ISBN 978-82-326-0482-1 (printed version)
ISBN 978-82-326-0483-8 (electronic version)
ISSN 1503-8181

Doctoral theses at NTNU, 2014:285

Brage Rugstad Knudsen

On Shut-In Based Production Optimization of Shale-Gas Systems

NTNU – Trondheim
Norwegian University of
Science and Technology

Doctoral theses at NTNU, 2014:285

NTNU

NTNU
Norwegian University of
Science and Technology
Faculty of Information Technology,
Mathematics and Electrical Engineering
Department of Engineering Cybernetics

NTNU – Trondheim
Norwegian University of
Science and Technology

Brage Rugstad Knudsen

On Shut-In Based Production Optimization of Shale-Gas Systems

Thesis for the degree of Philosophiae Doctor

Trondheim, October 2014

Norwegian University of Science and Technology
Faculty of Information Technology,
Mathematics and Electrical Engineering
Department of Engineering Cybernetics

NTNU – Trondheim
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Information Technology,
Mathematics and Electrical Engineering
Department of Engineering Cybernetics

ISBN 978-82-326-0482-1 (printed version)
ISBN 978-82-326-0483-8 (electronic version)
ISSN 1503-8181
ITK Report 2014-5-W

Doctoral theses at NTNU, 2014:285

Printed by Skipnes Kommunikasjon as

To Sigrid

Summary

The use of shale gas as an energy resource has increased extensively over the last decade, with emerging new technologies causing shale gas to become the largest source of growth in the US natural-gas supply. Efficient utilization of shale-gas resources is, however, very challenging, with wells being inherently dynamic and fields typically consisting of hundreds of geographically distributed wells and production pads. The characteristic system of low-permeable matrix blocks and interconnecting fracture networks makes shale-gas wells particularly suited for intermittent shut-in based production schemes, a property which can be utilized both as a means of preventing liquid loading and for meeting varying gas demands and prices. This thesis concerns the development of efficient optimization schemes and shale-gas well- and systems models for scheduling of well shut-ins in large-scale shale-gas systems. The thesis is divided into four parts.

The first part describes the development of novel shale well and reservoir proxy models, together with mixed integer models for computation of shut-in times for shale multi-well pads. The scheme applies short shut-ins in order to prevent liquid loading in late-life wells, and tracks an assigned reference rate for a set of wells producing at a shared pad. Following a spatial and temporal discretization of the PDE governing the reservoir proxy model, we use logic-based generalized disjunctive programming (GDP) as a basis for modeling well shut-ins and constraints on well switchings. This formulation lends itself both to a complete MILP reformulation and reduced size MINLP reformulations. A computational study shows that the complete MILP formulation retains best computational performance for increasing problem sizes. Moreover, we demonstrate how a structured shut-in schedule prevents the oscillating rates caused by a naive shut-in approach, however, at the expense of a complex shut-in pattern.

The second part of the thesis addresses shut-in scheduling for preventing liquid loading in large, distributed shale *multi-pad* systems. The proposed scheme optimizes shut-in times and a distinct reference rate for each multi-well pad, minimizing deviations from the individual pad rates while ensuring that the total produced rate tracks a given short-term gas demand for the entire field. An extended tuning procedure for the proxy models are presented, using prefiltering of prediction errors as weight-selection in a least-squares formulation for parameter estimation in the proxy models. By using an embedded GDP model for describing the states of the system, we derive an MILP reformulation of the entire shale multi-pad scheduling

problem. The resulting MILP model renders a block-separable structure, facilitating the solution by a Lagrangian relaxation scheme. The resulting decomposition scheme is developed with a trust-region cutting-plane method for solving the Lagrangian dual, and a combined fixing and Local-Branching based heuristic for recovering primal feasible solutions from the Lagrangian.

In the third part of the thesis, the use of shut-ins is further developed as a means of meeting seasonal varying gas demands from natural-gas power plants. To this end, the proposed scheme argues that shale-gas reservoirs can be used to shift storage of gas used for meeting varying demands, from separate underground storage units operated by local distribution companies to the gas producers themselves. This chapter contains slightly modified well and reservoir proxy models with a more accurate nonlinear tubing model. Moreover, a tight convex-hull reformulation of an GDP model is derived and utilized in the scheduling formulation. Model reformulations and SOS2 approximations are applied to render a large-scale MILP, which we solve by a Lagrangian relaxation scheme. Finally, we implement a receding horizon strategy to address operational uncertainties and varying gas spot-price. Illustrative case studies show a significant economic potential for shale-well operators by adopting the proposed scheme, and a viable approach for generating companies to secure a firm gas supply for meeting seasonal varying electric-power demands.

The last chapter of the thesis presents an Objective Feasibility Pump (OFP) designed for finding good feasible solutions in short computation times for difficult convex MINLPs. The algorithmic development uses a multi-objective optimization formulation to systematically design a heuristic that balances the two goals of quickly obtaining a feasible solution and preserving solution quality. A set of user-defined parameters allows for emphasizing either short computation time or solution quality. The efficiency of the proposed heuristic is evaluated by extensive computational testing on a large set of convex MINLP test problems. As the algorithmic framework of the OFP is quite generic, we further demonstrate applicability of the proposed heuristic on nonconvex MINLPs by solving a nonlinear shale-well scheduling problem.

Preface

This thesis is submitted in partial fulfillment of the requirements for the degree of Philosophiae Doctor (PhD) at the Norwegian University of Science and Technology (NTNU). The research has been conducted at the Department of Engineering Cybernetics (ITK) from August 2010 to 2014. Funding for the research has been provided by the Center for Integrated Operations in the Petroleum Industry (IO center) and partly by the department, for which I am very grateful. The financial support from an IBM PhD Fellowship Award is also highly appreciated.

First and foremost I would like express my deep gratitude to my supervisor Professor Bjarne Foss. I feel very fortunate who have had the opportunity to collaborate with you during the course of this thesis, and I am sincerely grateful for all the support, motivation and guidance you have given me. You have managed to perfectly balance structured guidance with the autonomy to pursue my own research ideas, thereby providing an excellent research and working environment.

I am very grateful for the collaboration with my two co-supervisors, Professor Curtis H. Whitson at Department of Petroleum Engineering and Applied Geophysics (IPT) at NTNU, and Dr. Andrew R. Conn at IBM T.J. Watson Research Center in New York. Observing and learning from Curtis' expertise in petroleum and reservoir engineering has been very motivating. I am really thankful for all the hours you have spent with me in developing and understanding shale-gas reservoir models, and for the way you have introduced me to and let me benefit from your network of contacts. A special thanks goes to Andrew for the exciting collaboration we have had during my stays in New York. It was the rewarding internship with you at IBM T.J. Watson in 2009 that initiated my interest in pursuing a PhD. Your insight and long experience from working with numerical optimization have lead to many ideas and fruitful discussions. I would also like to thank you and Barbara for the amazing hospitality you have shown during my visits to New York.

During the second half of 2012 I had the privilege of visiting Professor Ignacio E. Grossmann at Carnegie Mellon University (CMU) in Pittsburgh. I am deeply grateful for the collaboration with you and your research group; it was a great inspiration and an important contribution to my thesis. Thanks are further due to Lars Høier in Statoil for introducing me to their research group on shale gas, to Narayan Nair in Linn Energy, Oklahoma, for hosting me during a field visit in December 2012, to Alexander Juell at IPT for the help you have provided, and to Robert Hubbard in John M. Campbell & Co for value inputs.

I would like to thank the members of my thesis committee, Professor Stratos Pistikopoulos at Imperial College London, Dr. Remus Hanea in Statoil and Pro-

fessor Morten Hovd at ITK, for the efforts you have spent on revising this thesis.

I have been fortunate to supervise some very clever master's students at the department, Espen T. Nordsveen, Shaurya Sharma and Einar Wold. It has been a rewarding and educational experience. I would especially like to acknowledge the collaboration with Shaurya, with whom two papers from this thesis has been written.

During my time here at the department, I have enjoyed good friendship and many interesting discussions with Tor Aksel Heirung, Bjarne Grimstad, Joakim Haugen, Anne Mai Ersdal, Anders Willersrud, Kristian G. Hanssen, Christoph Backi, Silvya D. Rahmawati, Agus I. Hasan, Vidar Gunnerud, Anders Fougner, Espen Hauge and several others; it has been a great time. I also appreciate the great lunch environment with friends from Tindegruppa. Furthermore, I would like to express my gratitude for the warm hospitality and good friendship with fellow PhD students and visitors at CMU, including Sumit Mitra, Linlin Yang, Satyajith Amaran, Vijay Gupta, Francisco Trespalacios, Qi Zhang, German Oliveros and Pedro Mendéz.

My parents, Bente and Frode, deserve a special thanks for the unconditional support you have shown and for teaching me the value of education. Finally, and most important of all, I am endlessly thankful to Sigrid, for all the love, support and encouragement you have given me through these years.

Brage Rugstad Knudsen
Trondheim, August 2014

Contents

Summary v

Preface vii

Contents ix

Abbreviations xiii

1 Introduction 1
- 1.1 Shale gas . . . 1
  - 1.1.1 Shale-gas production and operational challenges . . . 5
  - 1.1.2 Shut-ins of shale-gas wells . . . 8
- 1.2 Simplified systems modeling of shale-gas wells . . . 10
  - 1.2.1 Shale-gas reservoir modeling . . . 10
  - 1.2.2 Proxy and black-box modeling . . . 13
  - 1.2.3 Dynamic proxy modeling of shale-gas reservoir . . . 14
  - 1.2.4 Well and surface system modeling . . . 18
- 1.3 Large-scale mixed integer programming . . . 19
  - 1.3.1 Decomposition by Lagrangian relaxation . . . 21
  - 1.3.2 Solving the Lagrangian dual . . . 23
- 1.4 Research objective and scope . . . 25
- 1.5 Outline and contributions of thesis . . . 26
  - 1.5.1 Publications . . . 28

2 Shut-in Based Production Optimization of Shale-gas Systems 31
- 2.1 Introduction . . . 31
- 2.2 Shale-gas reservoir modeling . . . 35
  - 2.2.1 Well model . . . 38
- 2.3 Fitting parameters of the proxy model . . . 39
  - 2.3.1 Cross-validation and practical considerations . . . 41
  - 2.3.2 Critical rate . . . 43
- 2.4 Formulation of the production optimization problem . . . 44
  - 2.4.1 Reformulation of nonsmooth nonlinearities . . . 48
  - 2.4.2 Exact linear reformulation . . . 50
- 2.5 Computational results . . . 51
- 2.6 Discussion . . . 57

2.7 Conclusions . . . 58

3 Lagrangian Relaxation Based Decomposition for Well Scheduling in Shale-gas Systems 59
3.1 Introduction . . . 59
3.2 Shale well and reservoir proxy modeling . . . 62
3.2.1 Critical gas rate . . . 65
3.3 Computing model parameters . . . 65
3.3.1 Updating model parameters . . . 70
3.4 Formulation of the optimization problem . . . 71
3.5 Lagrangian relaxation scheme . . . 76
3.5.1 Solving the Lagrangian dual . . . 77
3.5.2 Generating primal feasible solutions . . . 80
3.6 Computational Results . . . 83
3.7 Discussion . . . 86
3.8 Conclusions . . . 89
3.A Proof of bound for dual variable . . . 89

4 Shale-gas Scheduling for Natural-gas Supply in Electric Power Production 91
4.1 Introduction . . . 91
4.2 Shale well modeling . . . 96
4.2.1 Shale reservoirs as gas storage . . . 98
4.3 Formulation of the shale-well scheduling problem . . . 100
4.4 Decomposition by Lagrangian relaxation . . . 105
4.4.1 A novel primal recovery heuristic . . . 108
4.5 Receding horizon optimization . . . 110
4.6 Case studies . . . 113
4.6.1 Case 1: Test case with a 43 MW generator and 10 producing wells . . . 115
4.6.2 Case 2: A 60 wells shale-gas field supplying a 300 MW NGPP 118
4.6.3 Case 3: 105 shale-gas wells supplying a 400 MW NGPP . . 120
4.7 Discussion . . . 123
4.8 Conclusion . . . 125
4.A Reformulation of disjunctions to algebraic constraints . . . 125
4.B Piecewise linear approximations . . . 126

5 Towards an Objective Feasibility Pump for Convex MINLPs 129
5.1 Introduction . . . 129
5.1.1 The Feasibility Pump . . . 132
5.1.2 Algorithmic issues . . . 133
5.1.3 The Objective Feasibility Pump for MILPs . . . 133
5.2 An Objective Feasibility Pump for convex MINLPs . . . 134
5.2.1 Rounding . . . 138
5.2.2 Stall and cycle handling with SOS1 constraints . . . 139
5.3 Computational results . . . 141
5.3.1 Results of varying the user defined weight . . . 145

5.4 Example: Shale-well scheduling 147
5.4.1 Results 148
5.5 Concluding remarks 151
5.A Detailed computational results 151

6 Concluding Remarks and Recommendations for Further Work 155

Bibliography 161

A Discretization and Parameter Estimation in Slab-based Proxy Model 179
A.1 Grid selection and prefilters in WNLS 182
A.1.1 Discussion on the tuning scheme 186

Abbreviations

BB	Branch-and-bound
BHP	Bottomhole pressure
CP	Cutting plane
CH	Convex hull
DST	Decision support tool
DWD	Dantzig Wolfe decomposition
EIA	Energy Information Administration
EMPC	Economic model predictive control
EUC	Electric utility companies
FP	Feasibility Pump
GDP	Generalized disjunctive programming
GENCO	Generating companies
HF	Hydraulic fracturing
IBVP	Initial boundary-value problem
IEA	International Energy Agency
IIR	Infinite impulse response
IP	Integer program
LDC	Local distribution company
LNG	Liquified natural gas
LP	Linear program
LR	Lagrangian relaxation
mD	Milli-Darcy
MFR	Multi-fracture reference model
MHE	Moving horizon estimation
MIP	Mixed integer program
MILP	Mixed integer linear program
MINLP	Mixed integer nonlinear program
MPC	Model predictive control
NLP	Nonlinear program
NGPP	Natural-gas power plants
NPV	Net present value
NYMEX	New York Mercantile Exchange
OA	Outer approximation
OFP	Objective Feasibility Pump
ODE	Ordinary differential equation
PDE	Partial differential equation

QP	Quadratic program
RH	Receding horizon
RINS	Relaxation Induced Neighborhood Search
ROM	Reduced-order models
RTO	Real-time optimization
SOS	Special-order sets
SRV	Stimulated reservoir volume
SSE	Sum of squared errors
THP	Tubinghead pressure
UC	Unit commitment
WNLS	Weighted nonlinear least-squares

gc	gas critical
sc	standard conditions

Chapter 1

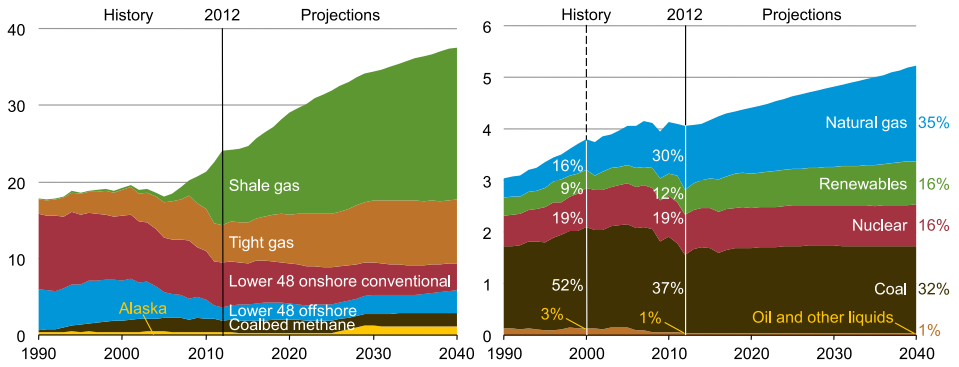
Introduction

This opening chapter presents a brief introduction to shale gas, its characteristics and important aspects of the development of shale-gas resources. We then focus on production related challenges in recovery of shale gas, and motivate the use of shut-ins during production. As the modeling approach applied for performing shut-in scheduling differs from conventional shale-gas modeling, we continue with a review of existing shale well modeling schemes to place our approach into context. We also include a brief introduction to large-scale mixed integer programming and decomposition techniques. At the end of the chapter, the research objective, outline, organization and contributions of the thesis are given.

1.1 Shale gas

Exploitation of unconventional gas resources has reshaped the distribution and understanding of the global natural-gas market. Shale gas is undoubtedly the main source of unconventional gas supply, estimated at 67% of the share of global technical recoverable unconventional gas resources (McGlade et al., 2013), the latter including tight gas sands and coal-bed methane. The remarkable development of the shale-gas industry has its roots in the last decade's shale-gas revolution in the US, driven by continued advances in horizontal drilling and hydraulic fracturing (HF). These technological developments have facilitated economic feasible utilization of vast shale-gas resources that earlier were regarded as uneconomical. Shale-gas currently constitutes more than 40% of the total US natural-gas production, and is expected to increase to 53% by 2040 (EIA, 2014). Fig. 1.1a, displaying a projection of US-natural gas production from different resources as given by the US Energy Information Administration (EIA, 2014), clearly shows the expected dominance of shale gas in the country's natural-gas supply. A consequence of the US shale-gas revolution has been a reduction in the country's dependency on gas imports, thereby increasing security of supply and energy self-sufficiency. In a global context, this shift in the US gas market has altered the global trade connections of conventional natural gas and LNG. Although resource estimates are more uncertain and available gas infrastructure is more scarce, the success of US shale gas has generated significant interest for exploring shale-gas resources other places in

the world, including Argentina, Australia, China and countries in North Africa.



(a) Current and projected US production of shale gas and other natural-gas resources.

(b) US electricity generation by fuel.

Fig. 1.1: Prospects for developments in US shale-gas recovery and electricity generation by fuel. From EIA (2014).

The entry of shale gas in the US has exhibited a substantial impact on both industry, economy, power production and the natural-gas market. On the one hand, shale gas has brought economic growth to the country with increased industrial manufacturing, and as mentioned, enhanced the security of supply for residential, electrical and industrial use. Concerns have though been raised about sustainability of shale-gas recovery, both in terms of environmental footprint of the extraction process and profitability due to low gas prices, creating a comprehensive and diverse debate on viability of the dry-gas shale industry, e.g. Engelder (2011); Hou et al. (2012); Hughes (2013); Jenner and Lamadrid (2013). Another challenge has been the public acceptance of shale gas in some areas, mainly due to concerns about groundwater contamination and the high volumes of water required for hydraulic fracturing stimulation. Water management is consequently an important aspect for sustainable shale-gas recovery, with the industry currently moving towards complete recycling of the flowback fluids (Engelder, 2011; Jenner and Lamadrid, 2013). Natural gas is expected to be a bridge fuel towards increased low-carbon generation and renewable energies (IEA, 2012), with abundant reserves of unconventional gas, and in particular shale gas, feeding the natural-gas value chain. As natural gas produces less than half of the carbon dioxide emissions from coal (Hou et al., 2012), an important part of the expected growth in natural-gas electricity generation, cf. Fig. 1.1b, is due to an expected replacement of coal-fired power plants with high efficiency combined-cycle natural-gas power plants to reduce CO_2 emissions (Macmillan et al., 2013). The shale-gas boom has caused US gas prices to plunge, putting high pressure on costs and profitability of dry-gas shale fields. However, without sufficiently high CO_2 taxes, the competitiveness of natural gas as a substitute for coal in electricity generation relies heavily on a sustained low gas price. As such, with the electric power sector constituting a major end-customer for shale-gas producers, and shale gas being a major feedstock for natural-gas power plants, there is a clear interdependency between these sectors in terms of economic feasibility.

Being an unconventional gas resource, shale gas distinguishes itself from conventional resources in several ways. From a geological point of view, shale-gas basins differ from conventional gas reservoirs since the rock is both the seal preventing gas from emigrating to the surface and the source rock storing the organic rich content at the same time. Gas can be stored in ultra-tight shale formations both as free gas in porous spaces and natural fractures of the shale rock, and as absorbed gas within organic material (Cipolla et al., 2010). With permeabilities ranging from 10^{-6} to 10^{-3} mD[1] (Cipolla et al., 2010; Shelley et al., 2008), the inherent tightness of shales impedes gas to flow freely through the formation, necessitating stimulation with hydraulic fracturing to create conductive channels and fractures transporting gas from the formation and into horizontal wellbores. Depending on the formation characteristics, existence of natural-fracture networks and major stress components, the high-pressurized HF stimulation may both crush the rock close to the wellbore and create longer and more extensive fractures penetrating the formation (Warpinski et al., 2005; Cipolla et al., 2010). In both cases, HF stimulation releases large volumes of gas stored in the stimulated area, causing the characteristic initial peak rate for shale-gas wells. At the end of this initially released gas, the rate declines rapidly, with production of gas now being driven by high pressure gradients between the highly conductive fractures and the low-permeable shale matrix.

Economic development of shale-gas assets contains many steps and involves massive logistics. During the initial phase of field development, the first challenge often met relates to land regulations and lease- contracts and costs. The composition of land ownership and regulations in the area where the wells are to be drilled are decisive for this initial phase. In the US, for example, the land is often privately owned, requiring fees for leasing of acreage, a lease bonus if gas is found and royalties for the gas produced (Kaiser, 2012). To compare, the land in Canada is mainly owned by the government, giving a different set up of leasing contracts, environmental regulations and often high federal taxes on the gas produced. Focusing on US shale gas, a common field-exploration strategy is to divide an area into blocks of fairly equal size, with subsequent drilling of an appraisal well in each acreage. Appraisal wells are often vertical to save drilling costs (Jenkins and Boyer, 2008), and serve the purpose of providing basin information such as gas content and saturations, organic-matter characteristics, permeabilities and formation thickness. The collected information is then used by the operating companies to construct some type of geological model to support well-placement decisions and design of stimulation procedures for wells to be drilled in the area.

Areas containing the most encouraging appraisal wells are further explored, either by extending the appraisal wells with a horizontal section to prepare for hydraulic fracturing, or drilling and completion of one or more pilot wells. This stage of the field development aims at demonstrating that sufficient gas rates can be recovered, in order to decide whether a full field development is economically viable before surface processing equipment is installed and the wells are connected with gathering pipelines to transmission networks.

[1] milli-Darcy is the established field unit for permeability. The SI unit for permeability is m^2, with conversion factor $1mD = 9.87 \times 10^{-16} m^2$ (Whitson and Brule, 2000), but it is rarely used for shale-gas reservoirs.

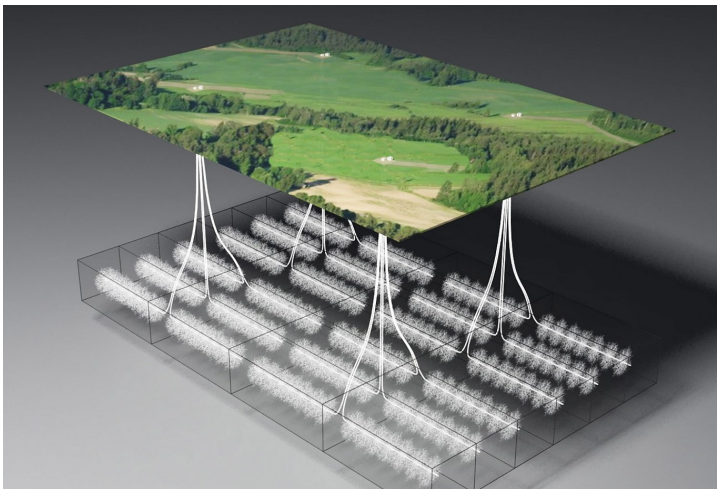


Fig. 1.2: Illustration of shale-gas multi-well pads. Source: Statoil.com.

