximately 3%, where the cost reduction due to a

reduced number of ICDs has not been taken into account. We do not have data about the actual costs (installation) of a single ICD. Based on discussions with field specialists the total costs saved as a result of installation of fewer ICDs used for this case is estimated, on the higher side, to be two orders of magnitude lower than the values of NPV shown in Fig. 7.8. Thus, for this example, the costs saved by the installation of fewer ICDs does not offset the gain in NPV achieved by using a higher number of ICDs.

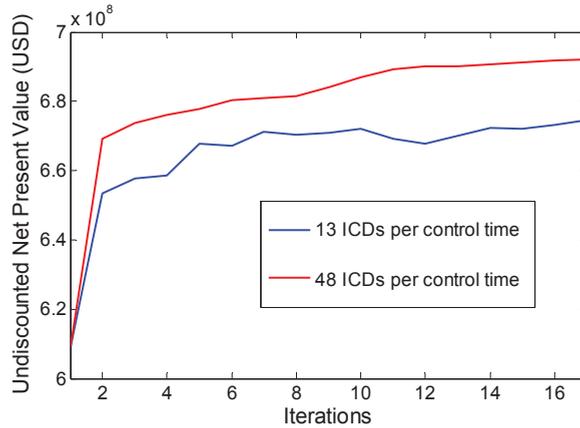


Fig. 7.8: Comparison of the impact of the use of a reduced number of ICDs based on geology and well path designs. A reduced set of ICDs (blue) results in reduced controllability and thus a lower NPV.

7.4.7 Hierarchical Switching Optimization

The hierarchical (bi-objective) switching optimization method is used to achieve multi-objective optimization under uncertainty as illustrated in Fig. 7.9. Following Van Essen et al. (2011), the primary objective is undiscounted NPV, displayed in Fig. 7.9, while the secondary objective is a highly (25%) discounted NPV to account for the short term gains. We observe a mean increase of approximately 9.2% in the secondary objective function compared to a maximum allowed mean decrease ε of 1% in the primary objective function. The switching optimization begins from the optimized solution achieved by the modified robust gradient formulation for life-cycle optimization. The modified formulation is also used for this hierarchical optimization. Additionally, the mean increase in cumulative oil produced over 20 years is marginally (2%) higher due to a higher (4.5%) increase in cumulative water injected compared to the solution achieved for life-cycle optimization. This confirms the general trend observed for this sector model, i.e., higher volumes of water injected will result in higher volumes of oil produced. The results illustrate the capacity of ensemble-based

multi-objective optimization under geological uncertainty to achieve results of practical importance.

Note: we have terminated the optimization for the secondary objective function after 15 iterations which translates to 1,500 reservoir simulations for the gradient estimate and 2,100 simulations for evaluation of the updated control set, i.e. a total of 3,600 reservoir simulations. While the total number of simulations is not high, due to the complex nature of the models, a forward simulation takes roughly 15 minutes. Thus, to obtain these results in a sequential manner would take roughly 37 days. An inherent advantage of EnOpt is that the gradient can be estimated using distributed computing, as has been done for the present study. The speed-up achieved in this case, using 25 cores, is roughly a factor 9, i.e. the robust hierarchical optimization was performed in approximately 4 days.

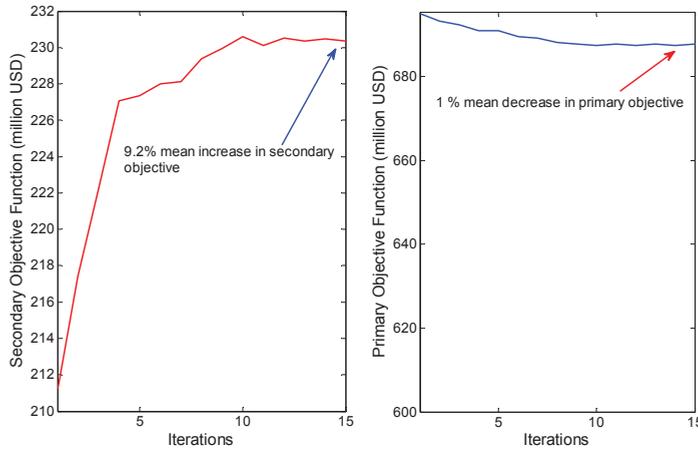


Fig. 7.9: Multi-objective optimization with an initial strategy based on life-cycle 0% discounted NPV, showing a mean 9.2% increase in the secondary objective function (25% discounted NPV) for a 1% mean decrease in primary objective function.

Fig. 7.10 depicts a comparison of the mean cumulative cash flow over time for the optimized solutions achieved by the switching algorithm (blue), life-cycle optimization (green) and reactive control (red). It is observed that after 2000 days the cumulative cash flow of the multi-objective optimization (blue curve) is approximately 12% higher compared to the life-cycle optimization (green curve) which will enable the project to achieve the break-even point faster, while the ultimate NPV of the two strategies is nearly equal. Similar to the results obtained in Fonseca et al. (2014), the ‘do-nothing’/reactive control strategy gives the best short-term performance at a price of the worst long-term performance.

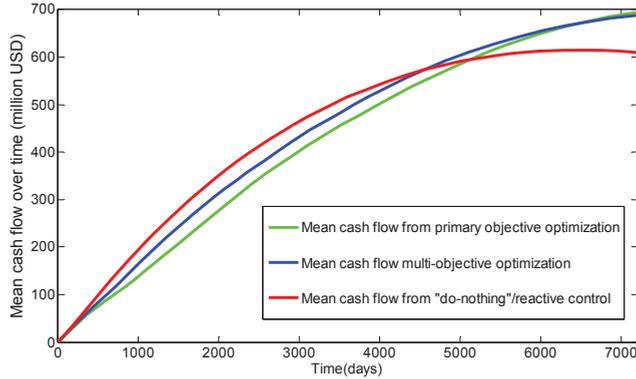


Fig. 7.10: Comparison of the mean cash flow over time for the entire ensemble of geological realization for the different optimization strategies.

7.4.8 Discussion

We have not considered the commonly used weighted-sum method in this work because it is impossible to know a-priori which weight combination will give the desired results. The advantage of the hierarchical approach is that the user decides the maximum allowable decrease in the primary objective. This feature is not known a-priori when using the weighted-sum method. We would need to perform a trial and error set of experiments which would be computationally demanding.

Also, we did not use a full Pareto curve approach in which a large number of weighted-sum simulations is performed with different weight combinations. Generating such a full Pareto curve would provide a decision maker a range of possible solutions to choose from. However, computing the full Pareto curve is computationally much more intensive and is outside the computational limits of this study. The choice of $\varepsilon = 1\%$ in our hierarchical switching approach is simply a choice; a user is free to decide his own choice depending on how much a user values the long-term targets. We do not claim that this is the correct choice, because, as with any multi-objective approach, especially those that aim to generate a Pareto front, the idea is to obtain a range of possible solutions to choose from. No single solution is necessarily the correct one. Van Essen et al. (2011), who first introduced the hierarchical switching method, have included an illustration of the principle and working of this hierarchical switching method. Due to the large drop in oil prices in 2015, we repeated the optimization for a much lower oil price of 50 \$/bbl while keeping the water costs the same. **Fig 7.11** depicts the results. Compared to Fig. 7.3 we now increase the objective function value much more (approximately 35% against approximately 12% in Fig. 7.3) which illustrates that at lower oil prices the effect of reducing water production and injection becomes more important.

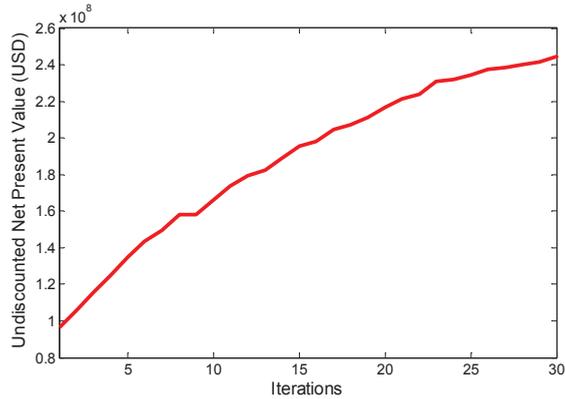


Fig 7.11: Mean increase in objective function value over the ensemble of realizations for a 50\$/bbl oil price

7.5 Conclusions

- ✓ An ensemble-based robust multi-objective optimization workflow tested on a sector model inspired from a real field case shows results of practical value against acceptable computational cost.
- ✓ Parameterization of ‘on-off’ type controls using a switching time interval method is efficient when working with stochastic gradient techniques, such as EnOpt.
- ✓ The hierarchical switching algorithm leads to an approximately 9% mean increase in the secondary objective function (short-term targets) against a mean decrease of 1% in the primary objective function (life-cycle targets)
- ✓ The modified formulation based on the 1:1 ensemble ratio is not only computationally attractive but also uses the entire ensemble of geological realizations i.e. captures all the uncertainty available.
- ✓ The optimized strategy obtained with the modified formulation achieved a 4.5% increase in the expected objection function value over the strategy obtained by the original formulation and a 3.5% increase over a reactive control strategy. The main scope of optimization for this example lies in the reduction of the volumes of water injected and produced.
- ✓ A reduction in the number of ICDs results in a loss of controllability and thus lower objective function values. However, it may still be a better strategy if the cost of the ICDs is incorporated into the objective function.
- ✓ The optimization is sensitive to the initial starting point. Two different optimized strategies have been provided for the decision maker to choose from.

CONCLUSIONS & FUTURE PERSPECTIVES

8.1 Conclusions

In this chapter we first provide general conclusions from this thesis and outline future perspectives of the applicability of EnOpt for model-based oil recovery optimization. In this thesis we have proposed a modified gradient formulation for ensemble optimization under geological uncertainty. We have shown that this modified gradient formulation is theoretically more ‘robust’ compared to the original formulation proposed by Chen (2008). This formulation performs significantly better than the original formulation and in terms of accuracy, potentially comparable to the adjoint method in an optimization context. We have shown that ensemble size may have a big impact on gradient quality depending on the region within the objective function space. We have also investigated the impact of gradient quality on an optimization experiment and illustrated the need to achieve high quality gradients for different cases. In this thesis we have also proposed an improvement to the gradient quality through an iterative updating of the covariance matrix used to generate the ensemble of controls. The CMA-EnOpt algorithm highlights the benefit of ‘learning’ about the objective function search space during the optimization. We have also investigated the applicability of using EnOpt to solve bi-objective optimization problems using a hierarchical structure for problems that take into account geological uncertainty. We have also illustrated the applicability of EnOpt to generate a Pareto front by solving a constrained optimization problem. Finally we have used some of the proposed modifications in EnOpt to achieve bi-objective optimization applied to an ensemble of realistic reservoir sector models inspired from a real field case. We have used the inherent parallelization possible to enhance the computational efficiency of

the method when using large scale models. Thus we show that ensemble based optimization workflows can be used to produce reservoir management strategies of significant practical value. The following sub-sections contain chapter specific conclusions.

Effect of Ensemble Size on Gradient Quality

- ✓ For the relatively simple Rosenbrock function we need an ensemble size equal to 900, 5 or 3 to satisfy the null hypothesis and achieve the desired confidence interval depending on the perturbation size used ($\sigma = 0.1$, $\sigma = 0.01$ and $\sigma = 0.001$ respectively) to generate the ensemble of controls.
- ✓ Including uncertainty within the Rosenbrock function we find that the modified formulation [based on equations (2.22) and (2.23)] with the computationally attractive 1:1 ratio outperforms the original formulation [based equations (2.19) and (2.20)], irrespective of the degree of uncertainty. However the degree of uncertainty does play a role in the quality of the gradient estimate. Higher-quality gradients can be obtained with an increasing ensemble size of model realizations i.e. using higher ratios.
- ✓ When working with higher ratios significantly better results are obtained through the use of the ‘Mean of Individual Gradients’ (MIG) formulation compared to the HPG formulation. This result is not surprising as this would be expected from the central limit theorem. Additionally the ratios required are smaller; for the Rosenbrock case, depending on the perturbation size, a ratio of 1:3 is sufficient to satisfy the hypothesis and meet the desired confidence interval using the MIG formulation.
- ✓ The analysis for the “Egg Model” case was divided into two regions: 1) a ‘steeper’ region, where with higher perturbation size ($\sigma = 0.01$) and lower ensemble sizes (150 samples) the hypothesis was satisfied at all points, and 2) a ‘flatter’ region where a smaller perturbation size ($\sigma = 0.001$) and a higher ensemble size (300 samples) were needed.
- ✓ In the flatter region, in addition to the perturbation size, a correlated covariance is needed where the choice of correlation length has a significant impact on the gradient quality. Through a trial and error procedure we observe that correlation lengths between 8 and 13 give the best results and lead to satisfying the hypothesis at most points.
- ✓ For the robust “Egg Model” case, we observe a similar trend in the results as for the uncertain Rosenbrock function. The modified formulation using the 1:1 ratio achieves significantly better results compared to the original formulation at any point along the optimization curve

- ✓ In the flatter part of the optimization curve the modified formulation with the 1:1 ratio, although performing better than the original formulation, never satisfies the hypothesis. Using higher ratios in the flatter region is necessary to achieve a good quality gradient. However, in the steeper part, the hypothesis is satisfied for all the points.
- ✓ As the results have shown, the developed methodology can, in theory, be used to quantify the ensemble size required to achieve a high-quality gradient. However, all the angles obtained in this work are always less than 90 degrees which suggests that with ensemble methods, irrespective of the ensemble size (in our case), we estimate the ‘correct’ uphill direction.
- ✓ Use of a gradient of inferior quality in an optimization experiment strongly influences the final achievable NPV for a finite number of optimization iterations.
- ✓ We recommend to use, out of the 30 potential robust gradient formulations identified in our paper, the single ‘smoothed’ modified formulation, i.e. equation (2.15) based on equations (2.22) and (2.23), using the 1:1 ratio for recovery optimization under uncertainty. In an optimization context, it is vital to use good quality

EnOpt with Covariance Matrix Adaptation

- ✓ A comparison between CMA-EnOpt and EnOpt for a simple five-spot model showed consistently (somewhat) higher objective function values and modest speed-ups for CMA-EnOpt, depending on the choice of user-defined parameters in both algorithms.
- ✓ The major benefit of CMA-EnOpt is its robustness with respect to the initial choice of the covariance matrix. A poor choice of the initial matrix can be detrimental to EnOpt, whereas the CMA-EnOpt performance is near-independent of the initial choice.
- ✓ Learning rates are crucial for the success of CMA-EnOpt. For both the simple five-spot model and the modified Brugge model, a 75%-25% update rule proved to be successful.
- ✓ For the simple five-spot model, the methods that explicitly use gradient information (EnOpt and CMA-EnOpt), performed better than the method that doesn’t do so (CMA-ES). (Not tested for the modified Brugge model).
- ✓ A comparison between CMA-EnOpt and EnOpt for the modified Brugge model revealed slightly lower to significantly higher (-1% - +9%) objective function values depending on choice of user-defined parameters in both algorithms.

- ✓ Updating a block-diagonal (i.e. time-correlated) covariance matrix leads to significant improvements in the results as well as in the efficiency of the algorithm, compared to using a prescribed correlation (smoothing) and compared to updating either diagonal elements only, or updating the full matrix.
- ✓ The different rank updates play different roles in the success of the optimization; for the simple five-spot model the rank μ update performed much better than the rank one update, with a combined rank update ending up in-between. For the complex Brugge model, however, the rank μ update performed worst, while the combined-rank update performed best. Further experience is needed to arrive at general recommendations.
- ✓ Robustness to the choice of the initial covariance matrix, and higher objective function values are the main advantages of CMA-EnOpt over EnOpt.

Hierarchical Bi-Objective Optimization

- ✓ Compromises made to short-term targets during life cycle optimization can be partly corrected for with an ensemble-based hierarchical multi-objective optimization method.
- ✓ The EnOpt method is a good alternative to achieve practical results when the adjoint formulation is not available for hierarchical multi-objective optimization.
- ✓ In our numerical simulation examples, two hierarchical multi-objective methods showed a 14.2% improvement in the secondary objective function (NPV @ 25% discount rate) approximately constrained to the primary objective function (NPV @ 0% discount rate). The results obtained with the null-space-based optimization algorithm are similar to those resulting from the switching algorithm, although for the case investigated here, the switching algorithm was found to be computationally somewhat more efficient.
- ✓ The BFGS algorithm, used to estimate the Hessian for the null-space method, is computationally attractive compared to a finite difference method especially when dealing with large control sets, and led to good results for the case reported here.
- ✓ Hierarchical multi-objective optimization of ICV settings shows significant scope for improvement in short to medium term goals approximately constrained to life cycle targets.

Pareto Fronts for Bi-Objective Optimization

- ✓ Approximate gradient techniques like EnOpt can be used to generate solutions which may lie on a pareto front for a bi-objective optimization problem within acceptable computational effort.
- ✓ Tracking the Pareto front using NBI is a computationally more efficient method and produces better solutions for the decision maker to choose from compared to the original NBI form. Different starting point have a significant impact on the optimal solutions achieved.
- ✓ The adjusted weighted sum produces a more even distribution of solutions and is marginally computationally more efficient compared to the traditional weighted sum technique for this case.
- ✓ For some weight combinations the NBI method produces solutions which dominate solutions obtained by the weighted sum variants and vice versa.
- ✓ A hierarchical switching method provides a single solution which satisfies the maximum allowable decrease in objective function value however the solution is always dominated by the solutions obtained by the other methods.

Robust Bi-Objective Optimization on On/Off Inflow Control Devices :

- ✓ An ensemble-based robust multi-objective optimization workflow tested on a sector model inspired from a real field case shows results of practical value against acceptable computational cost.
- ✓ Parameterization of ‘on-off’ type controls using a switching time interval method is efficient when working with stochastic gradient techniques, such as EnOpt.
- ✓ The hierarchical switching algorithm leads to an approximately 9% mean increase in the secondary objective function (short-term targets) against a mean decrease of 1% in the primary objective function (life-cycle targets)
- ✓ The modified formulation based on the 1:1 ensemble ratio is not only computationally attractive but also uses the entire ensemble of geological realizations i.e. captures all the uncertainty available.
- ✓ The optimized strategy obtained with the modified formulation achieved a 4.5% increase in the expected objection function value over the strategy obtained by the original formulation and a 3.5% increase over a reactive control strategy. The main scope of optimization for this example lies in the reduction of the volumes of water injected and produced.
- ✓ A reduction in the number of ICDs results in a loss of controllability and thus lower objective function values. However, it may still be a better strategy if the cost of the ICDs is incorporated into the objective function.

- ✓ The optimization is sensitive to the initial starting point. Two different optimized strategies have been provided for the decision maker to choose from.

8.2 Future Perspectives

In this thesis we have investigated a few important topics with respect to the theoretical understanding and applicability of EnOpt, however there exists further scope for research. The remainder of this chapter delves into some perspectives that could constitute topics of future research directions which focus on approximate gradient techniques such as EnOpt.

Algorithm for Adaptive Ensemble Size and CMA-EnOpt

The results from Chapter 3 illustrate that different ensemble sizes as well covariance matrices are required to obtain a certain desired gradient quality at different regions and points in the objective function space. Chapter 4 highlights the effectiveness in terms of objective function value of iteratively adapting the covariance matrix throughout the optimization. However, there does not exist a technique to iteratively adapt the ensemble size used in the gradient estimate. Investigation into the possibility of an adaptive scheme for the ensemble size and its effect on optimization results will be of significant interest. In addition to achieving high quality gradients throughout the optimization process such an adaptive algorithm will also make the method less dependent on user choices of these parameters, thus more ‘robust’ in applicability. A coupling of such an adaptive ensemble size algorithm with the CMA algorithm could probably lead to even more ‘robust’ algorithm. In the CMA algorithm a choice needs to be made for the number of ‘successful’ directions to be used in the matrix update, the effect of an adaptive ensemble size would be one of a few sub-topics which need to be investigated.