The progress of the subsequent part of a field development depends highly on the topology of surrounding terrain and landscape, accessibility to a gas transmission network and state regulations. Some wells are completed and produced as standalone wells without sharing surface production equipment, with other wells drilled in the vicinity. There is, however, a clear trend towards extensive use of multi-well pads (Stefik and Paulson, 2011), where multiple wells are drilled from the same location as illustrated in Fig. 1.2. Benefits from using multi-well pads are numerous. By sharing drilling and completion equipment and crew, water pipelines and storage pits for fracturing water, and production equipment such as separators and storage tanks for any co-produced gas-condensates, operators are able to reduce costs substantially. Many places, particularly in developments of the Marcellus formation, significant efforts have been made to minimize the footprint caused by shale-gas extraction. Multi-well pads help reducing the impact by requiring less roads, reduced gathering pipeline networks and less interference of the landscape surrounding the well cites. The number of wells drilled at a multi-well pad ranges from 3-4 wells to more than 20 (Kennedy et al., 2012). Completed wells are connected via gathering pipelines to a shared compression unit, or directly to a transmission pipeline if the wellhead pressure is sufficiently high. Compressor stations may either be operated by midstream companies in the natural-gas value chain, or by the well operator, particularly for larger companies.

Field development of shale-gas assets is a continuous process that lasts throughout the entire life-time of a field. The characteristic steep decline rate of shale-gas wells (Jenkins and Boyer, 2008; Baihly et al., 2010) require operators to continuously drill new wells and explore the organic rich formation to offset decline in overall field productivity. New wells are hence drilled at a high frequency, although some attempts have been made on refracturing of existing wells (King, 2010). The sustained low US gas price resulting from the shale-gas boom has put a high pressure on well economics in dry-gas fields, requiring operators to keep costs at

(a) Storage tanks. **(b)** Water pit.

Fig. 1.3: Example of a surface system with pipelines, valves and storage tanks for co-produced condensates (a), and a water pit for recycling of fracturing and co-produced water (b).

a minimum. Existing surface equipment and infrastructure are hence extensively utilized when new wells are drilled and set to produce, eventually reducing surplus capacity of the initially installed gathering and processing system when the number of wells in a field increases. Well operators end up with fields containing a large number of geographically distributed standalone wells and multi-well pads, thus adding significant challenges in terms of well maintenance, control and production optimization.

1.1.1 Shale-gas production and operational challenges

At the end of hydraulic fracturing completion of shale wells, it is common operational practice to shut in wells for some weeks or months, either immediately after completion or after flow-back for a few days (Cheng, 2012; Fakcharoenphol et al., 2013), causing a delay in selling the gas. Cheng (2012) observed that such extended initial shut-in periods may significantly reduce water rates and increase gas rates, while not affecting long-term gas production. Once the well is flown back, massive volumes of fracturing water must still be captured, cleaned and stored for later usage, although for many wells only 10-25% of the pumped water is recovered due to water imbibition in the shale matrix blocks (Cheng, 2012; Makhanov et al., 2013). The initial gas rate, whether directly after fracturing water is flown back or after an extended shut-in period, tends to be very high and hence important for the well economics. During this phase, the wells are often choked back for several reasons: Too high gas velocities may damage surface equipment, and possibly cause erosion on pipeline walls if gas rates exceed pipelines design-rates. Moreover, the capacity of shared pipelines and separators may be exceeded. Finally, it has been observed

that initial chocking of well flow may lead to improved long-term productivity compared to wells that are produced unrestricted (Okouma et al., 2012).

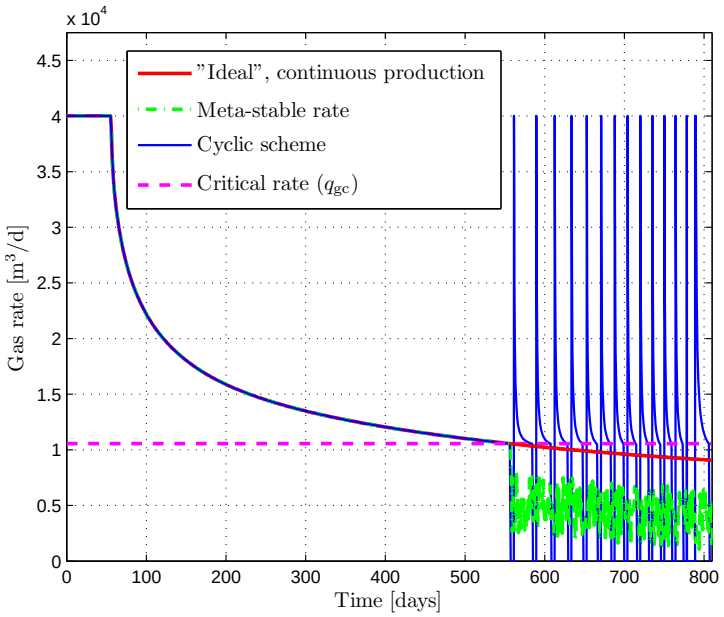


Fig. 1.4: Illustration of an "ideal", though non-realistic, continuous production (red line), the cyclic shut-in scheme with a *fixed* shut-in time (blue line), and an *illustration* of erratic meta-stable production (green line).

The initial peak or plateau rate from shale-gas wells, the latter enabled by initial choking, is followed by a characteristic steep decline and subsequent low, so-called pseudo steady-state rates as illustrated in Fig. 1.4. Shale-gas wells may produce at these pseudo steady-state rates for some hundred days or even for years, depending on the well and reservoir properties. Along with the slow but steady decline in the pseudo steady-state rates, comes one of the foremost operational challenges of shale-gas wells, which is to prevent the state of well *liquid loading* (Redden, 2012; Sutton et al., 2010; Al Ahmadi et al., 2010; Awoleke and Lane, 2011; Lea and Nickens, 2004; Whitson et al., 2012). This state is reached when the pressure support in the well is insufficient to lift co-produced liquids to the surface, causing accumulation of liquids in the wellbore and thereby increased bottomhole hydrostatic backpressure. Accumulated liquids may be due to some low saturation of water in the formation, gas condensates, or left-over water injected during HF stimulation (Whitson et al., 2012). Onset of liquid loading in gas wells can be recognized by highly erratic and unstable gas rates, a sharp drop in the decline curve (Al Ahmadi et al., 2010; Lea and Nickens, 2004) or as semi-stabilization of the production rate at significantly lower levels (Dousi et al., 2006; Whitson et al., 2012), i.e. at so-called meta-stable rates, illustrated by the green erratic rates in Fig. 1.4. Liquid loading severely deteriorates operations of gas wells, possibly causing large liquid slugs that must be captured by the wellhead separators, and requires some remedial operational procedures.

The most common remedies for gas-well liquid loading are installing tubing in the vertical part of the wellbore to increase flow velocity (if not installed initially), installing a pump or plunger-lift system, or using gas lift if water rates are high and the well is very productive. Gas-lift is rarely applied for dry-gas shale wells due to the installation cost and marginal profit of many wells, whereas plunger-lift seems to be widely used both in shale and tight gas systems, see e.g. Stevenson and O'Shea (2006); Tang (2009); Kravits et al. (2011). Plunger-lift is an artificial-lift technique, based on installing a metal plunger in the tubing to supply a solid interface between the gas and liquids to be lifted, and using the energy of gas expanding in the annulus during a shut-in to lift to the surface both the plunger and liquids accumulated in the tubing once the well is re-opened. Production with a plunger lift system implies an intermittent shut-in and production scheme. Efficient operations with plunger-lift systems rely on good control and optimization of the flowing, after-flowing and shut-in cycle, which may be complex (Tang, 2009; Gasbarri and Wiggins, 2001), particularly for multi-well systems producing to shared surface facilities. From the literature, however, it does not seem to be a common understanding of the actual *effect* of the solid interface provided by a plunger for increasing the volume of lifted liquids, as most models neglect liquid fallback when the plunger rises (Lea, 1982; Marcano, 1994; Gasbarri and Wiggins, 2001). Tang (2009), on the other hand, includes this effect in a high fidelity model, and reports that almost no liquid remains in front of the plunger when hitting the wellhead. Whitson et al. (2012) develop an alternative cyclic production strategy to plunger lift for eliminating liquid loading in low-permeability tight-gas wells. Compared to plunger lift, this scheme does not rely on installing a metal device in the tubing. Moreover, the scheme is based on shutting in wells *before* liquids accumulate in the wellbore, ensuring that production is always above the critical gas rate q_{gc} needed to continuously remove co-produced liquids in the wellbore (Turner et al., 1969). This scheme is illustrated in Fig. 1.4 for a *fixed* shut-in time, with the critical rate computed by the correlation given in Coleman et al. (1991). The scheme utilizes both pressure build-up in the well and in the stimulated close-wellbore region to regain a gas rate above q_{gc} once the well is re-opened. An added benefit of this scheme is that eliminates safety issues of a free moving metal-plunger moving at high velocities to the surface.

In addition to operational issues related to liquid loading, operators must also adapt daily routines and operations to various planned and unplanned events. This includes well and surface-equipment maintenance, fluctuations in surface line pressures, well testing, sudden fall in gas production, and shut down of production due to fracturing in nearby wells (King, 2010). Challenges in terms of adapting daily operations to such events depend on the field infrastructure, the size and distribution of the interconnected pipeline network between wells, available rate and pressure measurements, and the integrity of each well and production pad. That is, if the wellhead pressure or only a shut-in choke can be remotely controlled, how much surface production equipment that is shared, and if production scheduling is coordinated between the wells. Considering the distributed nature of shale-gas installations and the fact that most fields eventually contain several hundreds of wells, operators may clearly benefit from having a well defined production plan and a decision support tool (DST) for optimizing their operations.

1.1.2 Shut-ins of shale-gas wells

The characteristic system of highly conductive fracture networks and low permeable, ultra-tight shale matrix blocks enables the use of shut-ins of shale-gas wells in a distinct way compared conventional oil and gas wells. This reservoir characteristic is the basis for the previously mentioned cyclic production scheme proposed by Whitson et al. (2012) to prevent liquid loading. As demonstrated through the cyclic shut-in simulation shown in Fig. 1.4, shale-gas wells exhibit an immediate peak-rate after shut-ins. This phenomenon has been observed in production data with isolated shut-ins, e.g. King (2010); Valkó (2009); Jayakumar et al. (2011); Jayakumar and Rai (2014), although, as mentioned in the previous section, most frequently related to an initial shut-in period (Cheng, 2012; Makhanov et al., 2013). Shut-ins of shale-gas wells have also been performed and studied related to the plunger-lift technique, however, with durations of minutes or hours. The *effectiveness* of applying shut-ins to shale-gas wells, both as a means of preventing liquid loading and as a mechanism for production optimization with respect to varying demands and prices, has though received little attention.

The ability to efficiently perform shut-ins of shale-gas wells is associated with high pressure gradients caused by the difference in conductivity of low-permeable shale matrix blocks and the interconnected fracture network. After initial depletion of gas released by the HF stimulation, the shale matrix blocks act as a source term with almost no pressure depletion, feeding gas slowly into the fracture system. During shut-ins, the sustained high pressure gradients will cause gas to continue flowing into the fractures and wellbore. This recharge of gas in the fractures, as well as the gas expansion itself, increases the well and near-wellbore pressure. Once a well is reopened, the gas recharged in fractures causes the characteristic peak rate shown in Fig. 1.4.

The two main factors controlling the volume of gas recovered after a shut-in of dry and semi-dry shale-gas wells are the shut-in length and formation permeability. Fig. 1.5 demonstrates the impact of these parameters on the efficiency of performing cyclic shut-ins on shale-gas wells as illustrated in Fig. 1.4. The figure shows total recovery after a *fixed* 10 year simulation horizon, normalized against the cumulative production obtained if the well is continuously produced, i.e. with no shut-ins, as shown with a red dotted line in Fig. 1.4. This is the maximal recovery a well can obtain, and assumes implicitly an ideal, unrealistic production profile with no liquid loading occurring, or some "miraculous" means of continuously removing liquids from the wellbore. The simulations are performed using a high-fidelity shale-gas model implemented in the state-of-the-art reservoir simulation software SENSOR (SENSOR, 2011), see Knudsen et al. (2014c). Cumulative production is collected at the end of the fixed ten year prediction horizon. For each simulated formation permeability k_{m}, 0% shut-in time corresponds to continuous production, giving a normalized recovery of 1 as shown in the upper left corner of Fig. 1.5. Each time the gas rate hits the lower (critical) rate q_{gc}, computed by the correlation in Coleman et al. (1991), the well is shut in for a predefined, minimum time, and reopened once the pressure is sufficiently built-up to regain a flowrate above q_{gc}.

From Fig. 1.5, it is evident that the combination of low formation permeabilities and shale reservoir's characteristic fracture-matrix system allow shut-ins to

be applied with very little loss of cumulative production. For the lower range of simulated permeabilities, the given dry-gas shale-well realization used for the simulation allows the well to be shut-in more than 50% of total operation time while still recovering nearly 100% of the maximum recovery obtained by the idealized continuous production scenario. This property is unique for fractured shale and tight-gas reservoirs; the mentioned combination of fracture matrix systems and corresponding high reservoir pressure gradients cause the magnitude of the subsequent peak rate after a shut-in to immediately recover almost all production temporarily lost during shut-in. Thereby, there is a negligible difference in cumulative production compared to continued production at low pseudo steady-state rates as shown by the red line in Fig. 1.4. This property prevents any significant reduction in net present value (NPV) for shale wells applying the shut-in approach. In comparison, considering a conventional high-permeability gas reservoirs and a fixed horizon as used in Fig. 1.5, the temporary loss in production during a shut-in would not be recovered before *after* the end of the prediction horizon, that is, after the end of the time horizon shown in the shut-in illustration in Fig. 1.4, thus reducing NPV significantly. Observe that the given value of shut-in rate q_{gc} impacts the ratio of shut-in and production time; a lower value of this rate would generally cause a longer production time before the well must again be shut in, thereby reducing the total percentage time the well is shut-in. The critical rate is, however, in practice determined by properties of the well, see Turner et al. (1969).

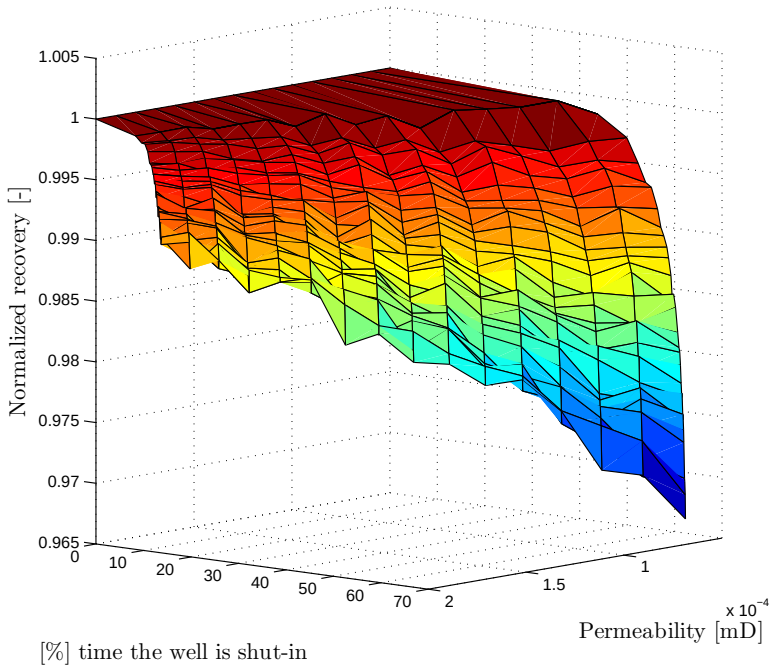


Fig. 1.5: Normalized relative recoveries as a function of the formation permeability k_{m} and the percentage time of total operation time the well is shut-in. From Knudsen et al. (2014c).

It is important to note that above a certain percentage total shut-in time the

recovery suddenly drops, eventually leading to loss in NPV and as such reducing the economic viability of performing this type of shut-in schemes on shale-gas wells. Moreover, shut-ins are generally undesirable for operators during the initial peak production and rate transient, where much of the revenue of a well is gained, cf. Fig. 1.4. Nevertheless, it is clear that fractured shale-gas wells with low permeabilities are particularly suitable for intermittent shut-in schemes. Shut-ins may as such both be used to prevent liquid loading and thereby stabilize production, but also as a means of scheduling production with respect to varying demands and prices, so as to increase revenue for well operators, while meeting predicted or varying seasonal gas demands from end-users. The latter type of utilization translates to using shale-gas reservoirs as a type of gas storage, a property which is explored in Knudsen et al. (2014c).

Performing well shut-ins in large, distributed shale-gas systems naturally present challenges for operators. Depending on well characteristics and duration of shut-ins, the wells may produce very high peak rates after being re-opened, and subsequently very low rates. This behavior of the wells may lead to undesirable pressure and flowrate oscillations in the surface pipelines. Wells that share surface equipment may also benefit from a structured shut-in schedule to prevent violating capacity constraints. With the number of wells running into the hundreds, some producing at high initial rates and some at low, possible erratic rates, there is a clear incentive to construct scheduling plans, either in terms of a decision support tool for operators, or eventually as an automatic, possibly closed-loop, scheme for multi-well shale-gas shut-in scheduling, which allows for including flexible production objectives and constraints.

1.2 Simplified systems modeling of shale-gas wells

This section contains a brief review of the most commonly applied techniques for modeling shale-gas wells and reservoirs. The section further contains a brief description of general-purpose modeling techniques in predictive control and system identification, as well as conventional proxy-modeling schemes in reservoir modeling. The chosen proxy-modeling approach applied for shut-in scheduling is then placed into context of these different modeling techniques.

1.2.1 Shale-gas reservoir modeling

Gas flow and pressure behavior in fractured shale and tight-gas reservoirs is complex and made up of several components. Much work has been done to understand and develop detailed numerical models of the reservoir flow mechanisms, as well as developing simple static forecasting models with the objective of estimating long-term recovery rates. Below we give a brief description and associated references, without being exhaustive, to these two modeling approaches.

The two most commonly applied basis models for matrix-fracture systems in shale and tight-gas reservoirs are the dual-porosity model for naturally fractured reservoirs, introduced by Warren and Root (1963) and Kazemi (1969), and fully discretized single-porosity dual-permeability models (Cipolla et al., 2010). Dual

porosity systems model the fracture and rock matrix-blocks as two distinct media with different properties. The main difference between the numerous variants of dual-porosity models that have been developed, lies in modeling of the flow between the two media, and the imposed geometry of fractures and shale matrix blocks (Carlson and Mercer, 1991; El-Banbi, 1998; Medeiros et al., 2010; Ozkan et al., 2011). Even though these models intuitively fit the matrix-fracture system of shale-gas reservoirs, their idealized cube-based geometry and need to set reservoir specific coefficients such as matrix storativity and interporosity flow, often leads to a lack of accuracy in the longterm transient flow behavior in shales (Cipolla et al., 2010).

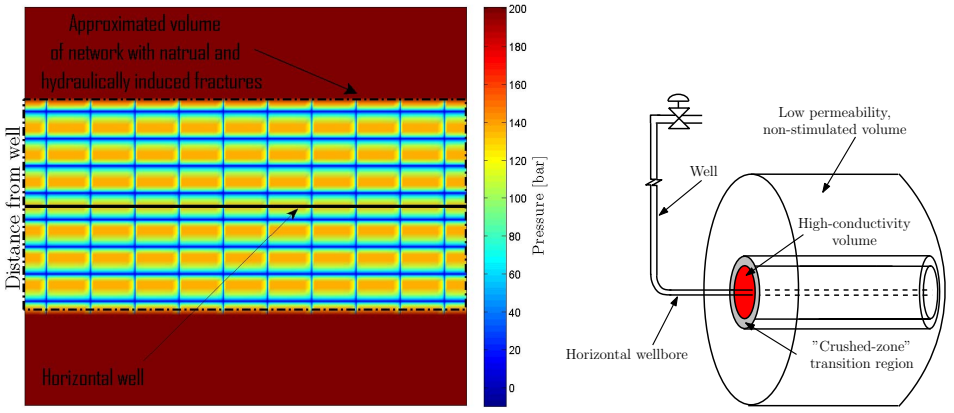
The approach of directly discretizing the entire system of shale-matrix blocks and fracture network (Cipolla et al., 2010) renders high modeling flexibility. In this approach, the reservoir is treated as *one* media or continuum, using a single porosity but dual permeability representation of the matrix-fracture system, i.e. one low formation-permeability for the rock matrix-blocks and a corresponding high permeability for fractures. The fracture permeability can be adjusted to achieve an appropriate fracture conductivity, defined as the product of the fracture permeability and width (Cipolla et al., 2010). Applying full discretization of the reservoir facilitates long-term shale reservoir characteristics such as gas desorption and stress-dependent fracture permeability, that is, reduction of permeability when pressure drops. Moreover, the scheme may account for complex natural and hydraulically induced fracture networks, and transient effects caused by non-Darcy flow in the fractures (Cipolla et al., 2010). The clear disadvantage of the full discretization scheme is its often poor numerical efficiency. Upscaling is hence normally applied (Cipolla et al., 2010; Wilson and Durlofsky, 2013; Knudsen et al., 2014b), using for instance logarithmic grid refinements around the fractures. Other references to this shale reservoir modeling approach include Mayerhofer et al. (2010); Economides and Wang (2010) and Freeman et al. (2013). Several general-purpose modeling schemes for fractured reservoirs have also been adapted and applied to shale-gas reservoir modeling, see for instance Karimi-Farad et al. (2004). Observe that very few numerical shale reservoir modeling approaches apply direct multi-well modeling. Although interference effects may occur during HF stimulation and hence alter the effective stimulated reservoir volume (SRV) (Mayerhofer et al., 2005), the common approach is to account for these effects by changing the fracture geometry and distribution in a single well model. Possible well and fracture interference effects consist nevertheless of slow dynamics with long transients.

A widespread technique for forecasting long-term recovery rates in shale-gas reservoirs is the use of static, semi-analytical or empirical models (El-Banbi, 1998; Ilk et al., 2008; Al Ahmadi et al., 2010; Bello and Wattenbarger, 2010; Nobakht et al., 2012). Many operators tend to avoid numerical models in favor of these types of models due to their ease of construction and maintenance (Ambrose et al., 2011), as well as ease of use in history matching of production data, i.e. parameter estimation, and estimation of original gas in place. These classes of forecasting tools are mainly based on either Arp's hyperbolic decline curves (Arps, 1945), or the so-called square-root-of-time analysis introduced by El-Banbi (1998). Although these models may be sufficiently accurate and good in matching production data, they all assume operation at constant bottomhole pressure or constant rate. Consequently, these models are not suited for optimization of intermittent shut-in schemes or

other dynamic production optimization schemes.

All of the above shale reservoir modeling schemes have been constructed both with Cartesian geometry, e.g. Cipolla et al. (2010); Medeiros et al. (2010), and cylindrical geometry (Economides and Wang, 2010; Larsen and Hegre, 1994; Knudsen and Foss, 2013). The two most commonly applied matrix-fracture geometries are the so-called slab model, illustrated in Fig. 1.7, and cube or fracture-network models (Al Ahmadi et al., 2010; Cipolla et al., 2010). Slab models construct the matrix-fracture geometry as a fixed number of hydraulically induced, planar fractures emanating from the horizontal wellbore and fully penetrating the reservoir thickness, cf. Fig 1.7. Cube models are similar, but also presumes existence of natural fractures that are parallel to the horizontal wellbore, hence creating a rectangular fracture network with the hydraulically induced fractures, cf. Fig 1.6a. The most suitable geometry will hence depend on the formation the well is drilled into, if larger networks of natural fractures are likely to exist, and the stimulation design for the well. More randomized fracture-patterns may also be considered (Knudsen, 2011; Jayakumar et al., 2011). On longer horizons, however, local pressure depletion in the rock matrix-blocks between the fractures may lead to a "homogenization" of the matrix-fracture system, causing different matrix-fracture geometries to lead to similar production profiles (Knudsen, 2011; Jayakumar et al., 2011).

Note that the backpressure equation and conventional inflow performance relationship models, see e.g. Fetkovich (1980) and (Golan and Whitson, 1991, Ch. 2), are normally not applied to shale-gas reservoirs, as these models assume steady state flow, whereas the long transient behavior of shale-gas wells cause these wells to never reach such a state. Consequently, a constant drainage area is essentially never established during the life-time of a shale-gas well.



(a) Snapshot of grid pressure (at 850 days) in a high-fidelity shale-gas reservoir model.

(b) Illustration of the radial composite shale-gas reservoir proxy model.

Fig. 1.6: Conceptual illustration of the SRV based proxy modeling approach (Knudsen and Foss, 2013)

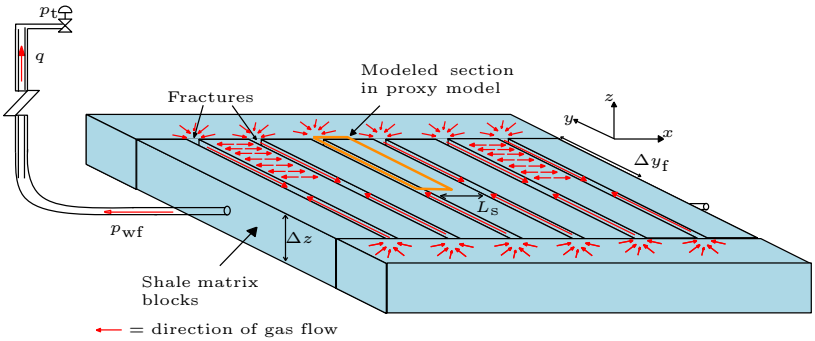


Fig. 1.7: Illustration of slab matrix-fracture geometry and the section used in the explicit fracture-modeling proxy (Knudsen et al., 2014b).

1.2.2 Proxy and black-box modeling

Proxy models, also known as surrogate or meta models, have been developed and applied in many contexts of reservoir modeling and management. Proxy modeling is applied within a wide range of statistics and engineering disciplines, and its definition is somewhat vague. However, common for all types of proxy models is the attempt to replace a complex model or system description with a simpler and more efficient model, while still retaining sufficient accuracy for the given application.

Within the domain of reservoir modeling and production optimization, commonly applied data driven proxy modeling techniques include polynomial regression, response-surface models and artificial neural networks (Bieker et al., 2007). Each of these types of proxy-modeling schemes are examples of nonlinear *black-box* model structures used within system identification (Ljung, 1999, Ch. 5.4-5.5). In comparison, the most commonly applied model structures in linear system identification are finite impulse response models, families of transfer functions and linear state-space models (Sjöberg et al. (1995); Ljung (1999, Ch. 4.2)). System identification with these types of model structures may be applied with very little a priori knowledge of the system, with an acceptable model fit often obtained in a trial-and-error fashion with parameter estimation and testing of different model structures. Although these types of black-box modeling approaches may be very efficient and in fact highly suitable for many applications, the intrinsic disadvantage lies in not considering relevant underlying physics of the system. System behavior that is not observable in the given input data may hence be lost, i.e. not captured in the model, causing loss of accuracy, particularly for longer prediction horizons. Black-box proxy models may both be constructed directly from (filtered) production data, or via construction of a high-fidelity model as illustrated in Fig. 1.8.

Gray-box modeling tries to combine physical insight and prior knowledge of the system when selecting model structure and parameters to achieve the required performance of the model (Tulleken, 1993; Thompson and Kramer, 1994; Sjöberg et al., 1995). This types of modeling is also referred to as a semi-physical approach, where the model structure may be mainly mechanistic, but augmented with stochastic elements, empirical relations and parameters that are not necessar-

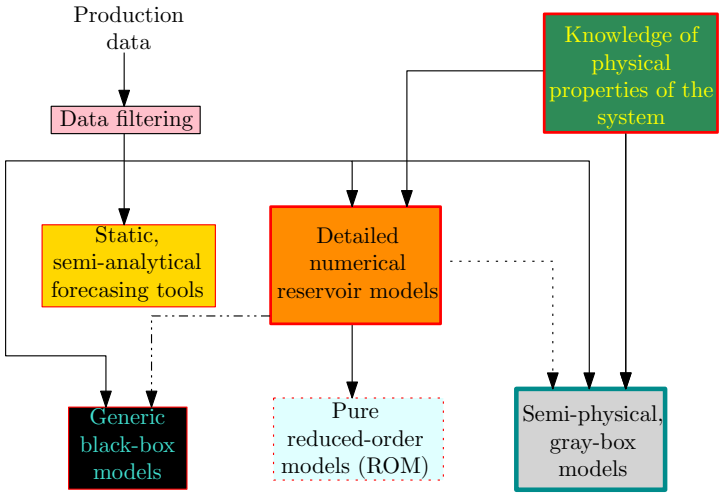


Fig. 1.8: Connections between different modeling approaches for shale-gas wells.

ily physically interpretable (Johansen and Foss, 1997). Gray-box modeling often provides a good compromise for constructing small sized and sufficiently accurate proxy models, compared to rigorous physical modeling, which may be expensive to construct and render prohibitive large problem sizes, and black-box models that are not compatible with physical reality (Tulleken, 1993).

Proxy models, belonging to the category of reduced-order models (ROMs), may also be considered in conjunction with model-reduction techniques, particularly for systems governed by partial differential equations (PDEs) such as shale-gas reservoirs, also referred to as distributed parameter systems in system identification. Construction of ROM from PDEs relies on explicit knowledge of the underlying model equations, using a model-reduction technique such as proper orthogonal decomposition to reduce the size of the model, e.g. Kerschen et al. (2005); Cardoso and Durlofsky (2010); Suwartadi (2012). The efficiency and accuracy of this type of ROM construction depend on several factors, including accuracy of the detailed high-fidelity model and its degree of nonlinearity (Cardoso and Durlofsky, 2010), the model-reduction technique's ability to sufficiently reduce the model size, and the level of details to be retained in the proxy model. The obvious drawback of this approach is as illustrated in Fig. 1.8 is the need to first derive and construct the high-fidelity model before generating the proxy/ROM model.

1.2.3 Dynamic proxy modeling of shale-gas reservoir

The two categories of commonly applied shale well modeling schemes described in Section 1.2.1, detailed numerical modeling and static decline-curve approaches, are both inadequate for optimizing time-varying shut-ins of shale multi-well systems in a scheduling or model predictive control (MPC) type of context. Shale-gas reservoirs do however contain special structures that can be exploited. This favors a gray-box approach with first principal physics, using knowledge of physical prop-

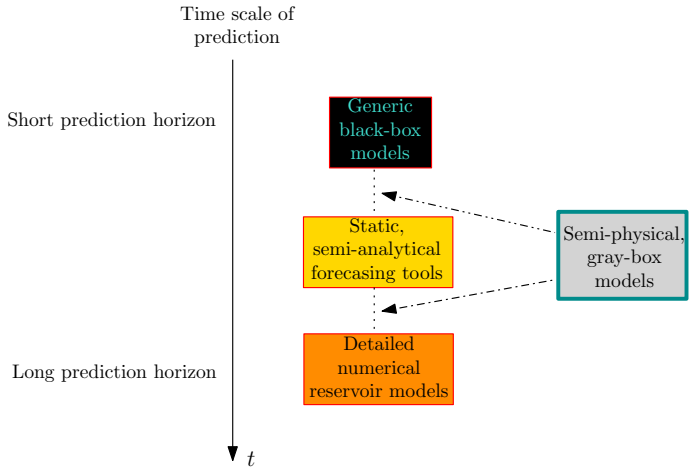


Fig. 1.9: The accuracy of different modeling approaches for shale-gas wells with respect to prediction time.

erties of gas flow and available input production data as shown in Fig. 1.8, rather than a pure black-box approach. This choice is also related to the appropriate time scale of prediction for different shale well modeling approaches as shown in Fig. 1.9. Depending on tuning and size of the model, as well as the actual behavior of the well, a semi-physical gray-box approach may be designed to be sufficiently accurate also for fairly long predictions as shown in the same figure. It is also worth commenting that from a practical perspective, a proxy model preserving physical interpretation is also more likely to be accepted by reservoir engineers and operators when integrated in a decision-support tool.

The development of the shale-gas reservoir proxy model uses as its basis the parabolic PDE describing one-dimensional isothermal flow of real gas through porous media,

$$\phi \frac{\partial}{\partial t}\left[\frac{p}{Z(p)}\right] = \nabla\left(k\frac{p}{\mu(p)Z(p)}\nabla p\right), \tag{1.1}$$

see e.g. Al-Hussainy et al. (1966) or Aziz and Settari (1979, Ch. 2). In (1.1), p is pressure, ϕ is porosity, $Z(p)$ is the real gas compressibility factor, k is permeability, $\mu(p)$ is gas viscosity and ∇ is the divergence operator. The PDE (1.1) is derived by using the law of mass conservation, and assuming Darcy's law, single phase gas and constant porosity. To account for the pressure variations of $\mu(p)$ and $Z(p)$ shown in Fig. 1.10, we use the variable transformation

$$m(p) := 2\int_{p_\mathrm{b}}^{p} \frac{p'}{\mu(p')Z(p')}dp', \tag{1.2}$$

known as gas pseudopressure (Al-Hussainy et al., 1966). The pseudopressure $m(p)$ hence corresponds to the integral of the red line in Fig. 1.10, with p_b being a low base-pressure. By further assuming negligible rock compressibility and using the

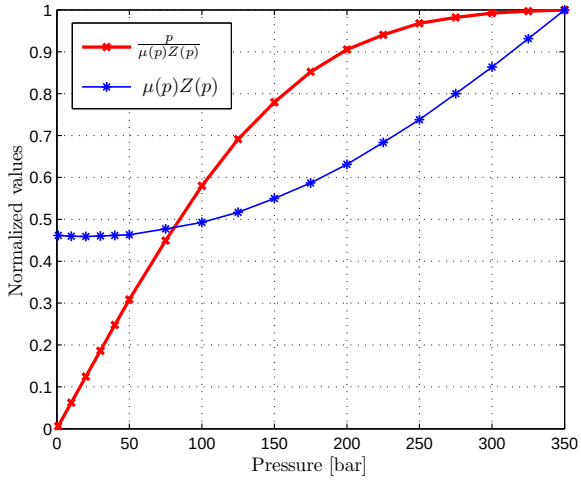


Fig. 1.10: Variations of $p/\mu(p)Z(p)$ and $\mu(p)Z(p)$ for real gas with specific gravity $G = 0.7$ and reservoir temperature $T = 366.3$ K. The $Z(p)$-values are based on the Hall-Yarbourough correlation, and the $\mu(p)$-values are based on the Lee-Gonzales correlation (Whitson and Brule, 2000, Ch. 3).

definition of isothermal gas compressibility (Whitson and Brule, 2000, Ch. 3),

$$c(p) = \frac{1}{\rho}\frac{d\rho}{dp}\bigg|_T = \frac{Z(p)}{p}\frac{d}{dp}\left[\frac{p}{Z(p)}\right] = \frac{1}{p} - \frac{1}{Z}\frac{dZ}{dp}\bigg|_T, \tag{1.3}$$

where T is temperature and ρ is real gas density, the PDE (1.1) can be reformulated in terms of $m(p)$,

$$\phi\mu(p)c(p)\frac{\partial m}{\partial t} = \nabla\left(k\nabla m\right). \tag{1.4}$$

Using the PDE (1.4) as basis, Knudsen et al. (2012); Knudsen and Foss (2013) and Knudsen et al. (2014b), respectively, constructs two differnt shale-gas reservoir proxy models of equal complexity but different approximations and geometries of the matrix-fracture system. The model developed in Knudsen et al. (2012) and Knudsen and Foss (2013) approximates the fracture system or stimulated reservoir volume (SRV) (Mayerhofer et al., 2010) as one big highly conductive volume as shown in Fig. 1.6a. The approximated fracture-system volume may either consist of a network with hydraulically induced and preexisting natural fractures assumed parallel with the horizontal wellbore, or the volume may consist of only multiple planar transverse fractures along the wellbore as shown in Knudsen and Foss (2013) and Knudsen et al. (2014a). Between this high-conductivity volume and the unstimulated outer volume shown in Fig. 1.6a, we include a crushed-rock volume with an intermediate permeability of the two other volumes. These three volumes are connected together as a radial composite model shown in Fig. 1.6b, consequently expanding the divergence operator on right-hand side of (1.4) in a cylindrical coordinate system.

The second shale-gas reservoir proxy model (Knudsen et al., 2014b) applies explicit modeling of a single transverse fracture and the adjacent shale matrix block. This type of shale reservoir proxy-modeling approach may be regarded as a simplified variant of the category Mohaghegh (2013) in a recent publication refers to as explicit hydraulic fracturing (EHF) models. The model applies the previously mentioned matrix-fracture slab geometry illustrated in Fig. 1.7, using a Cartesian geometry and assuming that the fractures are symmetric around the wellbore and equally spaced. The dominating direction of the flow in the low-permeable shale matrix blocks is orthogonal to the fractures, i.e. in the x-direction, referred to as linear flow (Bello and Wattenbarger, 2010; El-Banbi, 1998). As the pressure drop in the fractures are negligible due to high fracture conductivity (Knudsen, 2011), we can construct a proxy model for a *linear flow regime* by only considering a quarter section of the slab-fracture system as illustrated by the orange rectangle in Fig. 1.7. In this case, the divergence operator in (1.4) is naturally expanded in a one-dimensional Cartesian coordinate system. Note that this model does not account for radial inflow into the fractures (Bello and Wattenbarger, 2010), or external radial inflow into fractures from the unstimulated volumes outside the stimulated region, cf. Fig. 1.7.

The pseudopressure-based PDE (1.4) is still nonlinear due to pressure dependency of the product $\mu(p)c(p)$. Including the pressure variations of this product is more involved (Reynolds et al., 1987), causing further variable transformations, approximations or linearization techniques to be applied. Rather than applying the conventional technique of evaluating $\mu(p)c(p)$ at initial conditions (Al-Hussainy et al., 1966) or some other predefined pressure, we include this product in a parameter estimation scheme to achieve a preferred performance of the proxy model (Knudsen and Foss, 2013; Knudsen et al., 2014a). This approach is described for the cylindrical SRV-based proxy model in Chapter 2 and 3, and for the slab-based proxy model in Appendix A. Observe that by treating the compressibility $c(p)$ as a constant, we mimic the equation for a slightly compressible fluid (Aziz and Settari, 1979), leaving (1.4) as a semi-linear parabolic PDE (Ockendon et al., 2003, Ch. 1.3) as k is assumed to be independent of pressure.

Treating the product $\mu(p)c(p)$ as constant parameters renders the possibility of obtaining a linear discretization of (1.4). To this end we apply finite difference approximations of the time and spatial derivatives. Central difference approximations are applied to the divergence operator on the right-hand side of (1.4), the spatially varying permeability k, and associated Neumann boundary conditions, giving second-order accuracy. The first-order accurate Backwards Euler's method is applied for time discretization. Details on the discretization of the cylindrical proxy model can be found in Knudsen et al. (2012), while the reader is further referred to (Aziz and Settari, 1979, Ch. 3) and (Abou-Kassem et al., 2006, Ch. 4-5) for more details on finite difference discretization of the two reservoir proxy models. The complete discretization secures an implicit, consistent, unconditionally stable and dissipative scheme (Strikwerda, 2004, Ch. 6.3). Several different finite difference schemes for (1.4) may be considered such as the Crank-Nicolsons scheme, see Strikwerda (2004, Ch. 6) for more details. Another possibility is to use a semi-discretization (also known as the method of lines) or the finite element method for discretizing the right-hand side of (1.4), leaving a set of ordinary dif-

ferential equations (ODEs) where the time derivative of m may be subsequently discretized by a higher-order accurate orthogonal collocation on finite elements, see Biegler (2010, Ch. 10).

The two related proxy models with geometries shown in Fig. 1.6b and 1.7, respectively, may in general be applied interchangeably. The most suitable model structure will depend on the reservoir characteristics, in particular the assumed fracture distribution and geometry, and the application of the model. This is further elaborated in the chapters containing applications of the respective proxy models.

1.2.4 Well and surface system modeling

The inherent dynamic interactions between the low permeable shale matrix blocks and the highly conductive fracture networks are the most essential dynamics to describe gas production from shale wells. Shale-gas wells, however, rarely contain bottomhole control valves or measurement units, and are hence almost invariably controlled through wellhead valves. A complete *shale well* proxy model must therefore contain a coupling of the wellbore-tubing gas flow and the reservoir model, covering essential expected features of the well performance.

A commonly applied technique when coupling wellbore and reservoir flow is to use a static wellbore or tubing model in connection with a dynamic reservoir model (Chupin et al., 2007; Chen, 2007), or conversely, using a transient multiphase wellbore model with a static inflow relationship from the reservoir, e.g. Tang (2009); Stevenson and O'Shea (2006). Within scheduling and production optimization of petroleum systems, however, both the reservoir and wellbore are often described by static models, e.g. Kosmidis et al. (2005). A fundamental challenge when coupling *dynamic* well and reservoir flow is the different time scales of the models (Coats et al., 2004). Choosing whether to include dynamics in the entire or parts of the reservoir-to-surface system, and in the latter case, in which section of the system to include dynamics, depends consequently on the application and associated time scale of the model. Integrated multi-well reservoir and surface models of shale and tight gas systems are limited. Stevenson and O'Shea (2006) present a case study for modeling of a large-scale tight-gas gathering system, using a nodal-analysis technique with static models of the wells, pressure loss in pipelines, and the compressor-capacity curves. The model objective in this study is simulation of pipeline pressure loss and compressor fuel usage in a system with 1150 wells and many miles of pipelines. Tang (2009) presents a model of tight-gas multi-well pads running simultaneous plunger-lift systems, using a static reservoir inflow model and a dynamic multi-phase wellbore and plunger-lift model. In this application the shut-in lengths are in the range of minutes, hence enhancing the importance of including detailed wellbore dynamics. Moreover, the shut-ins are performed *after* water has accumulated in the wellbore, requiring multiphase flow modeling of the wellbore. The study, however, emphasizes the importance of including reservoir and wellbore characteristics in order to optimize production of shut-in based plunger-lift operations for these types of wells.

Due to the time scale of the shale well shut-ins we consider, we use simple static models of the tubing and wellbore. For single phase gas, the steady-state mechanical energy balance can be used to derive a quadratic tubing model (Katz and Lee, 1990,

Ch. 6.4). A linearizable aggregated well and wellbore model is applied in Knudsen and Foss (2013); Knudsen et al. (2014a), neglecting the friction term in the tubing model, but including a term for maximum allowed gas rate for each well, a value normally set during design and completion of the well and surface equipment. Improved proxy-model accuracy can be obtained by including the friction term (Knudsen et al., 2014b,c), however, leading to a nonlinear tubing model. The value of adding this term is most significant if the wells are producing at high gas rates. Whether to include the full nonlinear or the simplified tubing model in the shale well proxy model is hence a question of desired model accuracy, and the configuration and design of the actual well.

Surface systems for shale-gas wells normally consist of a wellhead choke, separation tanks, gathering pipelines and possibly pumping or plunger-lift units if the wells are liquid loading. The set-up of shale-well surface and gathering systems vary greatly between fields and operators. Each well is, however, equipped with its own wellhead choke, while separation tanks might be shared between multiple wells on a pad or installed for each separate well. Separation is more comprehensive for condensate rich shale wells than dry-gas wells. The separation performed in dry-gas wells mainly impacts the system by slightly lowering the gas pressure, and is hence omitted in the system model. Furthermore, the gathering pipelines between well pads and compression units or processing plants are normally quite short. As separation is performed at the well pad *before* the gas is sent through gathering pipelines, the pipeline flow is single phase gas. These two properties cause the gathering pipeline pressure-drop to be small, and is therefore also omitted in the description of the systems considered.

1.3 Large-scale mixed integer programming

The modeling of shut-ins for multi-well and multi-pad shale-gas systems is performed using mixed integer programming techniques. Generalized Disjunctive Programming (GDP)(Raman and Grossmann, 1994) is used to systematically derive numerically efficient mixed integer models. A block-diagonal structure in several of the mixed integer programs (MIPs) further favors use of a decomposition technique to efficiently solve the problems. This section contains a brief introduction to mixed integer programming and decomposition techniques. For a comprehensive description of GDP and related logic-based modeling techniques, see Grossmann and Trespalacios (2013) and Hooker (2000). The reader is further referred to Wolsey (1998) for a detailed introduction to mixed integer linear programming.

An MIP is an optimization problem containing both continuous and discrete variables. A wide range of codes and solution techniques have been developed for solving these types of optimization problems, see e.g. Lodi (2010) and Grossmann (2002). Mixed integer programming is naturally distinguished between problems with only linear constraints, e.g. mixed integer linear programs (MILPs), and problems also containing nonlinear, possible nonconvex constraints, rendering mixed integer nonlinear programs (MINLPs). Approaches for solving MILPs can roughly be divided into LP-based branch-and-bound methods (Land and Doig, 1960) and cutting-plane algorithms (Gomory, 1958). Combining these two approaches has

lead to the development of efficient branch-and-cut methods (Rinaldi and Padberg, 1991), the fundamental algorithm for high-fidelity MILPs solvers such as CPLEX, Gurobi and Xpress–MP. Major advances in computer hardware and MILP components, particularly presolving routines and cut generation (Bixby and Rothberg, 2007), has caused a tremendous speed up of these codes from their earliest versions.

Although significant advances also have been made in nonlinear mixed integer programming (Hemmecke et al., 2010; Grossmann, 2002), the available codes for MINLPs are far behind in efficiency compared to commercial MILP solvers. Convexity of the continuous NLP relaxation plays a fundamental role for the choice of solution approach to an MINLP, in the sense that NLP relaxation only provides a valid lower bound (for minimization problems) if the relaxed problem is convex. Methods for solving convex MINLPs include NLP-based branch-and-bound (Gupta and Ravindran, 1985), Outer Approximation (Duran and Grossmann, 1986), Generalized Benders Decomposition and more, see Grossmann (2002). All of these methods are exact for convex MINLP, while they may be applied as a heuristic methods to nonconvex MINLPs, that is, with no guarantee of global optimality of the solution or exact duality gap. Several rigorous global optimization algorithms have been developed for nonconvex NLPs and MINLPs, including spatial branch-and-bound (Smith and Pantelides, 1999) and branch-and-reduce algorithms (Sahinidis, 1996; Ryoo and Sahinidis, 1996). See e.g. Tawarmalani and Sahinidis (2004) for details.

While continuous improvements in MILP techniques and software have facilitated the solution of large and difficult problems, the problem size still has a major impact on achievable computational performance. This relates to the computational cost of solving large LP relaxations, and the fact that MILPs belong to the class of NP-hard problems (Nemhauser and Wolsey, 1988, Ch. 7). Many real-world, industrial optimization problems with discrete decisions lead to huge and complex mixed integer problems, causing a prohibitive long computation time with one of the aforementioned exact algorithms, consequently reducing the practical applicability on these problems. In order to overcome these limitations in the endeavor of solving such problems, one may consider several approaches. A first step is to try reducing the problem size or complexity, for instance by removing unnecessary dynamics and constraints, using surrogate models, or improving the existing formulation using reformulations and tighter variable bounds. For many problems, however, the only viable solution approach is to use a heuristic method. Heuristics optimization algorithms are diverse, ranging from *metaheuristics* such as Genetic algorithms, Tabu Search and Simulated Annealing, typically developed through experimental learning (Glover and Kochenberger, 2003), to pure MIP heuristics such as the Feasibility Pump (FP)(Fischetti et al., 2005), the Undercover heuristic (Berthold and Gleixner, 2013) and Pivot-and-Complement (Balas, 1985). In addition, there exists a variety of tailored, problem-specific heuristics. For several applications, such as decision support tools (DSTs), a local solution in short computation time is more important than the global optimal solution, rendering heuristics as suitable solution approaches. Note that metaheuristics make no assumption of the underlying model, and are as such particularly suited for problems containing simulator-based black-box models.