Further investigations into Gradient Quality

The results from Chapter 4 imply that the estimated gradient quality is improved using CMA-EnOpt compared to standard EnOpt. Thus it would be interesting to investigate if the gradient as a result of CMA-EnOpt is actually a better quality gradient. In chapter 3 we have investigated the impact of ensemble size on gradient quality for a 320 dimensional problem. Additionally, the impact of the problem dimension therefore needs to be investigated in terms of ensemble size. An increase in the problem dimension i.e., number of controls, means that the gradient equation that needs to be solved becomes more underdetermined for a fixed ensemble size. To the best of our knowledge the impact of problem dimension has never been investigated

for EnOpt and similar type of methods. Thus to provide general recommendations on ensemble sizes the impact of problem dimension needs to be investigated.

For robust optimization problems in this thesis we recommend the use of the modified EnOpt formulation with the 1:1 ratio for computational reasons. In our experiments, the ensemble size of geological realizations is 100 which is generally used. Thus another interesting sub-topic to investigate would be; *“Does the 1:1 ratio still produce high quality gradients when the ensemble size of geological models is significantly smaller, e.g. less than 10 ?”*. A secondary question would then be to investigate what would be the minimum ensemble size of geological realizations needed to use the 1:1 ratio with the modified formulation.

An economic objective (NPV) over a significant time horizon (10-20years) is generally the objective function used for model based optimization studies. NPV is a single number obtained at a particular end time while the controls usually change through time. This leads to the question: *“Can better quality gradient estimates be obtained if the gradient is estimated per control time step ?”* Thus instead of solving a single regression equation to estimate an “overall” gradient, multiple regression equations would need to be solved for each individual control time step. The final gradient to be used in the optimization would then be a concatenated gradient of the various gradients estimated per time step. This could be another factor to improve gradient quality and thus coupled with problem dimension this could form a future research direction into the understanding of approximate gradient techniques.

Use of an adaptive simplex Hessian matrix to improve gradient quality

All gradient based optimization exercises consists of two components. First, the gradient itself, either the adjoint gradient or an approximate gradient and second, an algorithm to choose a step in the direction of the gradient. In model based recovery optimization cases usually the steepest ascent method is used which only uses the gradient information. While the gradient contains information about the direction in which an objective function can be improved/increased it does not contain information about the local curvature of the objective function space which may be particularly important when working with approximate gradient techniques: In chapter 3 we have shown that in the flatter regions of the optimization space the gradient approximation is of inferior quality possibly because curvature information is not accounted for. In our problems it is not uncommon to postulate the presence of ridges in the objective function space which have ‘typical’ orientations. The curvature information or orientation of the ridges could possibly be accounted for through the use of the Hessian matrix. The Hessian is usually not available or estimating this matrix is computationally challenging. While there exists algorithms such as BFGS to iteratively

approximate the Hessian matrix, it requires many iterations before the approximate matrix is of sufficient quality to be used. Use of an inferior Hessian matrix within a Newton scheme may have a negative impact on the optimization. Thus it would be interesting to investigate the use of a quasi-Newton method for optimization with approximate gradients. Zhao et al. (2013) investigated the use of quadratic interpolation methods with EnOpt to account for the curvature in the optimization search space with limited success. The success of such a scheme will probably depend on the gradient quality as the iterative updating of the Hessian matrix uses gradient information. Conn et al. (2009) provide details of an algorithm which can be used to estimate a diagonal Hessian using the simplex method. The EnOpt formulation we employ is essentially the simplex method described in Conn et al. (2009). Therefore investigation into the use of such simplex Hessians within an optimization context would be interesting. This simplex Hessian would also need to be investigated in conjunction with gradient quality i.e. ensemble size and covariance matrices.

CMA-EnOpt for Robust Optimization

The CMA-EnOpt algorithm from Chapter 4 was shown to improve the robustness of EnOpt to the choice of the covariance matrix for deterministic optimization cases. When dealing with robust optimization problems it would be interesting to investigate the impact of using CMA-EnOpt. Currently CMA-EnOpt fails for robust optimization cases because the update of the covariance matrix uses a ranking scheme to choose the ‘best’ directions to be used. When using the 1:1 ratio every perturbed control ensemble member is coupled with a different geological model and thus the ranking takes into account the geological effects instead of only the perturbations in the controls. This has a negative impact on the optimization. Investigation into finding a solution to this problem will help in the applicability of CMA-EnOpt for robust optimization to find optimal well controls or optimal well placement strategies but also for simultaneous optimization of well placement and well controls. In this case as well, a few sub-topics for research exists such as should we use a subset of geological models for the optimization, and if so, should we update a single covariance matrix for all the models or individual matrices for the individual models, etc..

Use of CMA for multiscale control regularization

Recently there has been an increased focus on multi-scale control regularization based on the reasoning that the optimal number of control variables is not known a-priori. Various authors, Lien et al. (2008), Shuai et al. (2011) and Oliviera et al. (2014), have proposed different methods to iteratively find the optimum number of controls needed for the optimization. The methods are primarily based on refinement indicators

using gradient information and/or objective function information. In Chapter 4 we propose the CMA-EnOpt algorithm where the covariance matrix is adapted based on knowledge of the objective function space. Thus the use of the updated covariance matrix for multi-scale control regularization could be investigated as an alternative to the existing methods. The question that arises is “ *Does the updated Covariance Matrix provide information to quantify the optimal number of controls to be used in the optimization ?*”. In chapter 4 we show that during the update of the covariance matrix many of the controls have very small standard deviations and only a few retain larger standard deviation values. How this information could be used to refine or merge controls would be an interesting topic for research. The effect of such a scheme will enhance the applicability of EnOpt to a variety of problems.

Applicability of reduced order modeling/multiscale methods with EnOpt.

One of the major drawbacks of using any approximate gradient technique is the computational effort required to estimate the gradient. This is particularly challenging when working with realistic reservoir models of high dimension. Such models can sometimes take hours to days to complete a single forward simulation. Thus the use of approximate gradient methods which require multiple forward simulations to estimate a single gradient is challenging. On the other hand there has been a lot of research into the applicability of reduced-order models as a proxy to the high fidelity large scale models that approximately capture the relevant flow dynamics of the system. These reduced order models are fast to solve and thus computationally very attractive. Van Doren et al. (2005) used POD based reduced order models for production optimization using the adjoint formulation for the gradients while Cardoso and Durlofsky (2011) use TPWL for reduced order modeling with finite difference gradients. Thus it would be interesting to investigate the use of reduced-order models in conjunction with approximate gradient methods such as EnOpt for model based recovery optimization. This could potentially improve the computational efficiency of the method especially when working with large scale realistic reservoir model. While the computational benefit of such an approach is apparent, the gradient quality and optimization performance need to be investigated.

EnOpt for optimization of complex physical process

Most of the studies using EnOpt have focused on water flooding optimization. Raniolo et al. (2013) and Chen and Reynolds (2015) have applied EnOpt for polymer flooding and water alternating gas applications. They showed that significant scope exists to optimize such processes and EnOpt was able to achieve results of practical value. Namdar Zanganeh et al. (2014) investigated the applicability of an adjoint

gradient for optimization of a foam based EOR process. They observed small scale fluctuations in the objective function as a result of oscillations in time in the forward simulation caused by strong nonlinearities in the foam model. They suggest that similar behavior is expected for the simulation of other EOR techniques such as surfactant flooding and near-miscible flooding. These oscillations had a negative impact on the use of ‘local’ adjoint gradients. EnOpt on the other hand does not calculate the gradient at each simulation time step and thus the oscillations in the forward modelling process may not affect the ability to use a gradient based optimization algorithm to optimize foam EOR processes. Thus it would be very interesting to investigate the applicability of EnOpt for optimization of more complex physical problems such as different EOR processes especially for problems where the objective function is non-smooth. The effect of perturbation sizes for these types of problems will also need to be investigated. Additional smoothing formulations as proposed in Chen et al. (2009) could be an added advantage when using EnOpt for such problems.

Detailed comparison with derivate free methods

Approximate gradient techniques share many similarities with derivative free methods such as genetic algorithms, evolutionary strategies etc. In chapter 4 we have made a basic comparison between EnOpt and CMA-ES. Many derivative free methods, in addition to choosing an ensemble size and a covariance matrix for sampling require a user to a-priori choose variables such as learning rates, mutation coefficients etc. The choice of all these additional variables would naturally be case/problem dependent and a-priori knowledge of the values of these variables is impossible. Thus for a better understanding of CMA-EnOpt a sensitivity analysis on the number of variables such as learning rates etc., which need to be chosen will be interesting. While it is impossible to know a-priori which method will produce the best results, it would be interesting to make a detailed comparison between the advantages and dis-advantages between derivative free methods and approximate gradient techniques. This comparison would be very useful for robust optimization cases especially regarding the computational efficiency and optimized solutions achieved. These results could be used to investigate and propose new hybrid algorithms which combine the desired properties of the different methods.

Effect of time stepping and convergence of forward simulation on quality of approximate gradient

For any algorithm that is based on using a simulator as a black-box it is vital to understand the outputs from the black-box. Inaccurate outputs from the black-box will lead to inaccurate results. In any reservoir simulator convergence of the forward

simulation as well as the time stepping strategy used is vital to achieve accurate results. Volkov and Voskov (2013) showed the effect of different time stepping criteria on the quality of the adjoint gradient. When working with exact gradients for each time step this is even more important, however the impact of time stepping and convergence of the forward simulation on the accuracy of approximate gradients such as EnOpt could be considered as a topic for further research.

Constraint Handling with EnOpt.

In order to achieve results of practical value it is vital to include constraints for e.g., maximum oil production rate, maximum water injection rate etc., within the optimization. A few recent papers, Dehdari and Oliver (2012), Leeuwenburgh et al. (2015), etc., have addressed the use of EnOpt for constrained optimization problems. While there exists many methods to solve the constrained optimization problem it would be interesting to make a comparison of the performance of various constraint optimization techniques especially for EnOpt under geological uncertainty.

This chapter provides a list of a few possible research directions regarding the applicability and use of approximate gradient techniques such as EnOpt. Other, more general research areas within the optimization context such as visualization of the objective function space or different algorithms for solving a multi-objective optimization problem etc. also warrant investigation.