Block-diagonal structures in the constraint set, often arising in network-type,

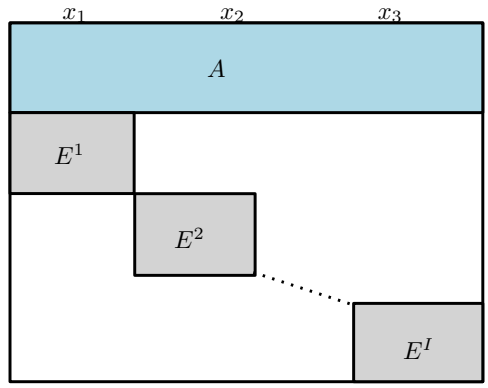


Fig. 1.11: Illustration of block-diagonal structure.

production planning, process synthesis and stochastic problems, may be utilized to facilitate the solution of large-scale MIPs. Problems with block-diagonal structures are typically linked together by only a few "complicating" constraints or variables as illustrated in Fig. 1.11. Lagrangian relaxation (LR) (Geoffrion, 1974; Held et al., 1974) provides a means of computing lower bounds for minimization problems with a block-diagonal structure by dualizing the linking constraints, thereby decomposing the original problem into smaller subproblems. The next section contains a brief description of Lagrangian relaxation in mixed integer linear programming and techniques for computing the Lagrangian dual. Problems with complicating linking variables may be solved using Bender's decomposition. This technique is not further addressed in this thesis, see Nemhauser and Wolsey (1988). Note that problems containing complicating variables can always be reformulated to a problem with complicating constraints by introducing duplicate variables.

1.3.1 Decomposition by Lagrangian relaxation

For description of Lagrangian relaxation (LR) in integer programming, consider the following common form of a linear integer program (IP)

$$\begin{aligned} Z = \min_x \quad & cx \\ \text{s.t.} \quad & Ax = b, \\ & Ex = d, \\ & x \geq 0 \text{ and } x \in \mathbb{Z}^p. \end{aligned} \tag{1.5}$$

MILPs with equal structure are treated similarly. In (1.5), the constraints $Ax = b$, with A being an $m \times p$ matrix, are assumed to be complicating linking constraints, while $Ex = d$ is assumed to have block-diagonal structure, i.e. $X = X^1 \times \ldots \times X^I$, with $X^i = \{x^i \in \mathbb{Z}^n : E^i x^i = d^i,\ x^i \geq 0\}$, for a subset of variables i. By dualizing the complicating constraints $Ax = b$ with Lagrangian multipliers λ, we obtain the Lagrangian

$$Z_{\text{LR}}(\lambda) = \min_x \{cx + \lambda (Ax - b) : x \in X\}, \tag{1.6}$$

and the corresponding Lagrangian dual

$$Z_{\mathrm{D}} = \max_{\lambda} \{ Z_{\mathrm{LR}}(\lambda) : \lambda \in \mathbb{R}^m \} . \tag{1.7}$$

The Lagrangian (1.6) can clearly be solved as I independent subproblems, with $Z_{\mathrm{LR}}(\lambda) = \sum_i Z^i_{\mathrm{LR}}(\lambda)$ providing a lower bound on Z as least as tight as the LP relaxation (Geoffrion, 1974). The Lagrangian dual (1.7) corresponds to the problem of finding the tightest Lagrangian lower bound on Z in the space of *dual variables* $\lambda \in \mathbb{R}^m$, which is equivalent with solving the following convexified problem in the space of the *primal* variables (Geoffrion, 1974):

$$Z^*_{\mathrm{P}} = \min_{x} \{ cx : Ax = b,\ x \in \mathrm{conv}(X) \} , \tag{1.8}$$

where $\mathrm{conv}(X)$ denotes the convex hull of the block-diagonal set X, which is a convex polyhedron if all entries of E and d are rational (Wolsey, 1998, Ch. 8). Solving the Lagrangian dual is hence equivalent with optimizing the primal objective over the intersection of the convex hull of non-complicating constraints and the LP relaxation of complicating constraints, see Geoffrion (1974); Guignard (2003); Frangioni (2005).

Solving large-scale (mixed) integer linear programs by Lagrangian relaxation is an iterative divide-and-conquer approach consisting of three major parts that must fulfill certain requirements for the scheme to be efficient. The Lagrangian subproblems must be significantly easier to solve than the primal problem, the Lagrangian dual must be solvable by a stable and efficient method, and recovery of primal feasible solutions must be possible by some heuristic method. Note that for Lagrangians like (1.6) formulated by dualizing an *equality* constraint, a solution to the Lagrangian that is also primal feasible, would yield the primal optimal solution, that is, the solution to (1.5). For Lagrangians with dualization of *inequality* constraints, a solution to the Lagrangian is a primal optimal solution only if complimentary slackness holds, i.e. if $\lambda (Ax - b) = 0$ (Guignard, 2003).

Construction heuristics for recovering primal feasible solutions after solving the Lagrangian for given input multipliers, typically consist of some greedy algorithm followed by a local search procedure to refine the solution (Frangioni, 2005), exploring the primal and/or dual space in proximity of the current solution points. Primal recovery heuristics in Lagrangian relaxation, sometimes denoted *Lagrangian heuristics*, are often problem dependent, although some more general methods do exist, e.g. Larsson et al. (1999); Feltenmark and Kiwiel (2000); Daniilidis and Lemarechal (2005); Sagastizábal (2012), based on certain conditions of the relaxation and choice of method for solving the dual. Deriving a suitable primal-recovery method for a given problem solved with Lagrangian relaxation may be a tedious and challenging task, and an important part of the overall efficiency of the scheme.

It is worth observing that *Lagrangian decomposition* (LD)(Guignard and Kim, 1987) is a special case of Lagrangian relaxation, where duplicate variables x' are

introduced so as to reformulate (1.6) as

$$\begin{aligned} Z = \min_{x} \quad & cx \\ \text{s.t.} \quad & Ax = b, \\ & Ex' = d, \\ & x = x', \\ & x \geq 0 \text{ and } x \in \mathbb{Z}^p. \end{aligned} \tag{1.9}$$

A Lagrangian relaxation of (1.9) is obtained by dualizing the constraint $x = x'$, rendering a relaxation with independent subproblems solved over the constraint set $Ax = b$ and $Ex' = d$, respectively. The lower bound by LD is at least as tight as the bound provided by any other Lagrangian relaxation (Guignard and Kim, 1987). However, this relaxation, also known as variable splitting, only makes sense if the variables entering in the complicating constraints also appear in the non-complicating constraints, and not just in the objective function. See Knudsen et al. (2014a) for an example on such problems.

1.3.2 Solving the Lagrangian dual

The Lagrangian $Z_{\text{LR}}(\lambda)$ (1.6) can be shown to be a piecewise linear and concave function of λ over the domain it is finite (Nemhauser and Wolsey, 1988, Ch. 6). $Z_{\text{LR}}(\lambda)$ is nondifferentiable at the breakpoints as shown in Fig. 1.12, where the solution to (1.6) is non-unique.

Methods for solving the Lagrangian Dual range from the basic subgradient method (Held et al., 1974) to more involved approaches, mainly based on cutting planes. For each solution x^n to (1.6) for given input multipliers λ^n, $(Ax^n - b)$ defines a subgradient of $Z_{\text{LR}}(\lambda)$ at the dual point λ^n (Hiriart-Urruty and Lemarechal, 1993, Ch. XII 2.1). The subgradient method of Held et al. (1974) generates a sequence $\{\lambda^n\}$ of Lagrangian multiplier updates by the rule

$$\lambda^{n+1} = \lambda^n + t^n(Ax^n - b), \tag{1.10}$$

for a suitable stepsize t^n, see Fisher (1981). Clearly, the method is very simple to implement, but suffers from lack of a termination criteria and depending highly on good stepsize adjustments. For many applications, however, this simple rule for updating the multiplier has been shown to give good practical performance.

Substituting (1.6) for $Z_{\text{LR}}(\lambda)$ in the Lagrangian dual (1.7), we obtain a maxmin problem which can be rewritten as the linear program (LP)

$$\begin{aligned} Z = \max_{\eta,\lambda} \quad & \eta \\ \text{s.t.} \quad & \eta \leq cx + \lambda\,(Ax - b)\,,\ x \in X, \\ & \eta \in \mathbb{R},\ \lambda \in \mathbb{R}^m. \end{aligned} \tag{1.11}$$

The LP (1.11) may, in principle, have as many constraints as there are solutions to (1.6), and as such be a semi-infinite programming problem (Hiriart-Urruty and

Lemarechal, 1993, Ch. XII 4.2). For computing $\lambda \in \mathbb{R}^m$ optimal Lagrangian multipliers, however, $m+1$ constraints would be sufficient. Cutting-plane (CP) based approaches for solving the Lagrangian dual are hence based on a constraint generation approach of (1.11), replacing $Z_{\mathrm{LR}}(\lambda)$ with a *cutting-plane model* $\tilde{Z}_{\mathrm{LR}}(\lambda)$, overestimating $Z_{\mathrm{LR}}(\lambda)$. Using this model, (1.11) is reformulated as the LP

$$\begin{aligned} Z = \max_{\eta,\lambda} \quad & \eta \\ \text{s.t.} \quad & \eta \leq cx^n + \lambda\,(Ax^n - b)\,, \quad n = 1 \ldots N, \\ & \eta \in \mathbb{R},\ \lambda \in \mathbb{R}^m. \end{aligned} \tag{1.12}$$

More precisely, this approach corresponds to iteratively generating a new constraint (or cut) for by solving (1.6) given input multipliers λ^n, and adding this cut to (1.12) to improve the polyhedral concave cutting-plane model $\tilde{Z}_{\mathrm{LR}}(\lambda)$ of the true Lagrangian $Z_{\mathrm{LR}}(\lambda)$. This is illustrated in Fig. 1.12, using an example from Knudsen et al. (2014a). Compared to the subgradient method, the CP method has a true termination criteria, while it is widely recognized that the method has poor numerical performance by suffering from an inherently instability, i.e. the sequence $\{\lambda^n\}$ of multiplier updates may oscillate. This may also be deduced from Fig. 1.12, seeing that the solutions to (1.12) with cuts iteratively added may jump significantly until a sufficient number of cuts is added to give a well-defined cutting-plane model of $Z_{\mathrm{LR}}(\lambda)$. Moreover, it is evident that the LP (1.12) will be unbounded during initial iterations unless bounds are assigned to λ.

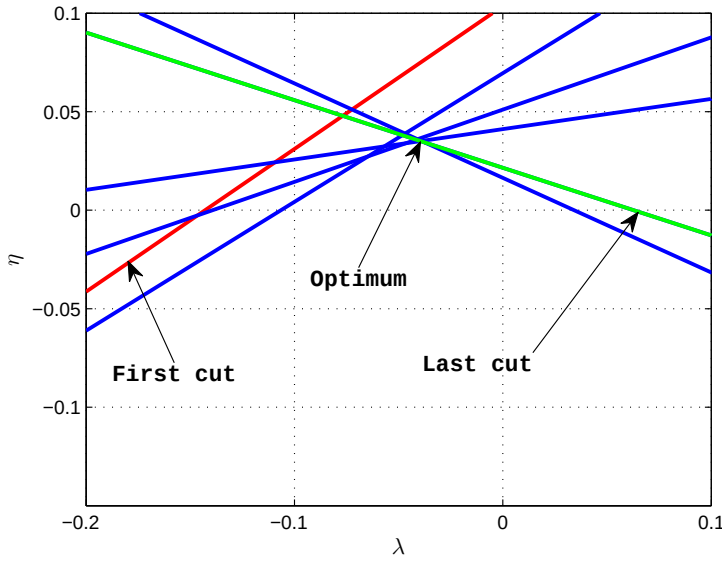


Fig. 1.12: The Lagrangian dual seen as a piecewise linear concave function with the constraints in the cutting-plane model generated by solving the LP (1.12). The example is from Knudsen et al. (2014a), using a trust-region approach to stabilize the CP method.

The drawbacks of the basic CP method have generated development of a variety of stabilized cutting-plane based approaches for solving (1.7). The methods

can roughly be divided into bundle-type and "center-based" methods (Frangioni, 2005). Bundle type methods (Hiriart-Urruty and Lemarechal, 1993, Ch. XV) seek to stabilize the CP method by introducing a stability or prox center $\bar{\lambda}$, and adding an associated stabilizing device, either in terms of a trust-region Δ around the cutting-plane model $\tilde{Z}_{\mathrm{LR}}(\lambda)$, or a quadratic penalization term in the objective function of (1.12). The latter class of methods goes under the name of proximal bundle methods, and are intimately related to the more general Moreau-Yosida regularization and proximal point algorithms, see Hiriart-Urruty and Lemarechal (1993, Ch. XV 4) and Daniilidis and Lemarechal (2005). Center based methods applied to (1.7), e.g. Goffin et al. (1992); Ouorou (2008), focus on guaranteeing a convergence speed for the sequence of multiplier updates by using so-called localization sets that contain all optimal solutions, and choosing the updates as some kind of center of this set of points (Frangioni, 2005). Generally, none of the above methods are solely superior to the others, and the choice of an appropriate method for updating the multiplier depends on the given problem to be solved.

Finally, we comment that solving the Lagrangian dual (1.7) with the cutting-plane method (1.12) is closely related to Dantzig-Wolfe decomposition (DWD). In particular, the LP dual of (1.12) is (Frangioni, 2005)

$$\min_{\theta}\left\{c\sum_{n=1}^{N}(x\theta)^n : A\left(\sum_{n=1}^{N}(x\theta)^n\right)=b,\ \theta\in\Theta\right\}, \tag{1.13}$$

with multipliers θ restricted to the unitary simplex, i.e. $\Theta=\{\theta^n\geq 0:\sum_{n=1}^{N}\theta^n=1\}$. The LP (1.13) is exactly the (restricted) primal master problem in DWD (Frangioni, 2005). DWD for solving (1.5) is hence a *column generation* approach, while Lagrangian relaxation solved with the CP method (1.11) is a constraint (cut) generation approach.

1.4 Research objective and scope

In light of the above presentation on the comprehensive development, distribution and production challenges in large multi-well shale-gas systems, a question arising is whether systematic analysis can be performed to exploit existing facilities in an optimal way, by developing tools to aid operational decisions. Optimal may be to stabilize production from one or more wells, maximizing recovery, producing according to a reference rate based on demands, or combinations thereof. These questions and targets yield the following main **research objective**:

To develop efficient models and algorithms for shut-in based production schemes, in order to enhance the utilization of existing large-scale shale-gas systems.

To approach this research objective, multiple aspects of both operational and computational character must be addressed. Significant effort is put into understanding and integrating existing and industrially established operational chal-

lenges in the problems considered, while also challenging what appears to be current best-practice by suggesting new and innovative approaches. From an computational point of view, shut-in scheduling of large-scale networks of shale-gas wells involves modeling and optimization of inherently dynamic and distributed systems with discrete decision variables.

The focus in this thesis is on *dry-gas* shale wells. Although several of the schemes developed may well be applied to condensate rich shale-gas assets, these types of wells may pose challenges not included in the model and system descriptions, while also requiring more comprehensive reservoir modeling to give accurate results.

1.5 Outline and contributions of thesis

This thesis addresses utilization of shut-ins to prevent liquid loading and stabilize late-life production in shale-multi well systems, and as a means of production scheduling with respect to varying gas demands and prices. Pursuing this research objective resorts to systematically developing suitable well and reservoir models, parameter estimation techniques, optimization models and solution algorithms.

Considering that shale-gas wells are inherently dynamic and that the systems are large, a major challenge has been to develop numerically efficient optimization schemes that can solve large-scale PDE-constrained problems with integer variables. To this end, a substantial part of the thesis addresses decomposition schemes, as several problems considered contains block-diagonal structures, as well as development of efficient heuristics.

The research conducted during the course of this thesis has been published in, or recently submitted to, international journals and conference proceedings. The four main chapters of the thesis each relates to a particular publication, included in its entirety. The main contributions in the different chapters are as follows

- **Chapter 2** introduces the concept of using shut-ins to prevent liquid loading for multiple late-life shale-gas wells producing at a single pad. The chapter presents the cylindrical SRV-based proxy model briefly introduced in Section 1.2.3, with cross-validation and description of the principals of imposing a critical gas rate. Using a logic-based model as the basis for describing shut-ins, the chapter explores different approximations and an exact MILP reformulation of a nonsmooth well and wellbore model. The chapter includes case studies on shut-in based production from a single pad with six wells, and discusses associated challenges of applying such schemes. The chapter consists of the paper Knudsen and Foss (2013), which includes preliminary results presented in Knudsen et al. (2012) and Knudsen and Foss (2012).
- **Chapter 3** describes a modified scheme for tuning parameters in the cylindrical reservoir proxy model introduced in Chapter 2. The presented parameter estimation scheme makes use of filters for selectively weighting residuals in a nonlinear least-squares problem. Using merits of the complete MILP formulation in Chapter 2 as a basis, we explore an embedded linear GDP formulation for shut-in scheduling in distributed shale multi-pad systems. A combined big-M and convex-hull reformulation is applied, rendering a large-scale MILP

which is solved using a Lagrangian relaxation scheme. The LR scheme includes an efficient combined fixing and local-search based Lagrangian heuristic, and a trust-region cutting-plane method for solving the Lagrangian dual. The chapter contains a comprehensive numerical study of the proposed LR scheme and analyzes its efficiency. The chapter consists of the paper Knudsen et al. (2014a).

- **Chapter 4** describes a novel scheme for tight integration of shale-gas producers in geographical proximity to natural-gas based electric power production. The chapter analyzes the efficiency of using shut-ins as a means of meeting varying gas demands and prices. The approach includes a more comprehensive shale-gas surface structure compared to the two previous chapters, using the slab-based proxy model introduced in Section 1.2.3 together with a GDP formulation for describing routing conditions in the model. A polyhedral convex-hull reformulation of the linear GDP is applied, utilizing disaggregated flow variables to derive a tight MILP formulation. The LR scheme applied renders a decomposition with one subproblem per well, and includes a proximal bundle method to solve the dual, a novel Lagrangian heuristic for primal recovery and a receding horizon strategy to handle operational uncertainties. Illustrative case studies demonstrate the potentials of integrating shale-gas production with varying gas demands from electric power production. Computational aspects of the proposed decomposition scheme is also analyzed. The chapter consists of the paper Knudsen et al. (2014c), with preliminary versions of models and decomposition schemes presented in Knudsen et al. (2013) and Knudsen et al. (2014b).
- **Chapter 5** presents the development of an Objective Feasibility Pump (OFP) for convex MINLP, using a structured multi-objective optimization approach. The chapter contains comprehensive computational results on a set of MINLP test problems, and a study on the applicability of the proposed OFP heuristic on a shale-gas scheduling problem. The chapter consists of the paper Sharma et al. (2014), with preliminary results presented in Knudsen et al. (2014b).
- **Chapter 6** contains concluding remarks, a discussion on connections between the different approaches in the thesis, and recommendations for further work on the considered problems and methods.
- **Appendix A** describes the parameter-estimation procedure for the slab-based proxy model applied in Chapter 4 and 5. The appendix further contains analysis of the spatial grid construction for the proxy model.

***Comments on the chapter outline*:** The thesis is organized such that each chapter is *self-contained*. Each paper is hence included in its original form in the respective chapters, and no parts are removed, except for the bibliography and acknowledgments which are collected in separate parts in the thesis. As such, there will be some repetition in the introduction and model description in some of the chapters.

***Comments on notation*:** The notation in Chapter 2 has been slightly modified from the original publication, Knudsen and Foss (2013), to make the equations and symbols consistent with the remaining chapters. A Boolean Y^d for a certain logical condition, and hence the corresponding binary y^d, may vary between the chapters, as the number and structure of disjunctions in the considered problems are different. Sub- and superscripts are symbolized such that italic references, e.g. m_{jk}, refers to *indices*. All indices are summarized in tables in the respective chapters. Latin letters (upright) are used to *symbolize* the relation of variables and parameters, e.g. p_{t} for tubinghead pressure.

1.5.1 Publications

The following is a list of publications that forms the basis for this thesis:

- Knudsen, B. R., Foss, B., Whitson, C.H., and Conn, A.R. (2012). Target-rate tracking for shale-gas multi-well pads by scheduled shut-ins. In *Proc. of IFAC Symp. on Advanced Control of Chemical Processes*, pages 107–113, Singapore.
- Knudsen, B. R. and Foss, B. (2013). Shut-in based production optimization of shale-gas systems. *Computers & Chemical Engineering*, 58:54–67.
- Knudsen, B. R., Foss, B., Grossmann, I. E., and Gupta, V. (2013). Lagrangian relaxation based production optimization of tight-formation wells. In *Proc. of IFAC Symp. on Dynamics and Control of Process Systems*, pages 147–152, Mumbai, India.
- Knudsen, B. R., Grossmann, I. E., Foss, B., and Conn, A. R. (2014a). Lagrangian relaxation based decomposition for well scheduling in shale-gas systems. *Computers & Chemical Engineering*, 63:234–249.
- Knudsen, B. R., Sharma, S., and Foss, B. (2014b). On MINLP heuristics for solving shale-well scheduling problems. In *Proc. of the 19th IFAC World Congress*, pages 2721–2726, Cape Town, South Africa.
- Knudsen, B. R., Whitson, C. H., and Foss, B. (2014c). Shale-gas scheduling for natural-gas supply in electric power production. *Energy* (accepted).
- Sharma, S., Knudsen, B. R., and Grimstad, B. (2014). Towards an objective feasibility pump for convex MINLPs. *Manuscript submitted for publication (Computational Optimization and Applications)*.
- Knudsen, B. R. and Foss, B. (2014d). On the design and tuning of dynamic shale well proxy models with applications to production optimization. *To be submitted*.

My contributions to the paper Sharma et al. (2014) have been posing the basic idea of the approach, algorithmic developments, construction of the shale-gas case, and a substantial part of the writing.

In addition to the above list of refereed publications, dissemination of results have among other been conducted through the following list of conferences presentations and seminars:

- Knudsen, B. R., Production optimization in shale gas reservoirs. The 6th International Conference on Integrated Operations in the Petroleum Industry, Trondheim, Norway, 2010, *Poster presentation.*
- Knudsen, B. R., Foss, B., and Whitson, C.H., Production optimization in shale gas systems : Reservoir models and problem formulation. International Symposium on Advanced Petroleum Production (ISAPP), Shell International E&P Premises, 2010, *Invited Talk.*
- Knudsen, B. R., Foss, B., and Whitson, C.H., Shut-in based production optimization of shale-gas wells. SPE Shale Gas Forum: Beyond Current Technologies and Going Global, 2011, *Poster Presentation.*
- Knudsen, B. R. and Foss, B., Mixed integer optimization of the latelife performance of shale-gas wells. Presentation at *the 21st International Symposium on Mathematical Programming (ISMP)* held in Berlin, Germany, 2012.
- Knudsen, B. R., Shut-in based optimization of late-life shale-well systems. Enterprise-Wide Optimization (EWO) Meeting, Carnegie Mellon University, September 2012, *Invited Talk.*
- Knudsen, B. R., Shut-in based production optimization of shale-gas systems. Seminar at IBM TJ Watson Research Center, Yorktown Heights, November 2012, *Invited Talk.*

Chapter 2

Shut-in Based Production Optimization of Shale-gas Systems

This chapter consists of the paper Knudsen and Foss (2013):

Knudsen, B. R. and Foss, B. (2013). Shut-in based production optimization of shale-gas systems. *Computers & Chemical Engineering*, 58:54–67.

Abstract

This paper presents a novel operational scheme for enhanced utilization of late-life shale multi-well systems. These systems are characterized by a large number of geographically spread wells and pads, where a substantial number of the wells are producing at low erratic rates due to reservoir pressure depletion and well liquid loading. By applying a cyclic shut-in and production strategy, the scheme avoids well liquid loading and optimizes the production from a set of late-life wells at a shared production pad. The scheduling of well shut-ins is formulated as a generalized disjunctive program (GDP), using a novel shale-gas well and reservoir proxy model. The GDP formulation lends itself both to a complete MILP reformulation and reduced size MINLP reformulations; a computational study indicates in favor of the MILP formulation. We include numerical examples to demonstrate the potential benefit of applying the proposed cyclic scheme compared to a non-optimized approach.

2.1 Introduction

Shale-gas is land-based, unconventional resources of natural gas stored in very tight rock formations. These formations may store enormous amounts of natural gas, and serve both as the source rock and the seal preventing the gas from migrating to the surface. The ability to recover profitable gas volumes from these extremely tight resources relies to a large extent on the drilling of long horizontal wells and stimulation with hydraulic fracturing. Due to pioneering technological developments, the U.S. shale gas production has experienced an average annual growth rate of 48 percent from 2006 to 2010 (EIA, 2013). Shale-gas has consequently increased

rapidly as an energy resource, and is projected to cover 47% of the total volume of domestic dry gas production in the U.S in 2035 (EIA, 2013).

Shale-gas fields are characterized by a large number of wells spread over a wide geographical area. The wells normally share surface equipment and are interconnected by comprehensive pipeline networks as illustrated in Fig. 2.1. There are at least two reasons for the need for many wells in a field. First, there is generally a lack of reliable seismic data in the exploration phase to locate the best and easiest recoverable resources within an organic rich basin. Hence, the initial drilling of wells in a field is often aimed at covering a relatively large geographical area. Seismic data and production data are then collected after an initial production period with possible well testing and used to identify the most promising well locations. Subsequent wells are then drilled next to the most promising wells, creating so-called multi-well pads which share surface production equipment and occupy less surface area than distributed single wells.

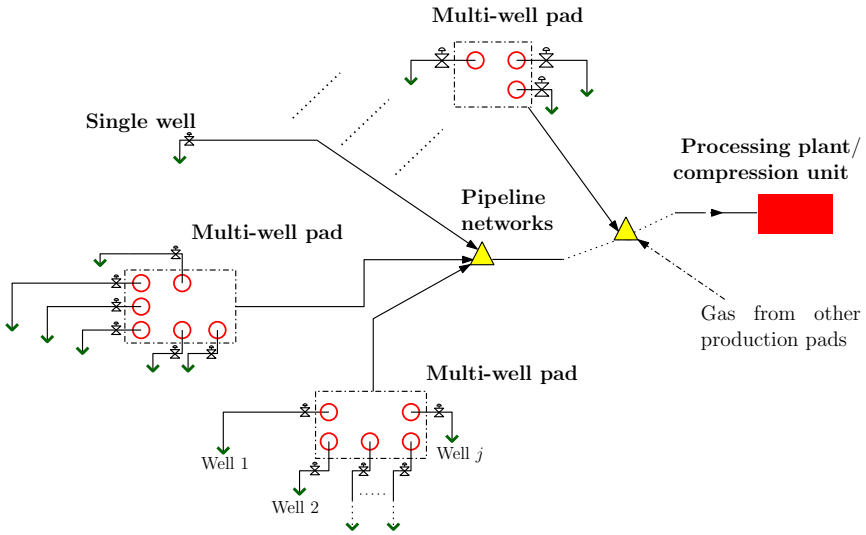


Fig. 2.1: Illustration of shale multi-well systems.

The second reason for many wells is the characteristic production decline profile of shale-gas wells (Awoleke and Lane, 2011; Baihly et al., 2010; Jayakumar et al., 2011; Mayerhofer et al., 2005). This production profile is particularly characteristic for *dry* shale-gas wells, and is seen by an initial peak rate or plateau level for some time (Jenkins and Boyer, 2008), particularly if the well is chocked back, followed by an early and steep decline with subsequent low pseudo steady-state rates. Clearly, to be able to maintain a sustainable total gas supply, then either a large number of wells have to be drilled initially with a clever scheduling of the wells to maintain the gas deliverability over time, or an extensive drilling program will have to be applied to prevent the total gas supply to drop to unacceptable levels.

The gas produced from dry and semi-dry shale-gas wells normally contains small quantities of liquids which are separated in tanks at the wellhead. The source of the liquids may be some low water saturations in the reservoir, gas condensates or left-

over water from the hydraulic fracturing. As the pressure in the reservoir depletes, the wells' ability to lift accumulated water or condensate from the bottomhole to the surface reduces, with a following increase in the hydrostatic backpressure in the wellbore. Eventually, there will be insufficient pressure to lift the accumulated or co-produced liquids to the surface, and the well will exhibit liquid loading. This state can be recognized by erratic and unstable rates, sharp drops in the decline curve (Lea and Nickens, 2004), or stabilization of the production rate at significantly lower levels (Whitson et al., 2012; Dousi et al., 2006), so-called meta-stable rates. Liquid loading is frequently observed at late and intermediate times for shale-gas wells (Al Ahmadi et al., 2010; Awoleke and Lane, 2011; Cipolla et al., 2010; Kravits et al., 2011; Ilk et al., 2008; Mayerhofer et al., 2005; Sutton et al., 2010), and will additionally worsen the already low late-life rates. Furthermore, operating multiple wells under liquid loading is generally undesirable, since it may give rise to sudden liquid slugs at the surface (Lea and Nickens, 2004) and thereby very unstable rates.

The common operational practice for shale-gas wells exhibiting liquid loading during low late life varies between fields and operators. Some wells are merely abandoned or left producing with erratic rates. If the wells are initially completed with casing only, then some operators choose to retrofit the wells with tubings. This may in some cases increase the well productivity significantly for months or up to a couple of years. Another practice is to install an artificial lift system, where a plunger lift system is the most commonly applied (Kravits et al., 2011). This technique is often applied to wells producing high water rates initially. A special property of fractured tight formation wells and shale-gas wells is that they can be shut-in for some time without losing significant cumulative production (Rahmawati et al., 2009; Whitson et al., 2012). This property can be utilized to avoid liquid loading, since shut-ins lead to pressure build-up in the wellbore and in the near-well region of the reservoir (Whitson et al., 2012; Knudsen et al., 2012). Shut-ins will further boost the late-life rate for a certain time once the well starts producing. The increasing use of multi-well pads (Stefik and Paulson, 2011; Harpel et al., 2012), however, requires careful scheduling of shut-ins in order to avoid multiple simultaneous shut-ins and start-ups of wells on a pad. This may be crucial for keeping the production within the constraints of the gathering system, and for meeting the total gas demand. The novel operational approach we therefore propose, is to schedule shut-ins for a given set of wells on a single production pad as illustrated in Fig. 2.1, so as to maintain stable total production while preventing liquid loading and thereby improving the utilization of late-life wells.

Cyclic shut-ins of shale-gas wells belong to the class of switched dynamical systems. These systems are also referred to as hybrid systems or discrete-continuous dynamical systems, and include systems with inherent logics, disjunctive formulations and other nonsmooth dynamics such as Coulomb friction. Several solution strategies for optimization of switching decisions in dynamical systems exist. Baumrucker and Biegler (2009) present a complementarity formulation for optimization of a class of hybrid dynamic systems, embedded into a nonlinear program (NLP) through an ℓ_1 penalty formulation. Although efficient on many problems, the complementarity condition introduces an inherent nonconvexity in the NLP and fails to satisfy constraint qualifications (Baumrucker and Biegler, 2009). These properties may restrict the applicability of the complementarity formulations to systems with

proper disjunctions, since the dependence on good initial NLP starting points may be strong. Sager (2009) provides an overview of solution methods for optimal control of switching decisions in nonlinear dynamical systems, including methods for direct switching time optimization. Such formulations may be practical for some problems, while the major drawback is that the number of switches normally must be assumed to be known a priori, as well as numerical issues related to congruent switching times. A more flexible and intuitive strategy is to formulate switching decisions as logical constraints using generalized disjunctive programming (GDP) (Raman and Grossmann, 1994). Using this formulation as a higher level modeling framework, the problem can be reformulated as a mixed integer dynamic optimization (MIDO) problem and solved using a mixed integer approach (Avraam et al., 1998; Barton and Lee, 2004; Bemporad and Morari, 1999). Oldenburg and Marquardt (2008) presents a combination of the GDP and the mixed integer approach by formulating a mixed logic dynamic optimization problem for linear hybrid systems.

The ability of efficiently solving mixed integer nonlinear programs (MINLP) depends inherently on the problem size, the number of integer variables, and the nonlinearities. A wide range of nonconvex MINLP can be reformulated to convex MINLPs or to mixed integer linear programs (MILP) by using exact or approximate reformulation or linearization techniques. One such example is MINLPs where the nonconvexities are separable and univariate, where a commonly applied technique is the use of piecewise linear approximations, eventually transforming the MINLP into an approximate MILP. In contrast, there are several nonconvex functions such as complementarity constraints and bilinear binary-continuous terms, that may be exactly reformulated or linearized, see e.g. Glover (1975). Both approximate and exact reformulations lead to additional constraints and auxiliary variables, in many cases additional binary variables. Different reformulation techniques for a given (nonconvex) nonlinearity therefore render a trade-off between the increase in the problem size and the possible loss of accuracy of the model. For nonlinear MIDO problems, and in particular those involving partial differential equations (PDE), the problem size increases rapidly with increased prediction horizons and finer spatial and temporal discretization. Hence, for these problems, the computational gain of reformulating nonlinearities depends strongly on the problem structure and the choice of the discretization intervals.

This paper focuses on short and midterm scheduling or production optimization for shale-gas multi-well pads. To efficiently solve shale-well planning problems with discrete decisions and dynamic models, it is necessary to develop well and reservoir models that are limited in size but captures the essential dynamics and expected performance of these wells. Hence, in Section 2.2 we derive a novel PDE-based shale-well proxy model, and in Section 2.3 show how certain model parameters can be tuned according to the goal of the proxy model. In Section 2.4 we embed the proxy model in a GDP formulation for finding optimal shut-in and start-up times. By reformulating the GDP to a nonsmooth MINLP, we then apply a set of smooth approximations for the nonsmooth functions, and subsequently develop an exact reformulation rendering an MILP formulation. In Section 2.5, we evaluate the efficiency of these formulations for increasing problem sizes, and demonstrate the potential of the proposed operational approach through numerical examples.

The paper ends with a discussion and concluding remarks in Section 2.6 and 2.7, respectively.

2.2 Shale-gas reservoir modeling

Storage and transport of gas in organic rich shale reservoirs is complex and consists of several mechanisms. The gas is stored both as free gas in porous space and natural fractures in the shale, but also as absorbed gas on the internal organic material (Jenkins and Boyer, 2008; Cipolla et al., 2010). Once a well starts producing and the fracturing water returns, the gas stored in the natural fracture networks and the adjacent pores will be released and flows into the network of interconnecting, hydraulically induced fractures. This sudden release of stored gas causes the high initial peak, with the following decline and low rates caused by the inherently low permeability of shale. Gas desorption, that is, release of absorbed gas on the internal organic material, generally only impacts the ultimate recovery of shale wells (Cipolla et al., 2010), and is a slow process that only occurs at very late times. Even though absorbed gas may account for a large part of the total organic content, the release of this gas critically depends on the reservoir characteristics and well completions. Occasionally absorbed gas may not be produced at all.

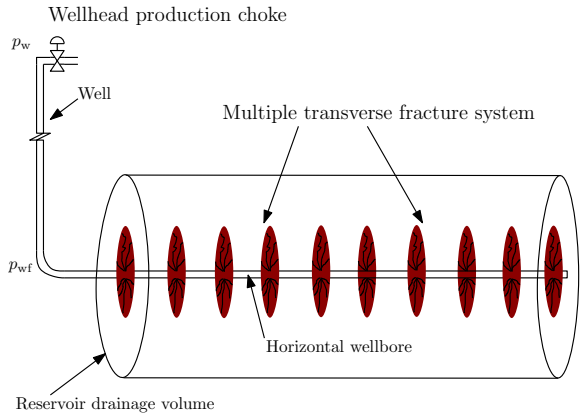


Fig. 2.2: Illustration of an idealized *multi-fractured* horizontal reservoir model with cylindrical geometry.

Forecasting and prediction of shale-gas production can be divided into two categories. The first is based on detailed numerical reservoir models. Fractured shale and tight gas reservoirs are modeled either as variations of the dual-porosity idealization introduced by Warren and Root (1963), as discrete fracture models (Karimi-Farad et al., 2004), or as fully discretized dual-permeability models (Cipolla et al., 2010). These models have been constructed both with Cartesian (Cipolla et al., 2010) and cylindrical geometries (Economides and Wang, 2010) as illustrated in Fig. 2.2, as well as flow in multiple dimensions. All these modeling schemes have different merits and drawbacks in terms of the ability to model complex fracture networks and to capture the dominating dynamics of the reservoir gas flow. Fur-

thermore, the numerical efficiency of these models varies significantly, and they typically come as black-box simulator models. Hence, using these detailed, high-fidelity models for production optimization of multi-well systems limits the possible number of decisions in optimization schemes.

A popular and commonly applied technique for production forecasting of shale-gas wells is to use simple semi-analytical or empirical models (Anderson et al., 2010; Al Ahmadi et al., 2010; Bello and Wattenbarger, 2010; Ilk et al., 2008). These techniques are based on interpolation of production data, and often give good correlation for a certain range of the available data. However, due to the wast variations encountered in shale gas reservoirs from the hydraulic fracturing and inherent heterogeneities, the techniques often loose accuracy in long-term production forecasting. Most importantly, the models assume continuous production at constant bottomhole pressure, and the models are therefore not suited for optimization of a cyclic production scheme.

With the objective of optimizing multi-well shut-in and start-up times, we derive a single shale-well *proxy model* aimed to capture the essential near-wellbore reservoir dynamics during switched operations. The proxy model is constructed as a radial composite, single-porosity dual-permeability model with cylindrical geometry as sketched in Fig. 2.3. The idea behind the proxy model is to represent the near-wellbore, highly conductive fracture network rising from the hydraulic fracturing by a compact "crushed rock" region. This concept has similarities with the shale-well representation recently proposed by Siripatrachai and Ertekin (2012). Between the high permeability propped fracture zone and the low permeability, unstimulated outer region, we include an intermediate transition region with some average properties. By using this three-region representation and estimating certain model parameters, the goal is to obtain a simple model replicating the multiple transverse fracture model in Fig. 2.2 during short time scales, with rapid pressure build-up and drawdown due to alternating shut-ins and start-ups of the well.

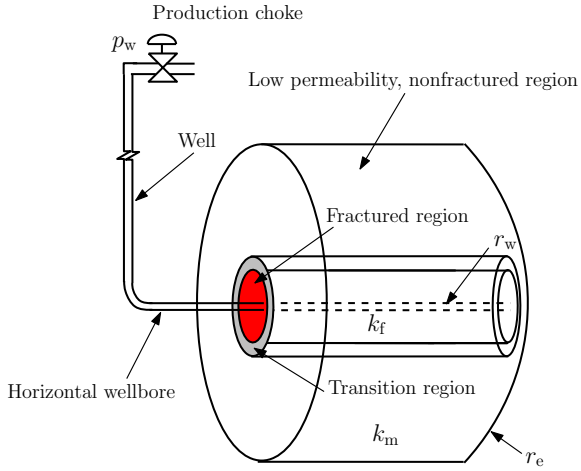


Fig. 2.3: Illustration of proxy model

We assume the gas to be dry (single phase), and any small amount of residual

water is hence negligible. Considering the time scale of the optimization, we do not include fracture unconsolidation or gas desorption. Assuming Darcy's law, using the law of mass conservation and the equation of state for real gases, the nonlinear PDE for the gas flow in the radial proxy model is given by

$$\phi \frac{p}{Z(p)} c(p) \frac{\partial p}{\partial t} = \frac{1}{r} \frac{\partial}{\partial r} \left(k(r) \frac{p}{\mu(p) Z(p)} r \frac{\partial p}{\partial r} \right), \tag{2.1}$$

where p is the pressure, $k(r)$ is a radially dependent permeability, $\mu(p)$ is the gas viscosity, $Z(p)$ is the gas compressibility factor and $c(p)$ is the total compressibility. By using Neumann boundary conditions with a producing well (sink) in the center and no-flow conditions at the outer boundary, we get the boundary conditions

$$r \frac{\partial p}{\partial r}\bigg|_{r_\mathrm{w}} = \frac{q B_\mathrm{g} \mu}{2\pi h k}, \tag{2.2a}$$

$$\frac{\partial p}{\partial r}\bigg|_{r_\mathrm{e}} = 0, \tag{2.2b}$$

where B_g is the gas volume formation factor at pressure p and temperature T, given by (Whitson and Brule, 2000)

$$B_\mathrm{g} = \frac{Z(p) T p_\mathrm{sc}}{p T_\mathrm{sc}}. \tag{2.3}$$

The subscript sc refers to evaluation at standard condition, 1 bar and 15.6°C. All gas volumes and volumetric rates are given in standard conditions in the rest of the paper. The pressure dependency of $\mu(p)$ and $Z(p)$ can be efficiently handled by the variable transformation known as *pseudopressure* (Al-Hussainy et al., 1966),

$$m(p) = 2 \int_0^p \frac{p'}{\mu(p') Z(p')} dp'. \tag{2.4}$$

Using (2.3) and the identity

$$\frac{\partial m}{\partial t} = \frac{\partial m}{\partial p} \frac{\partial p}{\partial t} = 2 \frac{p}{\mu Z} \frac{\partial p}{\partial t}, \tag{2.5}$$

with a similar expression for $\partial m / \partial r$, the PDE (2.1) with the boundary conditions (2.2) and a given initial reservoir pressure can be reformulated as an initial boundary-value problem (IBVP) in terms of the pseudopressure variable $m(p)$,

$$\phi \mu(p) c(p) \frac{\partial m}{\partial t} = \frac{1}{r} \frac{\partial}{\partial r} \left(k(r) r \frac{\partial m}{\partial r} \right), \tag{2.6a}$$

$$r \frac{\partial m}{\partial r}\bigg|_{r_\mathrm{w}} = q \frac{T p_\mathrm{sc}}{T_\mathrm{sc} \pi h k}, \tag{2.6b}$$

$$\frac{\partial m}{\partial r}\bigg|_{r_\mathrm{e}} = 0, \tag{2.6c}$$

$$m(r, 0) = m^\mathrm{init}. \tag{2.6d}$$

The above IBVP for the parabolic PDE (2.6a) is still nonlinear because of the pressure dependency of the product $\mu(p)c(p)$. The treatment of this product, and in particular the pressure variations of the compressibility $c(p)$ is generally challenging when using the pseudopressure formulation. A commonly applied technique is to evaluate this product either at initial conditions (Al-Hussainy et al., 1966), or at a given pressure between the initial reservoir pressure and the bottomhole pressure. We use a similar technique, but select the pressure to evaluate $c(p)$ at through the tuning of the proxy model against a reference model. Hence, the nonlinear PDE (2.6a) is converted to a semi-linear PDE, since the permeability $k(r)$ is assumed to be independent of the reservoir pressure. The initial pseudopressure (2.6d) is calculated by trapezoidal integration of the corresponding values of p, μ and Z which are obtained by gas correlations, see Whitson and Brule (2000). Consequently, all reservoir dynamics are calculated in terms of pseudopressure $m(p)$ instead of pressure p.

The IBVP (2.6) is discretized using a finite difference scheme. This scheme is described in Knudsen et al. (2012). Spatial discretization applies central difference approximations, i.e. with a second order accuracy. The time discretization uses the backward Euler scheme, or equivalently, a first order orthogonal collocation on finite elements with fixed element lengths (Biegler, 2010). The discretization leads to the following linear representation of the reservoir dynamics for one well.

$$Am_{k+1} = m_k + Bq_{k+1}, \tag{2.7a}$$

$$m_0 = m^{\text{init}}, \tag{2.7b}$$

m_k is a vector containing the grid pseudopressures in each grid block, A is a tridiagonal matrix and q is the scalar gas rate flowing into the well from the reservoir. The implicit scheme secures the numerical stability of the model.

2.2.1 Well model

The coupling between a reservoir model and a well model may be formulated in several different ways depending on the application, but should nevertheless represent the essential expected features of the well performance. Usually, a desired rate is specified with constraints on the bottomhole or wellhead pressure (Abou-Kassem et al., 2006), derived from surface boundary conditions such as minimum manifold or wellhead pressure (Chen and Baldick, 2007). The well will then normally be able to deliver the constant rate initially and for a short time after shut-ins. Once the wellhead pressure hits the lower bound of allowed pressures, the rate starts to decline with a constant wellhead pressure equal to the specified minimum constraint.

It is common practice when connecting well and wellbore models to reservoir models, to specify a constant wellhead pressure and connect the reservoir model to a steady-state wellbore model and a well inflow model. For gas wells, the steady-state wellbore model can be derived using a mechanical energy balance (Katz and Lee, 1990), which leads to a second order polynomial of the gas rate and the wellhead and bottomhole pressure. This generally requires a nonlinear, possibly piecewise mapping of pressure p to pseudopressure $m(p)$ to cover the entire pressure range

of the well. To circumvent this problem and to include the basic expected features of the well performance described above, we use an aggregated well inflow and wellbore model

$$q_k = \min\left\{q^{\max}, \beta\left(m_{k1} - m(e^S p_{\mathrm{w}})\right)\right\}, \tag{2.8}$$

where $q^{\max}$ is the specified desired rate, β is a skin dependent well index, m_{k1} is the pseudopressure in the innermost grid block, S is a tubing dependent constant and p_{w} is a specified constant wellhead pressure. The maximum rate also compensates for the model's lack of pressure drop due to wellbore friction, while the term $e^S p_{\mathrm{w}}$ yields the static vertical wellbore pressure (Katz and Lee, 1990).

2.3 Fitting parameters of the proxy model

Table 2.1: Proxy model parameter values.

Symbol	Parameter	Value	Unit
ϕ	porosity	6	%
k_{m}	matrix permeability	3.5×10^{-4}	mD
k_{f}	fracture permeability	100	mD
z_{w}	true vertical depth	2300	m
r_{e}	radial extent	564	m
p_{r}	initial reservoir pressure	2×10^7	Pa
p_{w}	wellhead pressure	6.9×10^5	Pa
$q^{\max}$	maximum rate	4×10^4	m^3/d
I	number of grid blocks	4	-

Proxy models are constructed to model certain parts of a more complex system, and always represent a trade-off between accuracy, complexity and numerical efficiency. The proxy model in Section 2.2 is constructed based on two objectives. The first is to capture the dominating dynamics during well shut-ins and start-ups at intermediate and late times of the well. The second is to be computationally simple enough to be included in a multi-well optimization problem, without sacrificing accuracy in the frequency range of interest. To meet these two objectives, we use a very coarse spatial discretization, and then tune this model realization using a multiple transverse fracture model as in Fig. 2.2 as a reference model. Proxy model parameters are given in Table 2.1.

The high-fidelity multi-fracture model is constructed with cylindrical geometry using ten equally spaced fractures, where each fracture is perpendicular to the horizontal wellbore. Note that a similar reference model with Cartesian geometry of a pure planar fracture model without interconnecting network fractures has essentially the same performance as the chosen reference model in Fig. 2.2. The high-fidelity multi-fracture model is implemented in the state-of-the-art simulation software SENSOR (SENSOR, 2011), using logarithmic grid refinements around the fractures with more than 2500 grid cells in total, and an added static wellbore model. The reference model geometry with multiple transverse planar fractures is

naturally an idealization of the true geometry in stimulated shale-gas wells. Nevertheless, this is one of the most commonly applied shale-well fracture geometries for long-term production forecasting (Al Ahmadi et al., 2010; Bello and Wattenbarger, 2010; Jayakumar et al., 2011; Economides and Wang, 2010; Anderson et al., 2010; Siripatrachai and Ertekin, 2012), and is shown through matching of production data and evaluation of seismic data to capture the most prominent fracture geometry for many shale-gas wells.