APPENDIX⁸

Degrees of Freedom

When dealing with an optimization problem, there may exist multiple sets of control variables for which we achieve similar results. Those sets of different optimal control variables are an indication of the presence of redundant degrees of freedom (DOFs) in the system. This existence of multiple solutions suggests that, when the optimum of a primary objective function has been reached, not all DOFs of the control variable space are fixed. This implies that there may exist redundant DOFs in the optimization problem. This conclusion formed the basis for development of the multi-objective optimization algorithm in Van Essen et al. (2011). Van Essen et al. (2011) approximate a Hessian matrix to find these DOFs, which is an integral part of the null-space-based hierarchical optimization method. Henceforth we denote $\mathbf{H} = \nabla^2 J(\mathbf{u}^*)$ where \mathbf{H} is the Hessian matrix, a matrix of second order derivatives of the objective function. Consider an objective function J and let \mathbf{u}^* be a control vector. If $\Delta\mathbf{u}$ is a vector (with same length as \mathbf{u}^*) of small perturbations then a Taylor expansion around the vector \mathbf{u}^* is given by

$$J(\mathbf{u}^* + \Delta\mathbf{u}) \approx J(\mathbf{u}^*) + \nabla J(\mathbf{u}^*)^T \cdot \Delta\mathbf{u} + \frac{1}{2}(\Delta\mathbf{u})^T \nabla^2 J(\mathbf{u}^*) \Delta\mathbf{u} + O(\|\Delta\mathbf{u}\|^3) \quad (\text{A-1})$$

If \mathbf{u}^* is a (local) optimum of J and \mathbf{u}^* is in the interior (i.e. not on the boundary) of the feasible domain for a constrained optimization then we can conclude that

$$\nabla J(\mathbf{u}^*) = \mathbf{0} \quad (\text{A-2})$$

Substituting equation (A-2) into equation (A-1) we obtain

$$J(\mathbf{u}^* + \Delta\mathbf{u}) \approx J(\mathbf{u}^*) + \frac{1}{2}(\Delta\mathbf{u})^T \nabla^2 J(\mathbf{u}^*) \Delta\mathbf{u} + O(\|\Delta\mathbf{u}\|^3) \quad (\text{A-3})$$

If we choose $\Delta\mathbf{u} \in \text{null}[\nabla^2 J(\mathbf{u}^*)]$ then $\nabla^2 J(\mathbf{u}^*) \Delta\mathbf{u} = \mathbf{0}$, and equation (A-3) reduces to

⁸ Taken from Fonseca, R.M., Leeuwenburgh, O., Van den Hof, P.M.J. and Jansen, J.D. 2014. Ensemble based Hierarchical Multi-Objective Production Optimization of Smart Wells. *Computational Geosciences* **18**(3-4).

$$J(\mathbf{u}^* + \Delta\mathbf{u}) \approx J(\mathbf{u}^*) + O(\|\Delta\mathbf{u}\|^3) \quad (\text{A-4})$$

Thus equation (A-4) implies that for a small perturbation $\Delta\mathbf{u} \in \text{null}[\nabla^2 J(\mathbf{u}^*)]$, the adjusted control vector $(\mathbf{u}^* + \Delta\mathbf{u})$ will have an objective function value very close to the objective function value $J(\mathbf{u}^*)$ which is an optimal value. This means that we can make an update to the control vector that is in the null space of the primary objective function to improve the secondary objective function. This proves the need to find the Hessian matrix at the optimum of the primary objective function and the set of vectors that span its null space. A singular value decomposition is used to obtain the null space and orthonormal basis \mathbf{B} used in the algorithm.

BIBLIOGRAPHY

- Alhuthali, A.H., Datta-Gupta, A., Yuen, B. and Fontanilla, J.P. 2009. Field applications of waterflood optimization via optimal rate control with smart wells. Paper SPE 118948 presented at the *SPE Reservoir Simulation Symposium*, The Woodlands, USA, 2-4 February. DOI: 10.2118/118948-MS.
- Arnold, D.V. and Hansen, N., 2010. Active Covariance Matrix Adaptation for the (1+1)-CMA-ES. In *Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation. GECCO'10*. Portland, Oregon, USA, 7-11 July 2010.
- Awotunde, A.A., and Sibaweih, N. 2012. Consideration of Voidage Replacement Ratio in Well Placement Optimization. Paper SPE 163354 presented at the *SPE Kuwait International Petroleum Conference and Exhibition*, Kuwait City, Kuwait, 10-12 December.
- Bailey, W.J., Couet, B. and Wilkinson, D., 2005. Framework for Field Optimization to Maximize Asset Value, SPE 87026-PA, *SPE Reservoir Engineering Journal*, **8** (1), pp.7-21
- Beck, J.V. and Arnold, K.J., 1997. *Parameter Estimation in Engineering and Science*. John Wiley & Sons, Cambridge, United Kingdom.
- BP Statistical Review of World Energy*, June 2014.
- Brouwer, R. and Jansen, J.D. 2004. Dynamic Optimization of Water Flooding With Smart Wells Using Optimal Control Theory. *SPE Journal* **9**(4) 391-402.
- Bouzarkouna, Z., Ding, Y.D. and Auger, A., 2011. Partially Separated Meta-models with Evolution Strategies for Well Placement Optimization. Paper SPE-143292-MS presented at the *SPE EUROPEC/EAGE Annual Conference and Exhibition*, Vienna, Austria, 23-26 May. DOI: 10.2118/143292-MS.
- Capolei, A., Suwartadi, E., Foss, B. and Jørgensen, J.B. 2013. Waterflooding optimization in uncertain geological scenarios. *Computational Geosciences* **17**(6) 991-1013.
- Cardoso, M.A. and Durlofsky, L.J., 2010. Use of reduced order modeling procedures for production optimization. *SPE Journal* **15** (2) 426-435. DOI: 10.2118/119057-PA.
- Chaudhri, M.M., Phale, H.A. and Liu, N. 2009. An improved approach for ensemble-based production optimization: application to a field case. Paper SPE-121307-MS presented at the *EUROPEC/EAGE Conference and Exhibition*, Amsterdam, The Netherlands, 8-1 June. DOI: 10.2118/121307-ms.
- Chen, Y. 2008. Efficient Ensemble based Reservoir Management. PhD Thesis, University of Oklahoma, USA.
- Chen, Y. and Oliver, D.S. 2010. Ensemble-Based Closed-Loop Optimization Applied to Brugge Field. *SPE Reservoir Engineering and Evaluation* **13** (1) 56-71. DOI: 10.2118/118926-PA

- Chen, Y., Oliver, D.S. and Zhang, D. 2009. Efficient Ensemble-Based Closed-Loop Production Optimization. *SPE Journal* **14** (4) 634-645. DOI: 10.2118/112873-PA
- Chen, Y. and Oliver, D.S., 2012. Localization of Ensemble-Based Control-Setting Updates for Production Optimization. *SPE Journal* **17**(1) 122-136. DOI: 10.2118/125042-pa.
- Chen, C., Li, G. and Reynolds, A. 2012. Robust Constrained Optimization of Short- and Long-Term Net Present Value for Closed-Loop Reservoir Management. *SPE Journal* **17**(3) 849-864.
- Chen, B., and Reynolds, A.C. 2015. Ensemble-based optimization of the WAG injection process. Paper SPE 173217-MS presented at the *SPE Reservoir Simulation Symposium*, Houston, USA, 23-25 February.
- Conn, A.R., Scheinberg, S., Vicente, L.N. 2009: *Introduction to Derivative-Free Optimization*. SIAM, Philadelphia.
- Das I. and Dennis, E. 1998. Normal Boundary Intersection: A new method for generating the Pareto surface in nonlinear multi-criteria optimization problems. *SIAM Journal of Optimization*, 8, 631-657
- Dehdari, V. and Oliver, D.S., 2012: Sequential Quadratic Programming for Solving Constrained Production Optimization--Case Study From Brugge Field. *SPE Journal* **17** (3) 874-884. DOI: 10.2118/141589-PA.
- Dekking, F., Kraaikamp, C., Lopuhaä, H., and Meester, L. 2005. *A Modern Introduction to Probability and Statistics: Understanding Why and How*. Springer London, DOI: 10.1007/1-84628-168-7
- Ding, Y., 2008. Optimization of Well Placement Using Evolutionary Algorithms. Paper SPE-113525-MS presented at the *Europec/EAGE Conference and Exhibition*, Rome, Italy, 9-12 June 2008. DOI: 10.2118/113525-MS.
- Do, S.T. and Reynolds, A.C., 2013: Theoretical connections between optimization algorithms based on an approximate gradient. *Computational Geosciences* **17** (6) 959-973. DOI: 10.1007/s10596-013-9368-9.
- Eclipse. 2011. <http://www.slb.com/services/software/reseng/eclipse>.
- Fonseca, R.M., Leeuwenburgh, O. Van den Hof, P.M.J. and Jansen, J.D. 2014. Ensemble based Hierarchical Multi-Objective Production Optimization of Smart Wells. *Computational Geosciences* **18** (3-4), 449-461.
- Fonseca, R.M., Kahrobaei, S., Van Gastel, L.J.T., Leeuwenburgh, O., and Jansen, J.D. 2015. Quantification of the Impact of Ensemble Size on the Quality of an Ensemble Gradient Using Principles of Hypothesis Testing. Paper SPE 173236-MS presented at the *SPE Reservoir Simulation Symposium*, Houston, USA, 23-25 February.
- Fonseca, R.M., Leeuwenburgh, O., Van den Hof, P.M.J. and Jansen, J.D, 2015: Improving the ensemble optimization method through covariance matrix adaptation (CMA-EnOpt). *SPE Journal* **20** (1) 155-168. DOI: 10.2118/163657-PA.

- Fonseca, R.M., Stordal, A.S., Leeuwenburgh, O. Van den Hof, P.M.J. and Jansen, J.D. 2014. Robust ensemble-based multi-objective optimization. *Proc. 14th European Conference on the Mathematics of Oil Recovery (ECMOR XIV)*, Catania, Italy, 8-11 September.
- Fonseca, R.M., Leeuwenburgh, O., Della Rossa, E., Van den Hof, P.M.J. and Jansen, J.D. 2015. Ensemble-based multi-objective optimization of on-off control devices under geological uncertainty. SPE 173268-MS presented at the 2015 SPE Reservoir Simulation Symposium, 22-25 February, Houston, USA.
- Forouzanfar, F, Della Rossa, E., Russo, R. and Reynolds, A.C. 2013. Life-cycle production optimization of an oil field with an adjoint-based gradient approach. *Journal of Petroleum Science and Engineering* **112** 351-358. DOI: 10.1016/j.petrol.2013.11.024.
- Forouzanfar, F, Poquioma, W.E. and Reynolds, A.C. 2015. A Covariance Matrix Adaptation Algorithm for simultaneous estimation of optimal placement and control of production and water injection wells. Paper SPE 173256-MS presented at the *SPE Reservoir Simulation Symposium*, Houston, USA, 23-25 February.
- Golub, G. and van Loan, C., 1989. *Matrix Computations*, 4th ed. The John Hopkins University Press, Baltimore, USA.
- Hansen, N., 2006. The CMA Evolution Strategy: A Comparing Review. In Lozano, J. A., Larranaga, P., Inza, I. and Bengoetxea, E. *Towards a new evolutionary computation. Advances on estimation of distribution algorithms*: 75-102, Springer.
- Hansen, N., 2011. The CMA Evolution Strategy: A Tutorial. <http://www.lri.fr/~hansen/cmatutorial110628.pdf>.
- Hansen, N. and Ostermeier, A., 1996. Adapting Arbitrary Normal Mutation Distributions in Evolution Strategies: The Covariance Matrix Adaptation. In *Proceedings of the 1996 IEEE International Conference on Evolutionary Computation*.
- Hansen, N. and Ostermeier, A., 2001. Completely Derandomized Self-Adaptation in Evolution Strategies. *Evolutionary Computation* **9**(2) 159-195.
- Hasan, A. and Foss, B. 2013. Optimal Wells Scheduling of a Petroleum Reservoir. *Proc. European Control Conference (ECC)*. Zurich, Switzerland, 17-19 July.
- Isebor, O.J. and Durlofsky, L.J. 2014. Bi-objective optimization for general oil field development. *Journal of Petroleum Science and Engineering*, **119**, 123-138, DOI: 10.1016/j.petrol.2014.04.021
- Isebor, O.J., Echeverría Ciaurri, D. and Durlofsky, L.J. 2014. Generalized field-development optimization with derivative-free procedures. *SPE Journal* **19** (5) 891-908. DOI: 10.2118/163631-PA.
- Jansen, J.D., Brouwer, D.R., Nævdal, G. and van Kruijsdijk, C.P.J.W., 2005. Closed-loop reservoir management, *First Break* **23** (2005), pp. 43-48.
- Jansen, J.D., Douma, S.G., Brouwer, D.R., Van den Hof, P.M.J., Bosgra, O.H. and Heemink, A.W., 2009. Closed-loop reservoir management. Paper SPE 119098 presented at *SPE Reservoir Simulation Symposium*, The Woodlands, Texas, U.S.A, 2-4 February.

- Jansen, J.D., 2011. Adjoint-Based Optimization of Multi-Phase Flow Through Porous Media – A Review. *Computers & Fluids* **46**(1) 40-51. DOI: 10.1016/j.compfluid.2010.09.039.
- Jansen, J.D., 2013. : *A systems description of flow through porous media*. SpringerBriefs in Earth Sciences, Springer. DOI: 10.1007/978-3-319-00260-6.
- Jansen, J.D., Fonseca, R.M., Kahrobaei, S.S., Siraj, M.M., Van Essen, G.M. and Van den Hof, P.M.J. 2013. The “Egg Model”. Research note and data set. Delft University of Technology, The Netherlands. [http://repository.tudelft.nl/view/ir/uuid:1b85ee17-3e58-4fa4-be798328945a4491/\(report\)](http://repository.tudelft.nl/view/ir/uuid:1b85ee17-3e58-4fa4-be798328945a4491/(report));
- Jansen, J.D., Fonseca, R.M., Kahrobaei, S.S., Siraj, M.M., Van Essen, G.M. and Van den Hof, P.M.J. 2014. The Egg Model- A geological ensemble for reservoir simulation. *Geoscience Data Journal* **(1)** 192-195.
- JOA Jewel Suite 2006 Training Manual, version 2.2, August 2007, JOA Oil & Gas B.V., Delft, Netherlands.
- Kraaijevanger, J. F. B. M., Egberts, P. J. P., Valstar, J. R. and Buurman, H. W. 2007. Optimal Waterflood Design Using the Adjoint Method. Paper SPE 105764 presented at the *SPE Reservoir Simulation Symposium*, Houston, USA, 26-28 February. DOI: 10.2118/105764-ms.
- Leeuwenburgh, O., Egberts, P.J.P. and Abbink, O.A., 2010. Ensemble Methods for Reservoir Life-Cycle Optimization and Well Placement. Paper SPE-136916-MS presented at the *SPE/DGS Saudi Arabia Section Technical Symposium and Exhibition*, Al-Khobar, Saudi Arabia, 4-7 April 2010. DOI: 10.2118/136916-ms.
- Leeuwenburgh, O., Egberts, P.J.P. and Chitu, A.G., 2015. An Ensemble-based Method for Constrained Reservoir Life-Cycle Optimization. Paper SPE-174318-MS presented at the *EUROPEC 2015*, Madrid, Spain, 1-4 June.
- Li, L., Jafarpour, B. and Mohammad-Khaninezhad, M.R. 2012. A simultaneous perturbation stochastic approximation algorithm for coupled well placement and control optimization under geologic uncertainty. *Computational Geosciences* **17**(1) 167-188.
- Lien, M., Brouwer, D.R., Manseth, T. and Jansen, J.D., 2008: Multiscale regularization of flooding optimization for smart field management. *SPE Journal* **13** (2) 195-204. DOI: 10.2118/99728-PA.
- Liu, X. and Reynolds, A.C. 2014. Gradient based multi-objective optimization with applications to water flooding optimization. *Proc. 14th European Conference on the Mathematics of Oil Recovery (ECMOR XIV)*, Catania, Italy, 8-11 September.
- Liu, X. and Reynolds, A.C. 2015. Multi-objective Optimization for Maximizing Expectation and Minimizing Uncertainty or Risk with Application to Optimal Well Control. Paper SPE 173216-MS presented at the *SPE Reservoir Simulation Symposium*, Houston, Texas, U.S.A, 23-25 February.
- Lorentzen, R.J., Berg, A., Naevdal, G. and Vefring, E.H. 2006. A New Approach for Dynamic Optimization of Waterflooding Problems. Paper SPE-99690-MS presented at the *Intelligent Energy Conference and Exhibition*, Amsterdam, The Netherlands, 11-13 April. DOI: 10.2118/99690-MS.
- Luenberger, D.G. and Ye, Y., 2010: *Linear and nonlinear programming 3rd ed.* Springer, New York.

- Marler, R.T. and Arora, J.S. 2004. Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization* **26** (6): 369-395. DOI: 10.1007/s00158-003-0368-6.
- Namdar Zanganeh, M., Kraaijevanger, J.F.B.M., Buurman, H.W., Jansen, J.D., Rossen, W.R. 2014. Challenges in adjoint-based optimization of a foam EOR process. *Computational Geosciences*. **18** (3-4) 563–577. DOI: 10.1007/s10596-014-9412-4.
- Nocedal, J. and Wright, S.J., 2006: *Numerical optimization*, 2nd ed., Springer, New York
- Nwaozo, J. 2006. Dynamic optimization of a waterflooding reservoir. MSc Thesis, University of Oklahoma, USA.
- Oliveira, D.F. and Reynolds, A.C., 2014. An Adaptive Hierarchical Algorithm for Estimation of Optimal Well Controls. *SPE Journal*. **9**(4) 391–402. DOI: 10.2118/163645-PA.
- Oliver, D.S., Reynolds, A.C. and Liu, N. 2008. *Inverse theory for petroleum reservoir characterization and history matching*, Cambridge University Press, Cambridge.
- Oliver, D.S., and Chen, Y. 2011. Recent progress on reservoir history matching: a review. *Computational Geosciences*. **15** (1) 185–221. DOI: 10.1007/s10596-010-9194-2.
- Pajonk, O., Schulze-Riegert, R., Krosche, M., Hassan, M. and Nwakile, M.M., 2011. Ensemble-Based Water Flooding Optimization Applied to Mature Fields. Paper SPE-142621-MS presented at the *SPE Middle East Oil and Gas Show and Conference*, Manama, Bahrain, 25-28 September 2011. DOI: 10.2118/142621-MS.
- Peters, L., Arts, R.J., Brouwer, G.K., Geel, C.R., Cullick, S., Lorentzen, R.J., Chen, Y., Dunlop, K.N.B., Vossepoel, F.C., Xu, R., Sarma, P., Alhuthali, A.H. and Reynolds, A.C., 2010. Results of the Brugge Benchmark Study for Flooding Optimization and History Matching. *SPE Reservoir Engineering and Evaluation* **13** (3) 391-405. DOI: 10.2118/119094-PA.
- Peters, E., Chen, Y., Leeuwenburgh, O. and Oliver, D.S., 2013. Extended Brugge benchmark case for history matching and water flooding optimization. *Computers & Geosciences* **50** 16-24. DOI: 10.1016/j.cageo.2012.07.018.
- Raniolo, S., Dovera, L., Cominelli, A., Callegaro, C. and Masserano, F. 2013. History matching and polymer injection optimization in a mature field using the ensemble Kalman filter. In *Proc. 17th European Symposium Improved Oil Recovery*, St. Petersburg, Russia, 16-18 April.
- Ros, R. and Hansen, N., 2008. A simple modification in CMA-ES achieving linear time and space complexity. *Parallel Problem Solving from Nature—PPSN X*, 296-305.
- Rosenbrock, H. 1960. An Automatic Method for Finding the Greatest or least Value of a Function. *The Computer Journal*. **3** 175-184
- Sarma, P., Aziz, K. and Durlofsky, L.J. 2005. Implementation of adjoint solution for optimal control of smart wells. Paper SPE 92864 presented at *SPE Reservoir Simulation Symposium*, The Woodlands, Texas, U.S.A, 31 January-2 February.
- Sarma, P., Chen, W.H., Durlofsky, L.J., and Aziz, K. 2006. Production optimization with adjoint models under nonlinear control-state path inequality constraints. *SPE Reservoir Engineering and Evaluation* **11** (2): 326–339.