The proxy model in Section 2.2 was used in Knudsen et al. (2012) with 12 instead of 4 grid blocks as shown in Table 2.1. However, the pressure profile in the grid blocks in the low permeability region was observed to be almost identical. In addition, due to very large time constants, the pressure in the outermost grid block was approximately unchanged. Hence, to reduce the size of proxy model, only four spatial grid blocks are used. We apply logarithmic ordering of the three outermost grid blocks as in Knudsen et al. (2012), but define the position r_1 of the innermost high permeability grid block as a tuning variable with a lower and upper bound. The remaining tuning parameters are the horizontal well length h, the permeabilities $\tilde{k}_2$ and $\tilde{k}_3$ in the two grid blocks in the transition region and the vector of compressibilities c^T in the four grid blocks. For simplicity, the radial and the fracture extent are equal for the two models. Let

$$\theta = \begin{bmatrix} r_1 & \tilde{k}_2 & \tilde{k}_3 & c^T & h \end{bmatrix}^T \tag{2.9}$$

be the vector of unknown parameters. To fit these parameters in the proxy model, we solve the constrained weighted least-squares problem

$$\min_{\theta} \sum_{k \in \hat{\mathcal{K}}} \omega_k \left(q_k^{\mathrm{MFR}} - q_k(\theta) \right)^2 + \tilde{\omega} \left(q_{\mathrm{cum}}^{\mathrm{MFR}} - q_{\mathrm{cum}}(\theta) \right)^2, \tag{2.10}$$

with the bounds

$$\theta^{\mathrm{lo}} \leq \theta \leq \theta^{\mathrm{up}} \tag{2.11}$$

on the parameter vector θ. In (2.10), q^{MFR} is the gas rate from the multi-fracture reference model (MFR). The weights ω_k are selected in order to match high frequency content of the reference model at intermediate and late times. The tuning problem (2.10)–(2.11) is as a nonsmooth (or nondifferentiable) NLP, due to the nonsmooth well model (2.8) defining the gas rate q_k for the proxy model. To make to problem differentiable at all points, we use a smooth square-root approximation of (2.8). This is detailed in Section 2.4.1. The last term in (2.10) is added to penalize deviations in cumulative production q_{cum} over the estimation horizon $\hat{\mathcal{K}}$. All of the parameters in θ appear nonlinearly in the reservoir model and the well model, see Knudsen et al. (2012) for details on these model equations. Inverse problems of the type (2.10) are generally ill-posed, and may lead to linear combinations of the parameters in (2.9) giving the same, possibly sub-optimal solution. The bounds (2.11) on θ, added from *a priori* physical evaluation of the estimated parameters, limits the solution space and hence possible combinations of the parameters in θ at the optimum. These bounds serve as a regularization term in a similar way as adding a Tikhonov regularization term $||\theta - \theta_{\mathrm{init}}||_P^2$ to the objective function for some positive definite matrix P and initial guess θ_{init} (Hansen, 1998). Weighted

least square problems are easy to construct to emphasize certain features of the solution. There is, however, no unique way of selecting the weights. We select the weight ω_k three times higher during transients than during the pseudo steady-state rate, while the weight $\tilde{\omega}$ is select much smaller than ω_k since the magnitude of q_{cum} is several orders higher than q_k.

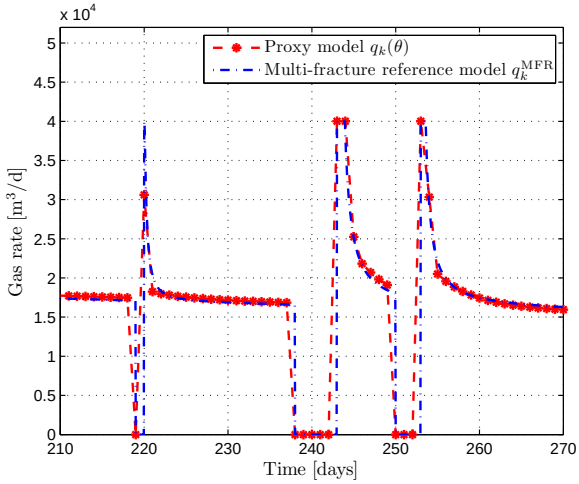


Fig. 2.4: Matching of the proxy model against the multi-fracture reference model (MFR). A particular good match is seen in the transients following the two latest shut-ins.

The result of the model fitting is shown in Fig. 2.4. The timestep for the proxy model is $\Delta k = 1$ day, while the timestep for the reference model is set to 1 hour, i.e., each 24th element is preselected for the least-squares regression. Decreasing the timestep of the reference model additionally did not increase the duration of the plateau rate after shut-ins significantly, and it is therefore assumed that the selected timestep is small enough to capture the high frequency content of the multi-fracture reference model. The match of the two latter shut-ins of five and three days, respectively, is observed to give better match than the one-day shut in. For the one day shut-in, the proxy model never reaches the maximum rate $q^{\max}$. The match in the pseudo steady-state rate is quite good following the three day shut-in at the end of the horizon, while there is a slightly higher discrepancy in the rates after the one day shut-in. The error in cumulative production between the two models is less than 10^{-3} m^3 over the horizon $\hat{\mathcal{K}}$. Increasing the weight decreases the cumulative error term in (2.10) marginally, while both the peak rate at day 220 increases and the discrepancy in the subsequent pseudo steady-state rate increases.

2.3.1 Cross-validation and practical considerations

The tuning of the proxy model is based on preselecting a fixed set of shut-in times for the reference model, and then solve the least-squares problem over a horizon $\hat{\mathcal{K}}$. Consequently, the tuning will depend on the chosen set of shut-in times, the initial reservoir pressure at the beginning of the horizon, as well as the length of

the tuning horizon. To evaluate the importance of the choice of shut-in times and initial conditions, we cross-validate the model with different data sets. The first, shown in Fig. 2.5, has a different initial pressure, but equal shut-in times. The match of the peak after the one day shut-in is still poor, while there is a little mismatch in the successive pseudo steady-state rate. The transients after the three and five day shut-ins match well with the reference model, but the discrepancy grows as the rates converge towards the pseudo steady-state rate.

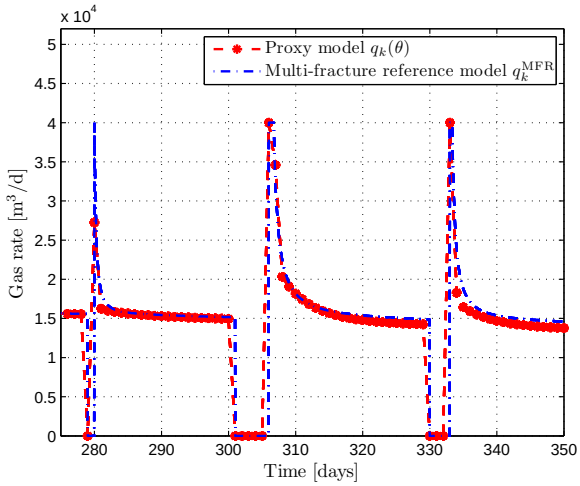


Fig. 2.5: Cross validation of proxy model using later and different times for shut-ins, but equal shut-in times to tuning.

We also perform cross validation with both different shut-in times and different initial pressure in Fig. 2.6. The rate decline following the peak from the two day shut-in is too steep as; similar to the one day shut-in in Fig. 2.5. This observed model mismatch may to some extent be related to the choice of a one day timestep in the proxy model compared to the one hour step size in the reference model. The match of the four and six day shut-ins is good, although the discrepancy between the models grows with time, particularly after the last shut-in. If this discrepancy persists for later shut-ins, then it may be necessary to re-tune the proxy model, depending on the lengths of the production times between the shut-ins.

In practice, if a well has been shut-in at times before the possible onset of liquid loading, then the tuning of the proxy model would be performed using these production data. The exceptions are wells with very early shut-ins, when the near wellbore pressure is still high or there is residual water from the hydraulic fracturing which requires clean up. Residual fracturing water may generally cause very erratic and unpredictable rates in the initial well production. Moreover, if shut-ins have been applied due to hydraulic fracturing of nearby wells, then fracture interference effects may occur with a possible shift of the pseudo steady-state rate. If a well has been producing continuously up to the time when a cyclic shut-in scheme is to be applied, then the tuning of the proxy can be performed as in Section 2.3, by running shut-in simulations on a high-fidelity reference model. This model would be

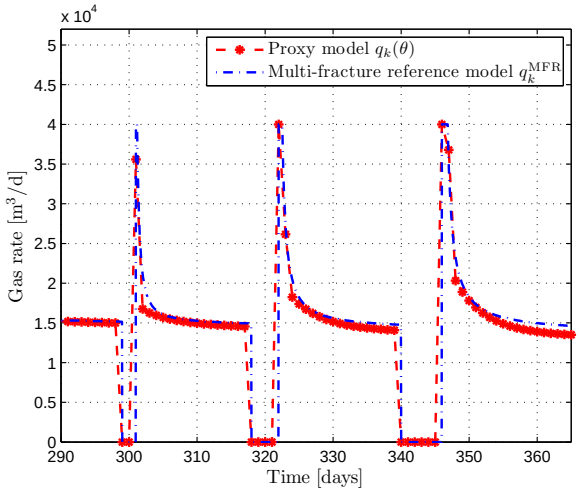


Fig. 2.6: Cross validation of the proxy model using both a later initial time for the first shut-in and different shut-in times than during tuning.

constructed from available geometrical data, and through matching of production data using one of the techniques described in for instance Al Ahmadi et al. (2010); Bello and Wattenbarger (2010); Anderson et al. (2010); Mayerhofer et al. (2005) to estimate certain reservoir parameters such as the permeability distribution and the fracture lengths.

Matching both drawdown and build-up pressure is generally difficult without incorporating pressure and time varying reservoir properties in the model (Jayakumar et al., 2011). Furthermore, the stimulation with hydraulic fracturing causes complex dynamic behavior initially and during the first major rate decline. Hence, there will naturally be a trade-off between matching the model initially with fast pressure dynamics and depletion of the fracture volume, and matching the reservoir pressure build-up during shut-ins. Since we emphasize to match transients at intermediate and late times, then the proxy model is not suited for long-term production forecasting. For this purpose, a multi-fracture simulator model similar to the reference model should instead be used.

2.3.2 Critical rate

Several papers have suggested models and methods for predicting the onset of liquid loading in gas wells. The most widely used model, sometimes referred to as the Turner rate Turner et al. (1969), is based on a spherical droplet entrained in the flowing gas. By balancing the gravitational and drag forces on a free falling droplet in the fluid medium, the critical velocity v_{gc} is given by

$$v_{\mathrm{gc}} = \kappa \frac{\left[\sigma\left(\rho_{\mathrm{l}} - \rho_{\mathrm{g}}\right)\right]^{0.25}}{\rho_{\mathrm{g}}^{0.5}}, \tag{2.12}$$

where ρ_{l} is the liquid density, ρ_{g} is the gas density, σ is the interfacial tension and κ is a constant. The onset of liquid loading is in most instances controlled by the wellhead conditions Turner et al. (1969); Coleman et al. (1991). Hence, by using the real gas law and evaluating the temperature, the compressibility factor and the pressure at wellhead conditions, the minimum rate necessary to ensure continuous removal of liquids in the wellbore is given by

$$q_{\mathrm{gc}}(p_{\mathrm{w}}) = C_{\mathrm{gc}} A_{\mathrm{t}} p_{\mathrm{w}}\, v_{\mathrm{gc}}(p_{\mathrm{w}}), \tag{2.13}$$

where A_{t} is the cross-section area of the tubing and C_{gc} is a constant. It is assumed that condensed water dominates the liquid accumulation, thus using the numerical value $\sigma = 0.02$ N/m (Turner et al., 1969). The model (2.12)- (2.13) for the critical rate is accepted to be a good estimate for the possible onset of liquid loading in gas wells due to good correlation with observed field data. The constant κ has been evaluated against field data in several papers, and the choice of its value depends on the droplet model used and the wellhead conditions for the wells against which the model is validated. Most of the wells used in the comparison in Turner et al. (1969) have wellhead flowing pressures above 34.5 bar, while Coleman et al. (1991) compared the model against wells with pressures below 34.5 bar and reduced the value of κ to 1.59 to match the observations. Since shale-gas wells require low wellhead pressures, we use the value of $\kappa = 1.59$ suggested by Coleman et al. (1991).

By using (2.13) as a lower bound on the flowing gas rate, we assure that the vertical part of the well stays unloaded with liquids, and hence avoid the need to use a multiphase model for a low liquid system like dry shale-gas wells.

2.4 Formulation of the production optimization problem

Optimization of upstream petroleum production is performed at many different levels of the value chain, and depends on the system topology, the available control and monitoring equipment, and the location of the system to be optimized. Conventionally, a multi-level hierarchy is used to organize the optimization of petroleum systems, where each level involves optimization of some objective over different decision horizons (Foss, 2012). The problem we study in this paper is shut-in based optimization of shale-gas wells over horizons spanning from weeks to a couple of months. Hence, the problem is categorized as a short to midterm production optimization problem. These problems typically aim at maximizing daily production rates or keeping production at some prespecified target or reference rate (Foss, 2012). Due to a large number of wells producing to a shared processing and control unit, cf. Fig. 2.1, we will assume that a *reference rate* q^{Ref} for a given horizon K has been *assigned* to the production pad. This reference rate would typically be assigned by some planning activity, based on a distribution of the total demand in a gas sales contract. The objective for the multi-well pad is then to produce as close as possible to the reference rate, that is, to track the reference rate.

Each well is assumed to operate on a fixed wellhead pressure $p_{\mathrm{w},j}$, and the downstream boundary is the pipeline leaving the pad. The separation tank at the pad is not included in the system description. This separation will only impact

Table 2.2: Set definition

Index	Interpretation	Set	Elements
k	discrete time	$\mathcal{K}$	$\{1 \dots K\}$
j	well number	$\mathcal{J}$	$\{1 \dots J\}$
i	spatial grid block	$\mathcal{I}$	$\{1 \dots I\}$

the system by slightly lowering the inlet pressure on the pipeline leaving the pad. The scheduling of individual well shut-ins and start-ups (i.e. switchings) is due to the following disjunction: either the well is producing at a rate higher than the critical rate q_{gc}, or the well must be shut in. To represent this disjunction, we define Boolean variables

$$Y_{jk} \in \{\text{True}, \text{False}\}, \tag{2.14}$$

for each well $j \in \mathcal{J}$ and for each timestep $k \in \mathcal{K}$. By using a Boolean representation of the aforementioned disjunction, we can, in an intuitive and flexible way, formulate the production optimization problem as a GDP. This allows for a structured evaluation of possible logic-based and mixed integer formulations. Moreover, as will be shown below, it simplifies the construction of logical constraints on the disjunctions themselves.

For many shale-well systems, it may be desirable to avoid too frequent shut-ins and start-ups of the wells. There may be several reasons to this. Limiting the number of switchings reduces possible wear and tear of the surface equipment such as the wellhead chokes and measurement devices. In other cases it may be a requirement by the well operators that the wells have to stay in a given production mode for a certain time for them to be able to monitor production adequately. Lastly, some wells may require a minimum shut-in time to achieve sufficient pressure build-up before they are re-opened. This may often be the case if the well exhibits signs of liquid loading before shut-in. The formulation of these types of constraints can be derived by using propositional logics (Raman and Grossmann, 1991). If we require a well j to be shut in for at least $\tau_1 \geq 2$ timesteps each time there is a transition from production to shut-in mode, we can define the logic constraint

$$(Y_{jk-1} \wedge \neg Y_{jk}) \Rightarrow \neg Y_{jk+1} \wedge \neg Y_{jk+2} \wedge \ldots \wedge \neg Y_{jk+\tau_1-1}. \tag{2.15}$$

Similarly, we can require that each well must be producing for at least $\tau_2 \geq 2$ timesteps between each succeeding shut-in by the logic constraints

$$(Y_{jk} \wedge \neg Y_{jk-1}) \Rightarrow Y_{jk+1} \wedge Y_{jk+2} \wedge \ldots \wedge Y_{jk+\tau_2-1}. \tag{2.16}$$

Using the objective described above, we can formulated the problem of finding optimal switching times for $\mathcal{J}$ wells as the following GDP:

$$\min \sum_{k\in\mathcal{K}} \left| q_k^{\mathrm{Ref}} - q_k^{\mathrm{Tot}} \right| \tag{2.17a}$$

s.t.

$$q_k^{\mathrm{Tot}} = \sum_{j\in\mathcal{J}} q_{jk}, \qquad \forall k \in \mathcal{K} \tag{2.17b}$$

$$A_j m_{jk+1} = m_{jk} + B_j q_{jk+1}, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K}^{\mathrm{m}} \tag{2.17c}$$

$$m_{j0} = m_j^{\mathrm{init}}, \qquad \forall j \in \mathcal{J} \tag{2.17d}$$

$$\begin{bmatrix} Y_{jk} \\ q_{jk} = \min\left(q_j^{\max}, \tilde{q}_{jk}\right) \\ \tilde{q}_{jk} = \beta_j \left(m_{jk1} - m\left(e_j^S p_{\mathrm{w},j}\right)\right) \\ q_{jk} \geq q_{\mathrm{gc}} \end{bmatrix} \vee \begin{bmatrix} \neg Y_{jk} \\ q_{jk} = 0 \end{bmatrix}, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K} \tag{2.17e}$$

$$(Y_{jk-1} \wedge \neg Y_{jk}) \Rightarrow \neg Y_{jk+1} \wedge \neg Y_{jk+2} \wedge \ldots \wedge \neg Y_{jk+\tau_1-1}, \qquad \forall j \in \mathcal{J}, k \in \{\mathcal{K} : k \leq K - \tau_1 + 1\} \tag{2.17f}$$

$$(Y_{jk} \wedge \neg Y_{jk-1}) \Rightarrow Y_{jk+1} \wedge Y_{jk+2} \wedge \ldots \wedge Y_{jk+\tau_2-1}, \qquad \forall j \in \mathcal{J}, k \in \{\mathcal{K} : k \leq K - \tau_2 + 1\} \tag{2.17g}$$

where $\mathcal{K}^{\mathrm{m}}$ is the set $\mathcal{K}$ shifted one step, i.e. $\mathcal{K}^{\mathrm{m}} = \{0, \ldots, K-1\}$. Note that m_{jk+1} refers to the vector of pseudopressures m for well j at timestep $k+1$, while m_{jk1} refers to the pseudopressure for well j at timestep k in grid block $i = 1$. To impose the shut-in and start-up constraints (2.17f)–(2.17g) also for the first timestep, $k = 1$, we infer the state Y_{jk} of each well at the preceding timestep of the prediction horizon.

The optimization problem (2.17) is a nonconvex nonlinear GDP. Furthermore, the problem is nonsmooth (or nondifferentiable) due to the 1-norm in the objective function and the min-function in the disjunction (2.17e). Convex nonlinear GDP problems can be solved by logic based methods such as the disjunctive branch-and-bound method (Lee and Grossmann, 2000) and the logic-based outer-approximation method (Türkay and Grossmann, 1996), or they can be reformulated and solved as an MINLP by formulating algebraic constraints of the disjunctions using either the big-M technique or a nonlinear convex hull relaxation (Lee and Grossmann, 2000). The presence of nonconvexities in the GDP problem leads to several difficulties. As for nonconvex NLPs and MINLPs, the fundamental difficulty lies in finding the global optimum. Solution approaches are therefore classified either as local methods or global methods. For the latter, the problem lies in finding tight relaxations as a means of predicting valid lower bounds (for minimization problems). The general purpose method for solving nonconvex GDPs to global optimality is a two-level branch and bound method proposed by Lee and Grossmann (2001). This procedure consists of deriving convex underestimators of the nonconvex functions in the GDP, and using a subsequent hull-relaxation of the resulting convex GDP in a spatial brand-and-bound framework. The method is shown to

work well for a large range of nonconvex GDP problems, but may be computationally intensive for larger problems like (2.17). For these types of problems, there is a trade-off between the time spent on searching for the global optimum compared to accepting a solution from a local solver which is within an acceptable value of the objective (2.17a) and solved within reasonable CPU time.

Instead of using the general approach in Lee and Grossmann (2001), we first reformulate the nonconvex GDP (2.17) as a nonsmooth MINLP, and then show how this problem can reformulated either as a smooth nonconvex MINLP, or as an MILP by an increase in the combinatorial part. For each disjunction pair (j,k), we define binary variables

$$y^1_{jk} = \{0,1\}\,, \tag{2.18}$$

with one-to-one correspondence with the Boolean variables (2.14). The proper disjunction (2.17e) can then be reformulated as the algebraic constraints

$$q_{jk} = \min\left(\tilde{q}^{\max}_j, \tilde{q}_{jk}\right), \qquad \forall j \in \mathcal{J}, k \in \mathcal{K} \tag{2.19a}$$
$$\tilde{q}_{jk} = \beta_j\left(m_{jk1} - m\left(e^S_j p_{\mathrm{w},j}\right)\right), \qquad \forall j \in \mathcal{J}, k \in \mathcal{K} \tag{2.19b}$$
$$\tilde{q}^{\max}_j = y^1_{jk} q^{\max}_j, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K} \tag{2.19c}$$
$$q_{jk} \geq y^1_{jk} q_{\mathrm{gc}}, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K} \tag{2.19d}$$

Note that $\tilde{q}^{\max}_j \in \mathbb{R}$ is an *auxiliary* variable which is substituted into the min-function (2.19a), while $q^{\max}_j$ is a constant. The variable $\tilde{q}_{jk}$ in (2.19b) is nonnegative, and in fact strictly positive due to the constraint (2.19d). Consequently, whenever $y^1_{jk} = 0$ for some (j,k) to relax the bound (2.19d), then $q_{jk} = \tilde{q}^{\max}_j = 0$, since $\tilde{q}_{jk} > 0$. Furthermore, by replacing the implication in (2.17f) and (2.17g) with a disjunction and manipulating the logic operators, we can convert these logic expressions to a conjunctive normal form (CNF), and eventually obtain the algebraic constraints (Raman and Grossmann, 1991)

$$y^1_{jk-1} - y^1_{jk} + y^1_{jk+1} \leq 1, \tag{2.20a}$$
$$y^1_{jk-1} - y^1_{jk} + y^1_{jk+2} \leq 1, \tag{2.20b}$$
$$\vdots$$
$$y^1_{jk-1} - y^1_{jk} + y^1_{jk+\tau_1-1} \leq 1, \tag{2.20c}$$

for the shut-in constraints (2.17f), and similarly

$$y^1_{jk-1} - y^1_{jk} + y^1_{jk+1} \geq 0, \tag{2.21a}$$
$$y^1_{jk-1} - y^1_{jk} + y^1_{jk+2} \geq 0, \tag{2.21b}$$
$$\vdots$$
$$y^1_{jk-1} - y^1_{jk} + y^1_{jk+\tau_2-1} \geq 0, \tag{2.21c}$$

for the start-up constraints (2.17g).

2.4.1 Reformulation of nonsmooth nonlinearities

Using general-purpose algorithms to solve MINLPs with nonsmooth functions as (2.17a) and (2.19a), may lead to unreliable solutions and oscillatory performance of the applied algorithms, and should in general be avoided. Generally, there are three types of approaches to this. The first is to apply special-purpose algorithms for nonsmooth optimization. These algorithms, however, often lack the efficiency of general NLP solvers for smooth problems. The second approach is to use some sort of smooth approximations for the nonsmooth functions. The third approach, if applicable, is to use an exact reformulation of the nonsmooth function, in most cases involving the introduction of nonconvex complementarity functions or additional combinatorial complexity by added binary variables. In the remaining of this section, we will consider the second and the third approach.

The objective function (2.17a) is the 1-norm, or the sum of the absolute values, of the total produce rate and the reference rate. By defining *nonnegative* variables z_k^+, z_k^- for all $k \in \mathcal{K}$, then (2.17a) can be reformulated to the smooth (differentiable) equivalent formulation

$$\min \sum_{k\in\mathcal{K}} z_k^+ + z_k^- \tag{2.22a}$$

s.t.

$$z_k^+ - z_k^- = q_k^{\mathrm{Ref}} - q_k^{\mathrm{Tot}}, \qquad \forall k \in \mathcal{K} \tag{2.22b}$$

Since (2.22a) is a minimization objective, then at least one variable for each pair $\left(z_k^+, z_k^-\right)$ will be driven to zero at the optimum. Hence, the reformulation (2.22a)-(2.22b) of the 1-norm minimization objective (2.17a) is exact *without* imposing the nonlinear complementarity condition $z_k^+ z_k^- = 0$.

The second nonsmooth function appearing in the formulation is the min-function in the well model (2.17e). This function is a disjunction itself. Any nonsmooth function $\min(g(x), f(x))$ can be reformulated as

$$\begin{aligned}\min\left(g(x), f(x)\right) &= -\max\left(-g(x), -f(x)\right) \\ &= g(x) - \max\left(0, g(x) - f(x)\right).\end{aligned} \tag{2.23}$$

We can therefore apply smooth approximations of the nonsmooth function $\max(0, h(x))$, with $h(x) := g(x) - f(x)$, to reformulate (2.17e). Note that the function $\max(0, x)$ is also referred to as the plus function. Balakrishna and Biegler (1992) proposed to rewrite $\max(0, h(x))$ as $|h(x)|/2 + h(x)/2$, and then apply a hyperbolic smoothing of the absolute value function. This leads to the approximation

$$\max(0, h(x)) \approx \frac{1}{2}\left(h(x) + \sqrt{h(x)^2 + \epsilon^2}\right), \tag{2.24}$$

where ϵ is a small positive constant typically in the range $10^{-2} - 10^{-4}$ to avoid singularity in the derivative of (2.24) when $h(x)$ is zero, i.e. at the point of the nondifferentiability. A different approximation was derived in Chen and Mangasarian (1996) by considering the sigmoid function approximation of the unit step function, together with probability density functions for smoothing the Dirac delta function.