- Sarma, P., and Chen, W. 2014. Improved Estimation of the Stochastic Gradient with Quasi-Monte Carlo Methods. *Proc. 14th European Conference on the Mathematics of Oil Recovery (ECMOR XIV)*, Catania, Italy, 8-11 September.
- Sarma, P., Durlofsky, L.J. and Aziz, K. 2008. Computational techniques for closed-loop reservoir modeling with application to a realistic reservoir. *Petroleum Science and Technology* **26** (10-11) 1120-1140. DOI: 10.1080/10916460701829580.
- Schulze-Riegert, R., Bagheri, M., Krosche, M., Kueck, N. and Ma, D., 2011. Multiple-Objective Optimization Applied to Well Path Design under Geological Uncertainty. Paper SPE-141712-MS presented at the *SPE Reservoir Simulation Symposium*, The Woodlands, Texas, USA, 21-23 February 2011. DOI: 10.2118/141712-MS.
- Shuai, Y., White, C.D., Zhang, H. and Sun, T., 2011. Using multiscale regularization to obtain realistic optimal control strategies. Paper SPE 142043 presented at the *SPE Reservoir Simulation Symposium*, The Woodlands, Texas, USA, 21-23 February.
- Spall, J.C., 1998. Implementation of the simultaneous perturbation algorithm for stochastic optimization. *IEEE Transactions on Aerospace and Electronic Systems* **34** (3) 817-823.
- Strang, G., 2006: *Linear Algebra and its Applications 4th ed.* Thomson Brooks/Cole, Pacific Grove.
- Su, H.-J. and Oliver, D.S., 2010. Smart Well Production Optimization Using An Ensemble-Based Method. *SPE Reservoir Engineering and Evaluation* **13**(6) 884-892. DOI: 10.2118/126072-pa.
- Sudaryanto, B. and Yortsos, Y.C. 2000. Optimization of fluid front dynamics in porous media using rate control. *Physics of Fluids* **12** (7), pp. 1656–1670. DOI: 10.1063/1.870417.
- Suwartadi, E., Krogstad, S. & Foss, B., 2012. Nonlinear output constraints handling for production optimization of oil reservoirs. *Computational Geosciences* **16** (2) 499–517.
- Stordal, A.S., Szklarz, S.P. and Leeuwenburgh, O. 2014. A closer look at ensemble based optimization in reservoir management. *Proc. 14th European Conference on the Mathematics of Oil Recovery (ECMOR XIV)*, Catania, Italy, 8-11 September.
- Tavakoli, R. and Reynolds, A.C., 2010. History matching with a parameterization based on the SVD of a dimensionless sensitivity matrix. *SPE Journal* **15**(12), 495-508.
- Van Doren, J.F.M., Markovinović, R. and Jansen, J.D., 2006: Reduced-order optimal control of waterflooding using POD. *Computational Geosciences* **10** (1) 137-158. DOI: 10.1007/s10596-005-9014-2.
- Van Essen, G.M., Zandvliet, M.J., Van den Hof, P.M.J., Bosgra, O.H. and Jansen, J.D. 2009. Robust waterflooding optimization of multiple geological scenarios. *SPE Journal* **14**(1), 202-210.
- Van Essen, G.M., Jansen, J.D., Brouwer, R., Douma, S.G., Zandvliet, M.J., Rollett, K.I. and Harris, D. 2010. Optimization of smart wells in the St. Joseph field. *SPE Reservoir Engineering and Evaluation* **13**(4), 588-595. DOI: 10.2118/123563-PA.
- Van Essen, G.M., Van den Hof, P.M.J and Jansen, J.D. 2011. Hierarchical long and short-term production optimization. *SPE Journal* **16**(1), 191-199.

- Van Essen, G.M., Van den Hof, P.M.J and Jansen, J.D. 2010. Determination of lower and upper bounds of predicted production from history-matched models. *Proc. 12th European Conference on the Mathematics of Oil Recovery ECMOR XII*, Oxford UK, 6-9 September 2010.
- Volkov, O. and Voskov, D.V. 2013. Advanced strategies of forward simulation for adjoint based optimization. Paper SPE 163592-MS presented at the *SPE Reservoir Simulation Symposium*, The Woodlands, Texas, USA, 18-20 February.
- Wang, C., Li, G., and Reynolds, A.C., 2009. Production Optimization in Closed Loop Reservoir Management. *SPE Journal* **14**(3), 506-523.
- Yan, X. and Reynolds, A.C., 2014. Optimization algorithms based on combining FD approximations and stochastic gradients compared with methods based only on a stochastic gradients. *SPE Journal* **19**(5), 873-890.
- Yang, C., Card, C., Nghiem, L.X. and Fedutenko, E. [2011] Robust Optimization of SAGD Operations under Geological Uncertainties. Paper SPE-141676-MS presented at the *SPE Reservoir Simulation Symposium*, The Woodlands, USA, 21-23 February.
- Yasari, E., Pishvaie, M.R., Khorasheh, F., Salahshoor, K., Kharrat, R. [2013] Application of multi-criterion robust optimization in water-flooding of oil reservoir. *Journal of Petroleum Science and Engineering* **109** 1–11.
- Yeten, B., Durlafsky, L.J. and Aziz, K. 2003. Optimization of nonconventional well type, location and trajectory. *SPE Journal* **8** (3) 200-210. DOI: 10.2118/86880-PA.
- Zandvliet, M.J., Bosgra, O.H., Van den Hof, P.M.J., Jansen, J.D. and Kraaijevanger, J.F.B.M., 2007. Bang-bang control and singular arcs in reservoir flooding. *Journal of Petroleum Science and Engineering* **58** 186-200. DOI: 10.1016/j.petrol.2006.12.008.
- Zhao, H., Chen, C., Do, S.T., Oliveira, D., Li, G. and Reynolds, A.C., 2013. Maximization of a Dynamic Quadratic Interpolation Model for Production Optimization. *SPE Journal* **18**(6) 1012-1025. DOI: 10.2118/141317-PA.

SYMBOLS AND NOTATION

Notation

b	=	discount rate
c	=	learning rate
$\mathbf{c}_{i,j}$	=	ensemble cross-covariance vector
$\tilde{\mathbf{C}}$	=	constant distribution covariance matrix
$\tilde{\mathbf{C}}_{uu}$	=	updated distribution covariance matrix
\mathbf{C}_{uu}	=	ensemble covariance matrix
\mathbf{g}	=	gradient vector
\mathbf{g}'	=	single-smoothed gradient vector
\mathbf{g}''	=	double smoothed gradient vector with ensemble covariance matrix \mathbf{C}_{uu}
\mathbf{g}'''	=	double smoothed gradient vector with distribution covariance matrix $\tilde{\mathbf{C}}$
\mathbf{g}''''	=	single smoothed gradient vector, using $\tilde{\mathbf{C}}$ and a straight gradient vector
\mathbf{g}'''''	=	double smoothed gradient vector, using $\tilde{\mathbf{C}}$ and a straight gradient vector
ℓ	=	iteration counter
\mathbf{H}	=	Hessian matrix
\mathbf{j}	=	vector of mean-shifted objective function values
J	=	objective function value
\bar{J}	=	mean objective function value
k	=	time step counter
K	=	total number of time steps
M	=	number of ensemble members
N	=	number of control variables
\mathbf{p}	=	evolution path vector
q	=	flow rates
r	=	price per unit volume
t	=	time
\mathbf{u}	=	vector of control variables
$\bar{\mathbf{u}}$	=	ensemble mean
$\tilde{\mathbf{u}}$	=	distribution mean
\mathbf{U}	=	matrix of ensemble mean-shifted control vectors
$\tilde{\mathbf{U}}$	=	matrix of distribution mean-shifted control vectors
STB	=	stock tank barrels, a volume measure
α	=	step size
σ	=	perturbation size
μ	=	number of 'best' ensemble members
ε	=	maximum allowed decrease in primary objective function value
τ	=	reference time for discounting

Subscripts

o	=	oil	wp	=	produced water
w	=	water	wi	=	injected water

LIST OF PUBLICATIONS

Journal Publications

- Fonseca, R.M., Leeuwenburgh, O. Van den Hof, P.M.J. and Jansen, J.D. 2014. Ensemble based Hierarchical Multi-Objective Production Optimization of Smart Wells. *Computational Geosciences* **18** (3-4), 449-461
- Fonseca, R.M., Leeuwenburgh, O. Van den Hof, P.M.J. and Jansen, J.D. 2015. Improving the ensemble optimization method through Covariance Matrix Adaptation (CMA-EnOpt). *SPE Journal* **20** (1), 155-168
- Fonseca, R.M., Kahrobaei, S., Van Gastel, L.J.T., Leeuwenburgh, O., and Jansen, J.D. 2015. Quantification of the impact of ensemble size on the quality of an ensemble gradient using principles of hypothesis testing. Under review @ *SPE Journal*
- Fonseca, R.M., Leeuwenburgh, O., Della Rossa, E., Van den Hof, P.M.J. and Jansen, J.D. 2015. Ensemble-based multi-objective optimization of on-off control devices under geological uncertainty. Accepted for publication in *SPE Reservoir Engineering & Evaluation*.
- Jansen, J.D., Fonseca, R.M., Kahrobaei, S.S., Siraj, M.M., Van Essen, G.M. and Van den Hof, P.M.J. 2014. The Egg Model- A geological ensemble for reservoir simulation. *Geoscience Data Journal* **(1)** 192-195.
- Fonseca, R.M., Chen, B., Jansen, J.D., and Reynolds, A.C., 2015. Theoretical understanding of an approximate ensemble-based gradient for optimization under geological uncertainty. *Submitted to International Journal of Numerical Methods in Engineering*
- Fonseca, R.M., Reynolds, A.C. and Jansen, J.D. 2015. Generation of a Pareto front for a bi-objective water flooding optimization problem using approximate ensemble gradients. *Submitted to Journal of Petroleum Science and Engineering*

Conference Proceedings

- Fonseca, R.M., Leeuwenburgh, O. Van den Hof, P.M.J. and Jansen, J.D. 2012. Ensemble based multi-objective production optimization of smart wells. Proc. 13th European Conference on the Mathematics of Oil Recovery (ECMOR XIII), Biarritz, France, 8-11 September
- Fonseca, R.M., Leeuwenburgh, O. Van den Hof, P.M.J. and Jansen, J.D. 2013. Improving the Ensemble Optimization Method through Covariance Matrix Adaptation (CMA-EnOpt). Paper SPE 163657

presented at SPE Reservoir Simulation Symposium, The Woodlands, Texas, U.S.A, 18-20 February.