Since $\max(0, x)$ is a ramp function, and the fact that the unit-step function and the ramp function are obtained by integrating the Dirac delta function once and twice, respectively, Chen and Mangasarian (1996) derived the smooth approximation

$$\max(0, h(x)) \approx h(x) + \frac{1}{\gamma} \ln\left(1 + e^{-\gamma h(x)}\right), \quad \gamma > 0. \tag{2.25}$$

This approximation converges to its exact nonsmooth counterpart as $\gamma \to \infty$. Other examples on smooth approximation of the function $\max(0, h(x))$ can be found in for instance Duran and Grossmann (1986). Note that none of these two smoothing procedures depend on bounds on the function arguments.

Requiring a minimum shut-in and production time of $\tau_1 = \tau_2 = 2$ days, and replacing the nonsmooth objective (2.17a) with the equivalent reformulation (2.22) together with one of the aforementioned smooth approximations, gives the following the smooth MINLP.

$$\min \sum_{k \in \mathcal{K}} z_k^+ + z_k^- \tag{2.26a}$$

s.t.

$$z_k^+ - z_k^- = q_k^{\text{Ref}} - q_k^{\text{Tot}}, \qquad \forall k \in \mathcal{K} \tag{2.26b}$$

$$q_k^{\text{Tot}} = \sum_{j \in \mathcal{J}} q_{jk}, \qquad \forall k \in \mathcal{K} \tag{2.26c}$$

$$A_j m_{jk+1} = m_{jk} + B_j q_{jk+1}, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K}^{\text{m}} \tag{2.26d}$$

$$m_{j0} = m_j^{\text{init}}, \qquad \forall j \in \mathcal{J} \tag{2.26e}$$

$$q_{jk} = f^{\text{S}}\left(\tilde{q}_j^{\max}, \tilde{q}_{jk}\right), \qquad \forall j \in \mathcal{J}, k \in \mathcal{K} \tag{2.26f}$$

$$\tilde{q}_j^{\max} = y_{jk}^1 q_j^{\max}, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K} \tag{2.26g}$$

$$\tilde{q}_{jk} = \beta_j \left(m_{jk1} - m\left(e_j^S p_{\text{w},j}\right)\right), \qquad \forall j \in \mathcal{J}, k \in \mathcal{K} \tag{2.26h}$$

$$q_{jk} \geq y_{jk}^1 q_{\text{gc}}, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K} \tag{2.26i}$$

$$y_{jk-1}^1 - y_{jk}^1 + y_{jk+1}^1 \leq 1, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K} \setminus K \tag{2.26j}$$

$$y_{jk-1}^1 - y_{jk}^1 + y_{jk+1}^1 \geq 0, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K} \setminus K \tag{2.26k}$$

The function f^{S} in (2.26f) is taken as one of the two following smooth approximations for $q_{jk} = \min\left(\tilde{q}_j^{\max}, \tilde{q}_{jk}\right)$: The MINLP - square root approximation with

$$f^{\text{S}}\left(\tilde{q}_j^{\max}, \tilde{q}_{jk}\right) = \frac{1}{2}\left(\tilde{q}_{jk} + \tilde{q}_{jk}^{\max} - \sqrt{\left(\tilde{q}_{jk} - \tilde{q}_{jk}^{\max}\right)^2 + \epsilon_1^2} + \epsilon_2\right), \tag{2.27}$$

where $\epsilon_2 \geq \epsilon_1$ is added to avoid negative flows when $\tilde{q}_{jk}^{\max} = 0$, shifting the error from the point of the nonsmoothness, or the MINLP - sigmoidal approximation with

$$f^{\text{S}}\left(\tilde{q}_j^{\max}, \tilde{q}_{jk}\right) = \tilde{q}_{jk}^{\max} - \frac{1}{\gamma} \ln\left(1 + e^{-\gamma\left(\tilde{q}_{jk} - \tilde{q}_{jk}^{\max}\right)}\right). \tag{2.28}$$

2.4.2 Exact linear reformulation

An exact linear reformulation of the disjunctive min-function defining the well model in (2.19a) can be derived by introducing additional binary variables. Let these binary variables be defined as

$$y_{jk}^2 \in \{0,1\}, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K} \tag{2.29}$$

with $y_{jk}^2 = 1$ if $q_j^{\max}$ is the smallest value and 0 if $\tilde{q}_k$ is the smallest. Furthermore, let

$$m^{\max} := \max_{(j,k)\in(\mathcal{J},\mathcal{K})} m_{jk1}, \tag{2.30}$$

be the least upper bound of the pseudopressures in grid block $i = 1$ for all time steps k in the prediction horizon $\mathcal{K}$. The reformulation (2.19) of the disjunction (2.17e) can then be extended with a logic reformulation of the min-function to obtain an exact reformulation of the nonconvex GDP (2.17) to an MILP formulation:

$$\min \sum_{k\in\mathcal{K}} z_k^+ + z_k^- \tag{2.31a}$$

s.t.

$$z_k^+ - z_k^- = q_k^{\mathrm{Ref}} - q_k^{\mathrm{Tot}}, \qquad \forall k \in \mathcal{K} \tag{2.31b}$$

$$q_k^{\mathrm{Tot}} = \sum_{j\in\mathcal{J}} q_{jk}, \qquad \forall k \in \mathcal{K} \tag{2.31c}$$

$$A_j m_{jk+1} = m_{jk} + B_j q_{jk+1}, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K}^{\mathrm{m}} \tag{2.31d}$$

$$m_{j0} = m_j^{\mathrm{init}}, \qquad \forall j \in \mathcal{J} \tag{2.31e}$$

$$q_{jk} \leq y_{jk}^1 q_j^{\max}, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K} \tag{2.31f}$$

$$q_{jk} \leq \tilde{q}_{jk}, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K} \tag{2.31g}$$

$$q_{jk} \geq y_{jk}^1 q_j^{\max} - q_j^{\max}\left(1 - y_{jk}^2\right), \qquad \forall j \in \mathcal{J}, k \in \mathcal{K} \tag{2.31h}$$

$$q_{jk} \geq \tilde{q}_{jk} - y_{jk}^2 \beta_j \left(m^{\max} - m\left(e_j^S p_{\mathrm{w},j}\right)\right), \qquad \forall j \in \mathcal{J}, k \in \mathcal{K} \tag{2.31i}$$

$$\tilde{q}_{jk} = \beta_j \left(m_{jk1} - m\left(e_j^S p_{\mathrm{w},j}\right)\right), \qquad \forall j \in \mathcal{J}, k \in \mathcal{K} \tag{2.31j}$$

$$q_{jk} \geq y_{jk}^1 q_{\mathrm{gc}}, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K} \tag{2.31k}$$

$$y_{jk}^1 + y_{jk}^2 \geq 1, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K} \tag{2.31l}$$

$$y_{jk-1}^1 - y_{jk}^1 + y_{jk+1}^1 \leq 1, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K} \setminus K \tag{2.31m}$$

$$y_{jk-1}^1 - y_{jk}^1 + y_{jk+1}^1 \geq 0, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K} \setminus K \tag{2.31n}$$

where the constraints (2.31f) - (2.31i) represent the equivalent mixed integer reformulation of the min-function in (2.19a). By inspection, it can be shown that the binary variables y_{jk}^1, y_{jk}^2 cannot both be zero at the same time. Adding the constraint (2.31l) hence avoids this infeasibility. As in (2.26), we have imposed a minimum two day alternating shut-in and production period, respectively.

The drawback of the mixed integer reformulation of the min-function is that the variable bounds on the function arguments must be known. The constant $m^{\max}$,

used to define the lower bound on $\tilde{q}_{jk}$, should be chosen as small as possible to provide a tight formulation. However, this constant will vary with the maximum shut-in time. Choosing $m^{\max}$ too small will therefore limit the possible maximum shut-in time, which may be disadvantageous for longer prediction horizons. Consequently, the value of this constant must be defined in some heuristic way. The approach we have applied is to run a simulation prior to the optimization with a long shut-in time, and then identify the largest value of m_{jk1} for all $j \in \mathcal{J}$ and $k \in \mathcal{K}$.

2.5 Computational results

In this section we consider the computational performance of the three different mixed integer formulations by using as example the single shale multi-well pad shown in Fig. 2.7. The single pad consists of 6 wells, where each well is assumed to have been operated different lengths of time, and hence have different initial pseudopressure m^{init}. Formally, we want to compare two properties of the formulations. The first property is how the MINLP formulations lend themselves to finding good local optimal solutions for different prediction horizons, as the problem is likely to have many local minima. That is, there may be several different sets of well switching times that give approximately the same deviation from the reference rate. The second compared property is the computational efficiency of the formulations, since the MILP formulation has additional combinatorial complexity, while the MINLP formulations are nonconvex with fewer binary variables.

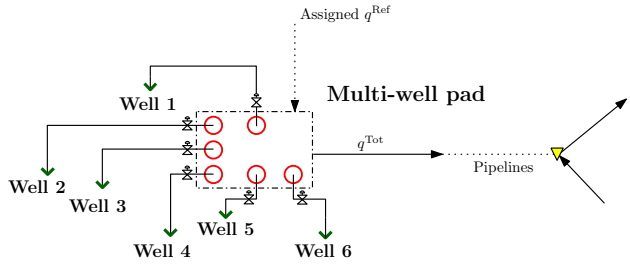


Fig. 2.7: System structure of numerical example.

We solve the nonconvex MINLPs by an NLP-based branch-and-bound method (Gupta and Ravindran, 1985), implemented in GAMS (Brooke et al., 2011) using the MINLP code SBB with the NLP solver CONOPT version 3.15. The MILPs are implemented and solved using the IBM ILOG CPLEX optimization studio (IBM ILOG CPLEX, 2011), version 12.3. Each problem is allocated a maximum CPU time of 2 hours, with the termination criteria set to a gap of less than 0.01%. The problems are solved using a Dell laptop with Intel Core i7 CPU and 8 GB of RAM. The starting points for the MINLP models are generated by running 100 *simulations* of the system with shut-in and start-up times generated by a uniform distribution. Each of these simulations require 1-5 seconds CPU time, depending on the given prediction horizon. We then use as starting point the simulation yielding

the lowest deviation in terms of the objective function (2.17a). To improve the performance of the NLP subsolver, we impose algorithmic bounds on the continuous variables in the MINLP formulation, inferred from the pre-simulation of the models.

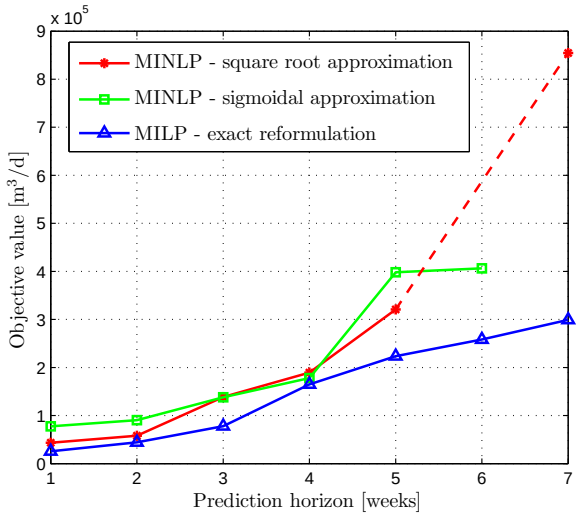


Fig. 2.8: Comparison of the three different formulations in terms of varying prediction horizon K. The dotted line indicates that no solution was found in this interval.

Fig. 2.8 compares solutions for the three different formulations as a function of the prediction horizon K. For a one and two week prediction horizon, there is a relatively small difference between solution of the MINLP square root formulation and the MILP formulation. However, the CPU time for the MILP model is significantly lower, cf. Table 2.3. For a three week horizon the difference between the two MINLP formulations and the MILP formulation is pronounced. However, for the four week case the solutions of the three different models are again close. Still, there is a relative difference of 15% and 8% in favor of the MILP vs. the MINLP square root and sigmoidal formulations, respectively. Beyond four weeks prediction horizon, the performance of the MINLP formulations is less robust and in a sense unpredictable, and there are certain problem sets where no solution is found within the two hour CPU timelimit. Comparably, the MILP model finds solutions for all prediction horizons K, and the objective value increases almost linearly with K. Hence, for this formulation, the per-day deviation in total produced rate compared to the reference rate is approximately constant.

The problem size and the CPU time is shown for one, two and seven weeks prediction horizon in Table 2.3. Beyond the two weeks prediction horizon all models are terminated by the maximum CPU time. For the one week prediction horizon, both solvers are able close the gap. It should be mentioned that CPLEX has a built-in (default) parallel computation mode. The real computation time will therefore be lower when the solver uses multiple threads. CPLEX is able to close the duality gap for the MILP model also for two weeks prediction horizon, while for longer

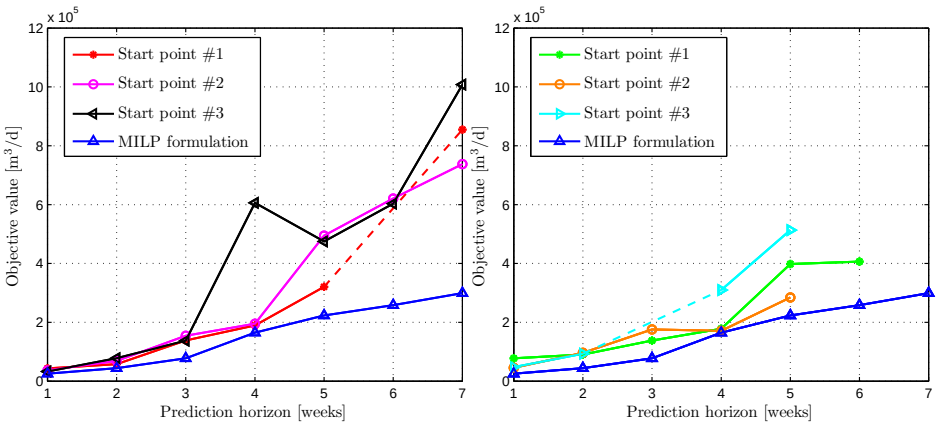
prediction horizons both CPLEX and SBB struggle with improving the lower bound and therefore terminate with large gaps. Note that the difference in problem size between the final MINLP and the MILP models is not very large. Since both of the smooth MINLP approximations include strong nonconvexities in terms of (2.27) and (2.28), respectively, we use intermediate variables and constraints to improve the implementation and the numerical robustness. The resulting increase in the problem size is handled by sparse matrix routines in the NLP solver.

Table 2.3: Test set statistics

	MILP - exact	**MINLP - square root approx.**	**MINLP - sigmoidal approx.**
	Prediction horizon $K = 1$ week		
Cont. var.	313	409	457
Binary var.	96	48	48
Constraints	628	509	557
CPU time [min]	0.3	6.3	0.5
	Prediction horizon $K = 2$ weeks		
Cont. var.	586	766	856
Binary var.	180	90	90
Constraints	1188	985	1075
CPU time [min]	86.2	120	120
	⋮	⋮	⋮
	Prediction horizon $K = 7$ weeks.		
Cont. var.	1951	2551	2851
Binary var.	600	300	300
Constraints	3988	3365	3665
CPU time [min]	120	120	120

Solving nonconvex NLP and MINLP models with local solvers usually relies on good initial starting points. To elaborate on the sensitivity of the two MINLP formulations to starting points, we compare in Fig. 2.9 the MINLP solutions, and the MILP solutions for different MINLP starting points. The starting points are generated from the pre-simulations using the three sets of shut-in and start-up times yielding the lowest objective function value. There is a slightly better match between the solutions for different starting points for the square root formulation than for the sigmoidal formulation, except for the four week prediction horizon. The sigmoidal MINLP formulation appears to be less robust than the square root formulation, particularly for longer prediction horizons where the sigmoidal formulation fails to find feasible solutions. One possible reason for this observation is that, for certain timesteps, the function arguments for the exponentials in (2.28) become relatively large. For longer prediction horizons, both of the MINLP formulations tend to find only a single integer-feasible solution, if any. In these cases, finding a feasible solution for a given starting point depends on good branching and node selection in the solver.

There are two sources of inaccuracy in the given mixed integer formulations. These are the error introduced by the smooth approximations of the min-function in (2.8), and the difference between the proxy model and multi-fracture reference model. To analyze the error from the applying the approximations (2.27) and (2.28), we compare the solutions of the two MINLP formulations for the four week prediction horizon with the equivalent exact model. The equivalent result with the exact



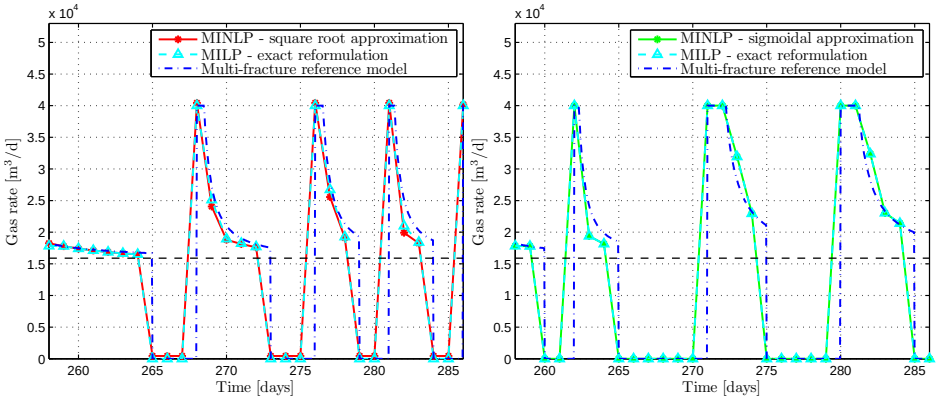
(a) MINLP - square root approximation. **(b)** MINLP - sigmoidal approximation.

Fig. 2.9: Sensitivity to starting points. A dotted line means that no solution is found in the interval between the two marked points.

model is obtained by a post-simulation using the exact reformulated model (i.e. the MILP model) with the the optimal switching times retrieved from the MINLP solutions. For the sigmoidal approximation, the optimal switching times simulated with the exact model yields a relative difference in the objective value of only 0.2%, while the equivalent comparison with the square root formulation gives a relative difference of 2.1%. The rate profile of well number 6 for the four weeks prediction horizon is shown in Fig. 2.10a for the MINLP square root approximation and in Fig. 2.10b for the MINLP sigmoidal approximation. Comparing the two figures, it can be seen that the discrepancy between the sigmoidal and the exact model is negligible, while there is a slightly larger discrepancy between the square root model and the exact model. Some of this difference is related to the choice of ϵ_1, ϵ_2 and γ in the approximations. The constants ϵ_1 and ϵ_2 in (2.27) was set to 0.01 after initial testing of the MINLP square root model to achieve acceptable numerical performance. Initial testing with the approximation (2.28) in the sigmoidal MINLP model revealed that a large value for γ was necessary in order to avoid accuracy problems related to the lower bound q_{gc}. Hence, this value was set to 75, giving good accuracy, but as observed in Fig. 2.9b, leading to more nonrobust numerical performance for longer prediction horizons. It can also be observed in Fig. 2.10a that the given square root approximation gives a small non-zero flow rate whenever the well is shut-in.

In the same Fig. 2.10, we compare both of the solutions with the optimal shut-in and start-up times simulated with the high-fidelity, multi-fracture reference model. The results from the use of the proxy models generally coincide well with the equivalent result simulated with the reference model. Still, the match is better for slightly longer shut-in times. This is as expected, cf. the tuning in Fig. 2.4 and the cross validation in section 2.3.1. Note that the observable discrepancy around the shut-in and the start-up times is caused by the use of a one hour discretization

interval for the multi-fracture reference model compared to a one day timestep for the proxy model.



(a) Well 6, MINLP - square root approximation.

(b) Well 6, MINLP - sigmoidal approximation.

Fig. 2.10: Comparison of the MINLP solution with the exact proxy model (MILP) and the high-fidelity, multi-fracture reference model.

The potential gain of applying optimized shut-in times for late-life shale multi-well pads is compared with a more naive, non-optimized operational approach in Fig. 2.11 for a seven week prediction horizon. In this naive, commonly applied operational approach (Dousi et al., 2006), we assume that each well is produced until signs of liquid loading is observed, and further assume that the signs in this case is a sharp drop in the gas rate, or the onset of erratic rates (Lea and Nickens, 2004). For simplicity, we assume that liquid loading occurs once the rate drops below q_{gc}, although some gas wells *may* produce stable sub-critical rates, including meta stable rates, for some time before signs of loading appears (Dousi et al., 2006; Li et al., 2002). The fictitious erratic rates from onset of loading are generated by simulation with uniformly distributed low-gain noise added to the rate once it drops below q_{gc}. Once the erratic rate occurs, the well is shut-in for a *fixed* number of days, in this case set to two days. Comparing the total produced rate using the optimized switching times against the naive operational approach, the latter leads to more than three times higher deviation from q^{Ref} than the optimized shut-in times, and the optimized approach avoids the high and the low peak rates as seen in Fig. 2.11. The high peak rates may be infeasible for the surface equipment due to limitations in the pipeline capacity or at the pipeline gathering points, cf. Fig. 2.1, or due to capacity constraints of a possibly connected compressor. High fluctuation in the rate as seen in the last part of the time horizon in Fig. 2.11 may cause substantial pressure variations in the surface pipelines, which may influence the line pressure for other pads and lead to oscillating flow in the pipelines. Hence, avoiding this by using optimized shut-in times may lead to more stable and better overall performance of the shale-gas surface gathering system.

The set of optimized start-up and shut-in times, however, quickly becomes

complex, as depicted in Fig. 2.12. There are 47 shut-ins for the optimized solution while the number is 28 for the naive approach. This complexity may be reduced by requiring a longer minimum shut-in and production time τ_1 and τ_2, respectively, with still better overall reference tracking than the naive approach. Comparing well number 2 and 3 in Fig. 2.12, which are the wells that have been producing for the longest and shortest time initially, it can be observed that well number 3 is producing for a significantly longer time in total than well number 2.

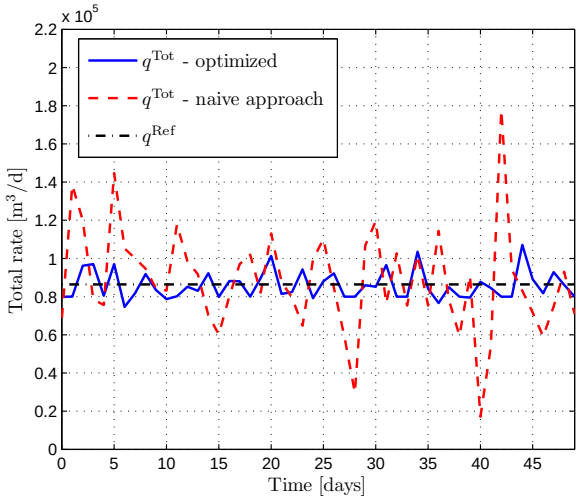


Fig. 2.11: The total rate for the pad in Fig. 2.7 with simulated liquid loading and a naive operational approach, compared with the total rate using optimized shut-in and start-up times.