Fonseca, R.M., Stordal, A.S., Leeuwenburgh, O. Van den Hof, P.M.J. and Jansen, J.D. 2014. Robust ensemble-based multi-objective optimization. Proc. 14th European Conference on the Mathematics of Oil Recovery (ECMOR XIV), Catania, Italy, 8-11 September.

Fonseca, R.M., Kahrobaei, S., Van Gastel, L.J.T., Leeuwenburgh, O., and Jansen, J.D. 2015. Quantification of the Impact of Ensemble Size on the Quality of an Ensemble Gradient Using Principles of Hypothesis Testing, Paper 173236-MS presented at the 2015 SPE Reservoir Simulation Symposium, 22-25 February, Houston, USA.

Fonseca, R.M., Leeuwenburgh, O., Della Rossa, E., Van den Hof, P.M.J. and Jansen, J.D. 2015. Ensemble-based multi-objective optimization of on-off control devices under geological uncertainty. SPE 173268-MS presented at the 2015 SPE Reservoir Simulation Symposium, 22-25 February, Houston, USA.

SUMMARY

Rapid industrialization combined with quest for higher standards of living especially in developing countries has fuelled an expeditious increase in worldwide energy demand. This energy demand is primarily met with energy from fossil fuels such as hydrocarbons located in the earth's subsurface. Of the total estimated hydrocarbon volumes present, currently on average, approximately only 30% of the volumes have/are being produced. Thus primarily due to technological limitations and sometimes economical limitations we leave behind more hydrocarbons than we produce. Some of the technological limitations were born out of a lack of knowledge about the hydrocarbon reservoir being developed which led to sub-optimal reservoir management strategies being used to produce the hydrocarbons. To counter this, the concept of Closed Loop Reservoir Management was introduced by Jansen et al. (2005) who proposed changing reservoir management from the traditional batch-type reactive process to a near-continuous proactive procedure using computer assisted history matching and optimization methods. This thesis focusses on the theoretical understanding to improve the efficiency of a particular optimization method, Ensemble Optimization (EnOpt), for problems incorporating geological uncertainties.

Chapter 2 provides an overview of the theory for Ensemble Optimization (EnOpt) as introduced by Chen et al. (2009). EnOpt uses a simulator, used to model the physical processes occurring in the subsurface, as a black box and thus does not require access to commercial simulator source code to use theoretically rigorous formal gradient based optimization techniques such as the adjoint method. A distinct advantage of EnOpt is the flexibility provided to different control types (rates, pressures, ICD settings, drilling order priorities, switching times) compared to the adjoint method. The stochastic nature of the method is a drawback due to the necessary use of a higher number of reservoir simulations needed to approximate the gradient compared to the adjoint method. Chen (2008) provided an argumentation which enabled EnOpt to be computationally as efficient compared to the adjoint for optimization problems under uncertainty. Chapter 2 provides theoretical insights and proof for a new modified gradient formulation for problems incorporating uncertainties. In this thesis we show that the new modified gradient formulation is not only theoretically more 'robust' compared to the original gradient formulation

proposed by Chen (2008) but also retains the same computational efficiency of the original formulation.

Chapter 3 provides a numerical experimental analysis into the impact of ensemble size on the quality of the stochastic ensemble gradient. This chapter investigates the impact of gradient quality on an optimization experiment and illustrates the importance to achieve high quality gradient estimates for different optimization problems. The results show that ensemble size has a big impact on gradient quality and behaves differently depending on the region within the objective function space. In this thesis we illustrate that the modified gradient formulation estimates a significantly better quality gradient compared to the original formulation and in terms of accuracy, is potentially comparable to the adjoint method in an optimization context.

Chapter 4 investigates a second factor which influences the quality of an ensemble gradient, the covariance matrix. The covariance matrix is used to generate the perturbations used to estimate the gradient was kept constant throughout the optimization in the original EnOpt formulation. In this thesis we have proposed the iterative updating of the covariance matrix inspired from an evolutionary strategy, CMA-ES, which leads to an improvement in gradient quality. This new algorithm is called CMA-EnOpt which highlights the benefit of ‘learning’ about the objective function search space during the optimization. CMA-EnOpt improves the robustness of the optimization to the initial choice of the covariance matrix and has been shown to be useful on multiple synthetic test cases including the Brugge model.

Chapter 5 delves into the applicability of EnOpt to solve bi-objective optimization problems using a hierarchical structure for problems with geological uncertainty to find a single optimal strategy which incorporates multiple objectives. There is no guarantee that there exists no other strategy better than this strategy. Thus a decision maker will require multiple strategies to make an informed decision. **Chapter 6** illustrates the applicability of EnOpt to generate a Pareto front for bi-objective optimization using constrained optimization techniques. The results show that EnOpt was successfully able to generate a Pareto front within acceptable computational effort.

Chapter 7 illustrates the applicability of EnOpt to achieve bi-objective optimization under geological uncertainty on an ensemble of realistic reservoir models based on a real field case. Some of the improvements proposed in this thesis were tested on this realistic model and an efficient parameterization of discrete ICD settings as controls to a continuous domain along with the flexibility of EnOpt to different control types was especially useful for this application.

The results from this thesis illustrates that ensemble based optimization workflows can be used to produce highly computationally efficient ‘robust’ reservoir management strategies of significant practical value.

SAMENVATTING

Snelle industrialisatie en een drang naar hogere levensstandaarden in ontwikkelende landen veroorzaken een toename in de wereldwijde energievraag. Aan deze energievraag wordt primair voldaan met behulp van fossiele brandstoffen zoals koolwaterstoffen die in de ondergrond zijn opgeslagen. Als gevolg van technische-, en in sommige gevallen economische beperkingen wordt momenteel maar ongeveer 30% van het totaal aan aanwezige koolwaterstoffen geproduceerd. Dit lage percentage is het gevolg van een tekort aan kennis van de te ontwikkelen koolwaterstofreservoirs dat leidt tot een niet-optimale strategie om de olie- en gasreserves te produceren. Om dit tegen te gaan is het concept van ‘closed-loop reservoir management’ geïntroduceerd door Jansen et al. (2005), die voorstelden om het traditionele reactieve management van reservoirs, gebaseerd op afzonderlijke processen, te veranderen in een bijna continue proactieve procedure, gebruikmakend van door computers ondersteunde methoden voor het conditioneren van reservoir modellen aan historische metingen en voor optimalisatie. Dit proefschrift richt zich op het verbeteren van het theoretische begrip en de efficiëntie van een optimalisatie methode, ‘Ensemble Optimization (EnOpt)’, in problemen waarin geologische onzekerheid een rol speelt.

Hoofdstuk 2 geeft een overzicht van de theorie voor Ensemble Optimization (EnOpt) zoals geïntroduceerd door Chen et al. (2009). EnOpt gebruikt een simulator om het fysische proces, zoals dat plaatsvindt in de ondergrond, te modelleren als een zogenaamde ‘black box’. Het vereist derhalve geen toegang tot commerciële simulatorcode om theoretisch rigoureuze, formele, op gradiënten gebaseerde optimalisatiemethoden te kunnen gebruiken, zoals de veelgebruikte ‘adjoint’ methode. Een onderscheidend voordeel van op stochastische gradiënten gebaseerde methoden ten opzichte van de adjoint methode is de flexibiliteit om verschillende soorten aansturingsvariabelen (debiets, drukken, ICD instellingen, prioriteiten van boorvolgorde, wisseltijden) te gebruiken. Het stochastische karakter van de methode is een nadeel vanwege het grote aantal simulaties dat vereist is om een gradiënt te benaderen ten opzichte van de adjoint methode. Hoofdstuk 2 levert theoretische inzichten en bewijs voor een nieuwe, aangepaste formulering voor de gradiënt bij problemen met onzekerheden. Dit proefschrift toont aan dat deze nieuwe formulering

niet alleen theoretisch robuuster is dan de oorspronkelijke formulering voor de gradiënt, voorgesteld door Chen (2008), maar ook efficiënter is in de berekening.

Hoofdstuk 3 levert een numerieke analyse van de invloed van de ensemblegrootte op de kwaliteit van de stochastische ensemble gradiënt. Dit hoofdstuk onderzoekt de invloed van de kwaliteit van de gradiënt op een optimalisatie-experiment en illustreert het belang van het bereiken van schattingen van de gradiënt met een hoge kwaliteit voor verschillende optimalisatieproblemen. De resultaten laten zien dat ensemblegrootte een grote invloed heeft op de kwaliteit van de gradiënt, afhankelijk van de regio in de doelfunctieruimte. Dit proefschrift laat zien dat de aangepaste formulering voor de gradiënt een gradiënt schat met een significant betere kwaliteit dan de oorspronkelijke formulering. Het laat ook zien dat, in termen van nauwkeurigheid, de aangepaste formulering voor de gradiënt in potentie vergelijkbaar is met de adjoint methode in een optimalisatie context.

Hoofdstuk 4 onderzoekt een tweede factor die de kwaliteit beïnvloed, namelijk de covariantiematrix. In de oorspronkelijke formulering van EnOpt wordt de covariantiematrix, die wordt gebruikt om de perturbaties te genereren, constant gehouden gedurende de optimalisatie. In deze dissertatie wordt een verbetering van de kwaliteit van de gradiënt voorgesteld door middel van een iteratieve aanpassing van de covariantiematrix, geïnspireerd door een evolutionaire strategie: CMA-ES. Het CMA-EnOpt algoritme verbetert de robuustheid van de optimalisatie voor de beginkeuze van de covariantiematrix. Het nut daarvan wordt aangetoond in meerdere synthetische testgevallen, waaronder het Brugge model.

Hoofdstuk 5 diept de toepasbaarheid uit van EnOpt voor het oplossen van problemen met tweevoudige doelfuncties, gebruikmakend van een hiërarchische structuur, in gevallen waar rekening wordt gehouden met geologische onzekerheid, om een enkele optimale strategie te vinden. In werkelijkheid heeft een beslissingsmaker meerdere strategieën nodig om een goed geïnformeerde beslissing te kunnen maken.

Hoofdstuk 6 illustreert de toepasbaarheid van EnOpt om een Pareto-grens te genereren voor optimalisatie van tweevoudige doelfuncties, gebruikmakend van beperkte optimalisatietechnieken. De resultaten laten zien dat EnOpt met succes een Pareto-grens kan genereren met een acceptabele berekeningsinspanning.

Hoofdstuk 7 schetst de toepassing van EnOpt op een ensemble van realistische reservoirsectormodellen, gebaseerd op een bestaand veld, met als doel het optimaliseren van twee doelfuncties in aanwezigheid van onzekerheid. Enkele van de in dit proefschrift voorgestelde verbeteringen zijn getest op dit realistische model. Een efficiënte parametrisering van discrete ICD instellingen intermen van continue aansturingsvariabelen tezamen met de flexibiliteit van EnOpt bij verschillende aansturingstypen was in het bijzonder bruikbaar bij deze toepassing.

De resultaten in dit proefschrift laten zien dat op ensembles gebaseerde werkprocessen voor optimalisatie gebruikt kunnen worden om strategieën voor reservoiraansturing te genereren die efficiënt zijn in termen van rekentijd, robuust in de aanwezigheid van onzekerheid en praktisch van significante waarde.

ACKNOWLEDGEMENTS

*“Feeling gratitude and not expressing it is like wrapping a gift and not giving it”
- William Arthur Ward*

Four years later, after having run over a million reservoir simulations combined with the roller coaster ride every PhD student experiences, it feels rather surreal to write these final few words as gratitude for this thesis. This thesis would have never been possible without the guidance, support, encouragement and friendship of very many people who for some strange reason choose to believe in me and my abilities. It is thus befitting, but also belittling, for me to attempt to express my gratitude to everyone who has been part of my journey in the windy plains of the Netherlands.

I would like to thank **Prof. Jan-Dirk Jansen** for giving me the opportunity to pursue a PhD and thereafter his valuable supervision, expert guidance, motivation and constant support during the past 4 years. Prof. Jansen, I truly appreciate your constant guidance in shaping me to become more professional (in research and life), pushing me to look deeper into subjects, write meaningful scientific papers and make presentations in front of large audiences. Thank you for your patience and giving me the freedom as well as encouragement to pursue many other extra-curricular activities outside of the PhD, be your teaching assistant as well as sending me to Tulsa for 3 months. Thank you for purchasing for the project a cluster to perform the simulations without which most of the work would not have been completed.

I would like to thank **Prof. Paul Van den Hof** for his valuable supervision, guidance and encouragement during the past 4 years. I learnt many concepts about the systems and control world from valuable discussions during the monthly meetings. Thank you for encouraging me to constantly take a step back and think about the larger research questions being investigated which helped significantly to formulate a structured research plan.

I would also like to thank **Dr. Olwijn Leeuwenburgh** for his guidance, advice and initiating me into the world of Ensemble Optimization during my MSc thesis at TNO. Olwijn, thanks for constantly pushing me to think about different aspects about my

research and patiently listening to my stupid ideas. In spite of your busy schedule you always found time to answer my questions and to be a sounding board about the thesis and challenges I encountered. Thank you very much for translating the summary of this thesis at very short notice.

I would also like to acknowledge **ENI** who were the main sponsors for my research. Ernesto, Alberto, Stefano, Laura, Riccardo and all the many others with whom I have interacted with at ENI many thanks for all the interesting work related discussions, support, guidance as well as the social events in Milano. I always looked forward to my trips to Milano which were highly productive and thoroughly enjoyable.

I would like to thank **Prof. Albert Reynolds** for giving me the opportunity to spend the summer of 2014 under his guidance at the University of Tulsa. Dr. Reynolds thank you for spending so many long afternoons discussing various aspects of ensemble optimization which led to many interesting discussions on life in general which has been a massive learning experience for me.

I gratefully acknowledge the support and opportunities **Dr. Remus Hanea** has given me during these past four years. Remus your infectious personality is simply amazing and thanks for initiating the process for me to spend time in Tulsa with Prof. Reynolds as well as now at my current position in TNO. The opportunity to teach at your summer school in Romania and the experiences there is something that will always stay with me. Thanks for being my paranymph at the defense.

To my doctoral **defence committee** members Prof. Heemink, Prof. Weiland, Prof. Macbeth, Prof. Rossen and Prof. Bruining, thank you for making the time to be on the committee, reviewing and providing valuable feedback about the thesis.

A special thank you to **Dr. Susanne Rudolph** who convinced me to do my MSc thesis with Prof. Jansen inspite of me considering an experimental based thesis under her supervision. If I had not listened to her advice I would probably never have written this thesis or gotten a chance to pursue this PhD.

Most of the results presented in my thesis and publications would not have been possible without ECLIPSE licenses. I would like to acknowledge **Schlumberger** for providing us at TU Delft a donation of academic licenses which helped me significantly speed-up the pace of my work. In particular, I would like to acknowledge Prof. Stefan Luthi and Alexander van Noort for helping us acquire the donation for the multiple academic licenses.

Dr. Willem Schulte, Prof. Bill Rossen and the **FDP team** for giving me the opportunity to take my mind away from my thesis for 3 weeks every year. Those 3 weeks helped me sharpen my reservoir engineering skills and was a lot of fun to help the students iterate towards some interesting innovative field development plans.

A big thank you to all my professors/lecturers at **TU Delft** for all their efforts and guidance during my MSc degree which continued during the past 4 years as a teaching assistant. To all the MSc. Students who undertook their final research projects in collaboration with me, Luc, Danny, Marcel, Arnaud, Chia. To Dr. Wolf, Marc, Karel and all the staff of the Dietz laboratory for making the working atmosphere so enjoyable. A special thank you to the engine room of the Geoscience and Engineering department, Anke, Marlijn, Lydia, Margot, Ralf and Hannie for always helping with every bureaucratic issue possible and always going out of their way to make it easier for us to work in an fun enjoyable atmosphere.

To all my professors/lecturers at **MIT Pune**, and **St. Vincent's High School** who have molded my thought process and positively influenced my career path. To all **MSc. classmates** at TU Delft for the wonderful initial 2 years of my journey in The Netherlands, the fun trips and friendships I will always cherish. To all my classmates and friends from the bachelor program at MIT Pune. The bonds forged during our initial forays into the world of petroleum engineering will always be cherished.

During my time in Tulsa I was fortunate to meet and make friends with some very special people, Konark, Amarinder, Soham, Ashwin, Jayant and everyone in Tulsa who made my stay there so comfortable and welcoming and Dr. Fahim Forouzanfar for all the interesting discussions about our research and the scientific community.

During these 4 years I was fortunate to serve on the SPE Delft Student chapter Board, MV PhD Committee and World Petroleum Council European Young Professionals committee. Working in these varied groups, organising many events and trips has been a steep though thoroughly enjoyable learning curve. To all my fellow board and committee members who I had the pleasure to serve with, Thank You !

To my new colleagues at **TNO**; Alin and Joanna who made travelling to Utrecht so much more fun, Frank for all the mentoring session and initiating me into the TNO way, Lilya, Farid, Dries, Yanhui, Paul, Lies, Daniel, Cor, Rohith, Chris, Ymke, Richard and many others; for making the past few months a smooth and very enjoyable transition for me from university to the real world.

To the many wonderful friends from the EWI department at TU Delft, Andreas and Slavek for all our research collaborations and insightful discussions as well as initiating me into the world of beer, Mohit and Shruti for the great dinners and debates as well as the occasional tennis game, Tudor for making conferences such fun especially the trips in the US, never would I have thought that mathematicians can be so much fun.

To all my friends at the petroleum engineering department whose friendship and support is one of the prime reasons for the completion of this thesis. Siavash, my first friend in Delft who has been a constant stream of advice support and motivation,

Durgesh, my closest Indian friend in Delft who kept me grounded, was my punching bag when needed and constantly reminded me of my Indian roots, Nikita for the super fun times and trying your best to make me dance Salsa, Christian for the terrific fishing evenings, riebekuchen and German beer, Rafael for all the support and honesty especially during the writing of this thesis, Eduardo for the long discussions about the results and life in the Netherlands and Roozbeh for all the encouragement and support during the tough times especially while setting up the SPE board. The unforgettable Tuesday dinners, barbeques, ice-skating etc. were a result of a great bunch of friends each of whom hold a special place in my life; Saskia, Jiakun, Daniel, Matteo, Ehsan, Elisa, Grigori, Negar, Jakolien, Leon, Amin, Maryam, Anna, Amin F., Matei, Mojtaba, Tianqi, Jinyu, Bander, Rodrigo, Ahmed, Mohammed and all the many others who made these past 4 years memorable.

To **Fr. Avin** for his encouragement, support and friendship during the past 6 years as well the extended ISC community which is truly a home away from home.

To my two of my closest friends in delft, the Kardashians of my scientific world, Koen and Kevin ! **Koen** many thanks for all the translations in this thesis and the numerous phone calls to the Belastingdienst. **Kevin**, the cover would not have been possible without you. More importantly though thanks to you both for the support, countless translations and the amazing evenings in Delft as well as visiting us in India which was truly memorable.

To my best friends from childhood Hasan, Neeraj, Rajeev, Taronish, Jonathan, Mikka and the entire cricket team at Park View, many thanks for your friendship which continues until today and hopefully forever. To my extended family, all my aunts, uncles and cousins for all their love, affection and generosity through the years.

I would like to express my deepest and most sincere gratitude and thanks to my parents and sister for their unconditional love and support, most of all for always encouraging me to do my M.Sc. and PhD at Delft University of Technology. They have always been my greatest sources of inspiration without whose constant help and support I would have never ever been able to undertaken this journey. Their everlasting patience and belief in my abilities have encouraged me always. I truly appreciate their efforts and contributions since I've been born and will always remain indebted to kindness and generosity !

Last but not least to the “Almighty” for the wonderful journey thus far.....

Rahul-Mark Fonseca,
Delft, December 2015