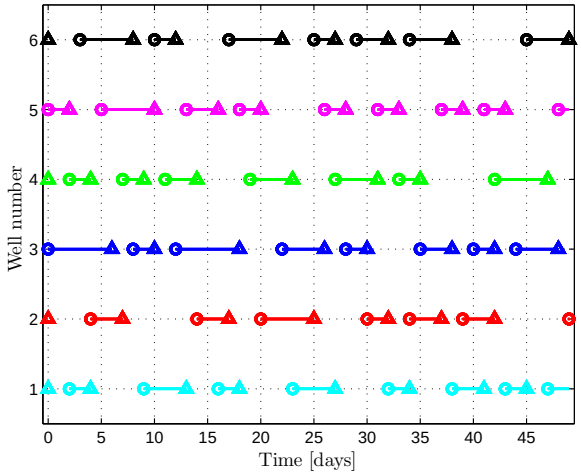


Fig. 2.12: Shut-in and start-up pattern for the optimized switching times giving the total rate shown in Fig. 2.11. The circle ◯ symbolizes a start-up of the well, while a △ symbolizes a shut-in.

2.6 Discussion

The two main contributions in this paper are the development of a novel multi-well production strategy for late-life shale-gas systems, and the derivation of mixed integer models for optimization of shut-in and start-up times in a single shale multi-well pad. Furthermore, the paper extends the tuning method and the efficiency of the shale single-well proxy model presented in Knudsen et al. (2012).

The proposed cyclic scheme may be used to improve the utilization of existing late-life resources. Applying this type of well operations may therefore serve as a complement to a continuous drilling of new wells to maintain the total gas supply rate from a field. By always producing above the critical rate q_{gc}, late-life erratic flow rates are avoided, while keeping the loss of cumulative production at a minimum by performing short shut-ins. If the shut-in scheme is applied to wells that are already loaded, then it may be necessary to apply longer shut-in times to get enough pressure build-up in the near wellbore region. Note that the commonly applied plunger lift technology is a form of cyclic shut-in control (Whitson et al., 2012). The proposed operational scheme may possibly be integrated with a plunger-lift system. Still, other issues like the wellbore curvature and safety issues with the rising plunger will then have to be taken into account. The cyclic, optimized production scheme may also be applicable to tight-gas systems, as these systems and wells share many of the same characteristics as shale-gas systems.

Scheduling shut-ins of late-life shale-gas wells according to a given reference rate rather than maximizing the production may be beneficial for several reasons. The first is a more stable total rate with reduced high and low peaks as demonstrated in Fig. 2.11. The second reason is related to lower sales prices as a result of increased availability of natural gas, partly due to increased shale-gas production. Gas prices may vary substantially with seasonal demands, and produced natural gas may be stored and sold at times with higher prices. As mentioned earlier, the loss in cumulative production by performing short shut-ins may be kept very low for shale-gas wells due to the matrix-fracture system acting as a refill and storage system. Consequently, by applying clever shut-in scheduling according to varying demands, the wells and reservoir are in a sense used for gas storage to account for seasonal variations in demands and gas prices.

The second contribution in this paper is the developments of mixed integer models from a GDP. For the given system structure in Fig. 2.7 with the selected initial well conditions, the MILP was shown to be superior for all prediction horizons in terms of less required CPU for short horizons and improved solutions for longer horizons. The exactly reformulated MILP model is the only formulation which can guarantee globality of the solution within the given duality gap. Further, it does not rely on function approximations. The MILP model is naturally more robust than the two MINLP formulations and does not rely on good starting points. The problem size of the MILPs, cf. Table 2.3, are generally small enough to be expected to outperform the noncovex MINLP. Still, there are drawbacks and challenges associated with the MILP formulation. As mentioned, the formulation depends on the bounds on m_{jk1} given in (2.30), which have to be decided heuristically, or set to a high number, giving a weaker formulation. Furthermore, the optimization problem is based on switching wells on and off to the meet a given reference rate. This

problem is computationally challenging, and it will almost certainly be impossible to meet the reference rate exactly.

By using a linear reservoir and well proxy model, one naturally omits certain nonlinear well and reservoir effects like friction and pressure-varying compressibility. Some of this is compensated through the tuning of the proxy model. Extending the proxy model with nonlinear effects may however be desirable or necessary in some cases to increase the accuracy of the model. In these situations, the MINLP formulations may be necessary. The comparison of the MILP and MINLP formulations may therefore serve as a *preliminary* indication of how well the formulations lend themselves to compute optimal shut-in and start-up times. Still, adding nonlinearities may change the performance of an MINLP formulation substantially, and the use of linearization techniques such as piecewise linearization should therefore be considered.

The use of a decomposition scheme may aid the optimization of both single pad systems as in Fig. 2.7, as well as multi-pad systems as in Fig. 2.1. The structure of these systems lend themselves to decomposition techniques since there are many local constraints linked to one or a few pads, and only a few global constraints. Hence, decomposition techniques which embed the methods studied here can be used to optimize larger shale-gas systems similar to Gunnerud and Foss (2010) for subsea oil production. Further, the proposed scheme may benefit from being integrated in a closed loop scheme where the proxy models are updated online to account for variations in well and reservoir properties.

2.7 Conclusions

This paper presents a complete MILP formulation for scheduling of shut-ins of shale-gas multi-well systems. The numerical examples proved the MILP formulation to be computationally superior to the corresponding MINLP formulations, while also maintaining better model accuracy. The novel PDE-based proxy model has been shown to be well suited for optimization of shut-in times by applying a thorough tuning of the model. Through the numerical examples, we have demonstrated how individual-well shut-ins can be efficiently scheduled and coordinated in a multi-well pad-system in order to reduce erratic production rates while meeting a shared production objective.

Chapter 3

Lagrangian Relaxation Based Decomposition for Well Scheduling in Shale-gas Systems

This chapter consists of the paper Knudsen et al. (2014a):

Knudsen, B. R., Grossmann, I. E., Foss, B., and Conn, A. R. (2014a). Lagrangian relaxation based decomposition for well scheduling in shale-gas systems. *Computers & Chemical Engineering*, 63:234–249.

Abstract

Suppressing the effects of liquid loading is a key issue for efficient utilization of mid and late-life wells in shale-gas systems. This state of the wells can be prevented by performing short shut-ins when the gas rate falls below the minimum rate needed to avoid liquid loading. In this paper, we present a Lagrangian relaxation scheme for shut-in scheduling of distributed shale multi-well systems. The scheme optimizes shut-in times and a reference rate for each multi-well pad, such that the total produced rate tracks a given short-term gas demand for the field. By using simple, frequency-tuned well proxy models, we obtain a compact mixed-integer formulation which by Lagrangian relaxation renders a decomposable structure. A set of computational tests demonstrates the merits of the proposed scheme. This study indicates that the method is capable of solving large field-wide scheduling problems by producing good solutions in reasonable computation times.

3.1 Introduction

The use of shale-gas as an energy resource has increased extensively over the last decade (EIA, 2013). Even though shale-gas recovery has primarily been a U.S. driven industry, there is currently an increasing exploration of shale and other tight formation resources in the Middle East, in North Africa, in China as well as in Europe (Dong et al., 2012). The long horizontal wells and stimulation with hydraulic fracturing needed to obtain profitable recovery rates from the very tight

rock formations, is however very costly. In recent years, low gas prices have made many shale-gas fields barely economical. This clearly increases the need for efficient planning, exploration and production techniques in shale-gas recovery.

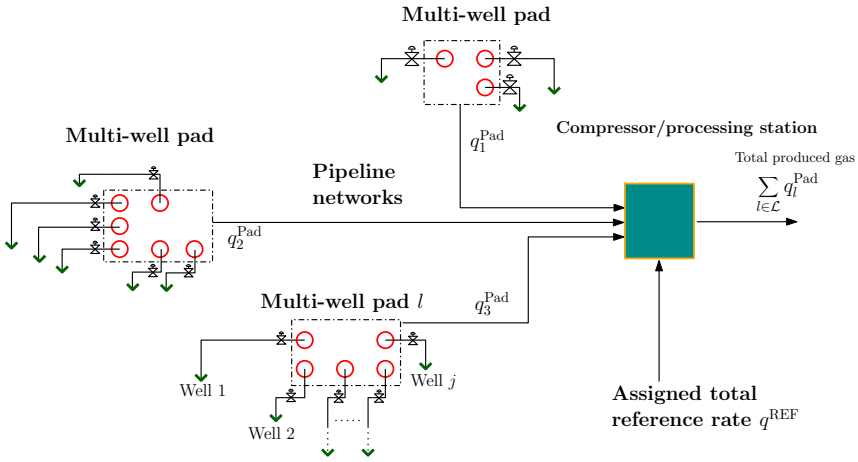


Fig. 3.1: Illustration of three pad shale-gas system.

Modern shale-gas developments are characterized by an increasing use of multi-well pads for sharing of surface infrastructure, both in the production from the wells and during the completion (Stefik and Paulson, 2011). Multi-well pads consist of several wells drilled at a single location, hence reducing the foot-print at the well locations by requiring less space for surface equipment, while at the same time encompassing a substantial part of the valuable shale formation. The pads are normally spread over a wide geographical area, with interconnecting pipeline networks transporting the gas to a compressor or processing station. This infrastructure is illustrated in Fig. 3.1. The gas-rate production profile from dry and semi-dry shale-gas wells are characterized by an initial high peak-rate or plateau rate if choking is performed, with an early steep decline followed by low pseudo steady-state rates (Jenkins and Boyer, 2008; Baihly et al., 2010). Shale-gas wells may produce at this pseudo steady-state rate for some hundred days or even for many years. All shale and tight formation gas wells will, however, eventually reach the state where the wells' pressure is insufficient to lift co-produced liquids to the surface, causing accumulation of liquids in the vertical wellbore (Whitson et al., 2012). This state is known as liquid loading, and detected by a very erratic and unstable rate, sharp drops in the decline curve (Al Ahmadi et al., 2010; Lea and Nickens, 2004) or semi-stabilization of the production rate at significantly lower levels (Dousi et al., 2006; Whitson et al., 2012), i.e. at so-called meta-stable rates. The accumulated liquids may be due to low saturations of water in the formation, condensates or oil, or left-over water from the hydraulic fracturing. Liquid loading severely deteriorates the performance of gas wells, and requires some remedial operational procedures. Whitson et al. (2012) showed that liquid loading can be eliminated by performing short regular shut-ins when the producing gas rate reaches the minimum rate required for continuous removal of liquids in the wellbore. This scheme is illus-

trated for a single well in Fig. 3.2. Whitson et al. (2012) also demonstrated that a cyclic scheme leads to minimal loss in cumulative production for low permeability wells when comparing with continuously-unloaded ("ideal" and non-realistic) production, while a significant improvement was obtained when comparing the cyclic shut-in scheme with production at meta-stable rates.

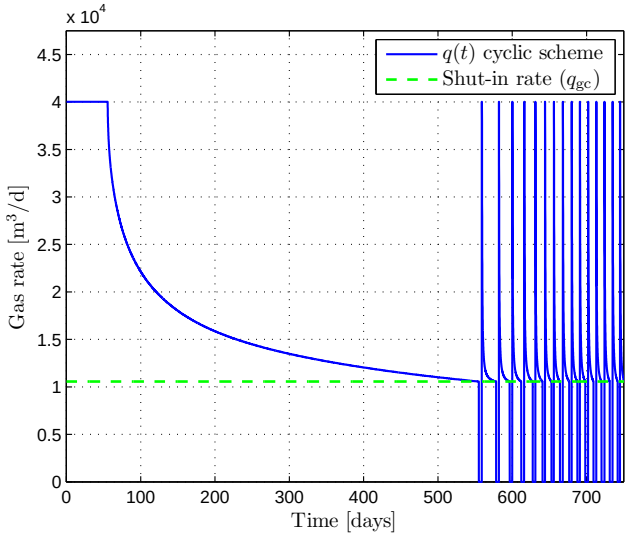


Fig. 3.2: Illustration of a cyclic shut-in scheme, showing the simulated gas rate for a multi-fractured horizontal shale-gas well using a *fixed* shut-in time.

The cyclic shut-in scheme introduced by Whitson et al. (2012) was extended to multi-well systems and time-varying shut-ins in Knudsen and Foss (2013). The problem of optimizing shut-in times for shale multi-well pads was formulated as a mixed integer linear program (MILP), using simplified well and reservoir proxy models (Knudsen et al., 2012). For field-wide multi-pad systems, however, the dimensionality of the MILP becomes computationally prohibitive. The distributed shale-gas multi-pad system shown in Fig. 3.1 renders a network-type structure with limited coupling between the pads. For these types of systems, the use of a Lagrangian relaxation method provides an efficient solution approach by decomposing the original problem into smaller subproblems, see e.g. (Foss et al., 2009; Sagastizábal, 2012). Nevertheless, any efficient implementation of a Lagrangian relaxation requires a stable and reliable method for finding optimal Lagrangian multipliers, as well as a heuristic for finding primal feasible solutions based on the solution from the Lagrangian and the dual. Furthermore, for the type of mixed-integer dynamic system considered in this paper, an efficient optimization scheme relies on using sufficiently accurate and numerically efficient proxy models, together with a good problem formulation such that the subproblems can be solved in reasonable computation time. The goal of this paper is to address these issues in the context of shut-in based production planning for the shale-gas multi-pad system shown in Fig. 3.1.

The remainder of this paper is organized as follows. Section 3.2 presents a review

of the shale well and reservoir proxy model. In Section 3.3, we present a parameter estimation problem for the tuning of the proxy model, together with a simple scheme for updating the model parameters. Section 3.4 presents the formulation of the fullspace optimization problem, while the Lagrangian relaxation scheme is presented in Section 3.5. Subsequently, results from computational testing of the scheme will be presented in Section 3.6. A discussion on the applicability and the numerical efficiency of the proposed scheme is given in Section 3.7, with concluding remarks in Section 3.8.

3.2 Shale well and reservoir proxy modeling

Hydraulically fractured shale-gas reservoirs are conventionally modeled either as variations of the dual-porosity idealization of naturally fractured reservoirs (Warren and Root, 1963), as discrete fracture models (Karimi-Farad et al., 2004), or as fully discretized single-porosity dual-permeability models (Cipolla et al., 2010). These modeling techniques have different merits and drawbacks in terms of accuracy, detail level and numerical efficiency, while they are all constructed with the incentive of both capturing highly dynamic short-term effects as well as predicting long-term recovery rates. In contrast to these detailed, numerically demanding modeling schemes, is the widespread use of simple semi-analytical or empirical models for gas-rate forecasting (Sutton et al., 2010; Al Ahmadi et al., 2010; Bello and Wattenbarger, 2010). These models, however, assume continuous production at a constant bottomhole pressure, and are therefore not suited for optimization of a cyclic production scheme.

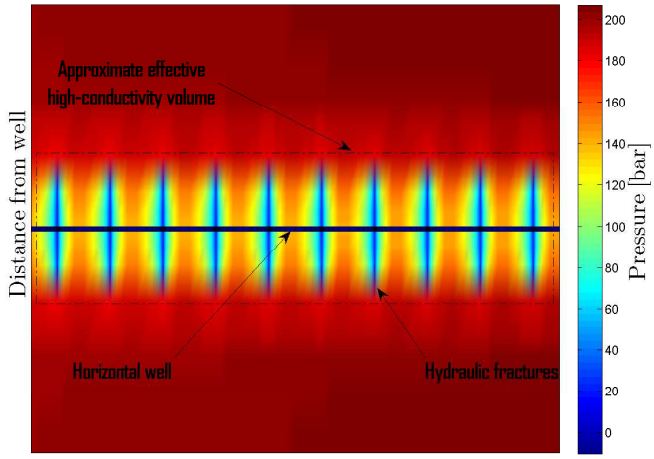


Fig. 3.3: The topside view of the grid pressure after 450 days of continuous production from simulation of a multi-fracture shale-gas reservoir model.

Both of the aforementioned groups of shale-gas modeling schemes are efficient

for their purposes, but lack applicability with respect to the goal of efficiently optimizing shut-ins for large multi-well systems. A simplified well and reservoir proxy model for dry and semi-dry shale-gas wells was therefore derived in Knudsen et al. (2012), with modifications and extensions in Knudsen and Foss (2013). Consider Fig. 3.3 showing the grid pressure after 450 days of continuous production in a simulated multi-fracture shale-gas reservoir model. The pressure depletion around and between the fractures will over time "homogenize" the stimulated reservoir volume encompassing the fracture system, leading to a determination of average properties and pressures in this region of the reservoir. During shut-ins, gas will flow from the outer low permeability region with high pressure (red in Fig. 3.3) and recharge the effective low-pressure fracture system volume with gas. The reservoir proxy model is based on approximating the effective low-pressure high-conductivity volume marked out in Fig. 3.3 as *one*, compact volume. By using a radial composite model as shown in Fig. 3.4, we construct the proxy as a three-region model with an intermediate crushed-rock type volume and a low permeable outer region representing the tight shale rock. The effective high-conductivity volume in Fig. 3.3 will vary and clearly increase at late times of the well.

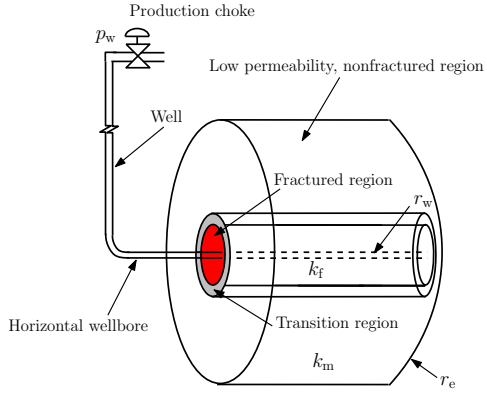


Fig. 3.4: Illustration of proxy model.

Assuming Darcy's law and single phase gas together with the law of mass conservation and the equation of state for real gases, the nonlinear partial differential equation (PDE) for the gas flow in the radial composite proxy model in Fig. 3.4 is given by

$$\phi \frac{p}{Z(p)} c(p) \frac{\partial p}{\partial t} = \frac{1}{r} \frac{\partial}{\partial r} \left(k(r) \frac{p}{\mu(p) Z(p)} r \frac{\partial p}{\partial r} \right), \tag{3.1}$$

where p is the pressure, $k(r)$ is a radially dependent permeability, $\mu(p)$ is the gas viscosity, $Z(p)$ is the gas compressibility factor and $c(p)$ is the total compressibility. The pressure dependency of $\mu(p)$ and $Z(p)$ can be efficiently handled by the variable transformation known as *pseudopressure* (Al-Hussainy et al., 1966), given by

$$m(p) = 2 \int_0^p \frac{p'}{\mu(p') Z(p')} dp'. \tag{3.2}$$

By using Neumann boundary conditions with a producing well (sink) at the center, r_{w}, no-flow conditions at the outer boundary, r_{e}, and a given initial reservoir pressure, we obtain the following initial boundary-value problem (IBVP) in terms of the reservoir pseudopressure, $m(p)$ (Knudsen et al., 2012).

$$\phi\mu c\frac{\partial m}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\left(k(r)r\frac{\partial m}{\partial r}\right), \tag{3.3a}$$

$$r\frac{\partial m}{\partial r}\bigg|_{r_{\mathrm{w}}} = q\frac{Tp_{\mathrm{sc}}}{T_{\mathrm{sc}}\pi hk}, \tag{3.3b}$$

$$\frac{\partial m}{\partial r}\bigg|_{r_{\mathrm{e}}} = 0, \tag{3.3c}$$

$$m(r,0) = m^{\mathrm{init}}, \tag{3.3d}$$

where q is the gas rate, T is the reservoir temperature and h is the horizontal length of the well. The subscript sc refers to evaluation at standard condition, 1 bar and 15.6°C. All gas volumes and volumetric rates are given in standard conditions in the rest of the paper. The product μc in (3.3a) is evaluated at given pressures, hence rendering a semi-linear PDE. This technique is elaborated in Section 3.3 below. The initial pseudopressure (3.3d) is calculated by trapezoidal integration of the corresponding values of $p, \mu(p)$ and $Z(p)$ which are obtained by gas correlations, see Whitson and Brule (2000).

The IBVP (3.3) is discretized in space and time in Knudsen et al. (2012), using a finite difference scheme. Spatial discretization applies central difference approximations, i.e. with a second order accuracy. The time discretization uses the backward Euler scheme, or equivalently, a first order orthogonal collocation on finite elements with fixed element lengths. In the completion and design of shale-gas wells, a maximum rate, $q^{\max}$, is normally specified, based on the surface equipment specifications together with long-term strategic planning of the wells. Moreover, a minimum wellhead pressure is required with respect to the given line pressure. The well rate will initially, or after a shut-in, deliver a peak or plateau rate for some time until the wellhead pressure is equal to the line pressure and the rate starts to decline. Combing these expected features, a simple aggregated well and wellbore model was derived in Knudsen et al. (2012). Using this aggregated model together with the discretized reservoir proxy model leads to the following simplified single well and reservoir proxy model.

$$Am_{k+1} = m_k + Bq_{k+1}, \tag{3.4a}$$

$$m_0 = m^{\mathrm{init}}, \tag{3.4b}$$

$$q_k = \min\left\{q^{\max}, \beta\left(m_{k1} - m(e^S p_{\mathrm{w}})\right)\right\}, \tag{3.4c}$$

where m_k is a vector containing the grid pseudopressures, A is a tridiagonal matrix, B is a single-column matrix, β is a constant, possibly skin-dependent, well index, e^S is the gravity term and p_{w} is the wellhead pressure. Note that m_{k1} is consequently the pseudopressure in the innermost grid block.

3.2.1 Critical gas rate

A critical gas rate, q_{gc}, sometimes referred to as the minimum rate to lift, can be specified as a lower bound on the flowing gas rate, q_k, in order to ensure continuous removal of liquids in the wellbore. The most widely applied model for calculating q_{gc} is given by Turner et al. (1969), with the onset of liquid loading observed for the majority of gas wells to be controlled by the wellhead conditions. Coleman et al. (1991) modified Turner's criterion for onset of liquid loading for gas wells with lower wellhead pressures. As shale-gas wells normally operate at low wellhead pressures, we use their model for calculating q_{gc}. By ensuring that the producing gas rate is always greater or equal to q_{gc}, we avoid accumulation of liquids in the vertical part of the wellbore. Furthermore, we limit the valid region of our model and avoid the need to use a multiphase model for low liquid systems such as dry and semi-dry shale-gas wells.

3.3 Computing model parameters

The gas-rate profile from shale and fractured tight formation wells is often observed to be defined by three distinct sets of transients. The first is a steep decline following the early-time peak or plateau rate, during which the gas initially stored in the fracture network is drained. The second is a very long transient with low rates, often characterized as a pseudo steady-state condition. The last category is pressure build-up transients observable during re-openings of wells after shut-ins. Modeling all of these transients accurately is recognized as inherently difficult (Jayakumar et al., 2011), even for very complex and detailed numerical models. Consequently, there are fundamental trade-offs when constructing shale-gas models, and clearly a proxy modeling scheme must take into account which type of transients the proxy should replicate.

The basis for the scheme we propose is to perform short, regular shut-ins to prevent liquid loading. Clearly we are most interested in constructing a model replicating the last of the aforementioned categories of transients, i.e. the transients from re-opening a well after a shut-in. During shut-ins, the gas stored in the low permeable shale matrix will flow into the highly conductive fractures and refill these small volumes with gas. The transients in the gas rate when the well is subsequently started up will exhibit slightly different characteristics depending on the length of the shut-ins and the pressures at the time of shut-in. The duration of these transients depends on the pressure build-up in the near-wellbore region, which in turn depends on several factors. The first is the essential reservoir characteristics such as the formation permeability, the conductivity of the fractures and the effective volume of the hydraulically induced and natural fractures. Some of these properties, in particular the formation permeability and the average fracture half-length, can be estimated using the techniques described in Al Ahmadi et al. (2010) and Bello and Wattenbarger (2010). Other effects that impact the magnitude of the pressure build-up during shut-ins include the total compressibility in (3.3a). Assuming negligible rock compressibility and using the equation of state for real gases, the total compressibility for gas reservoirs under isothermal conditions

can be expressed as follows (Whitson and Brule, 2000),

$$c(p) = \frac{1}{\rho}\frac{\partial \rho}{\partial p}\bigg|_T = \frac{1}{p} - \frac{1}{Z(p)}\frac{\partial Z(p)}{\partial p}\bigg|_T . \tag{3.5}$$

The compressibility of the gas decreases the initial pressure build-up during shut-ins, as some of the pressure change is used to decrease the volume of the gas in the pores of the reservoir. These pressure effects are strongly nonlinear, and are recognized as difficult to include in semi-analytical and proxy models, particularly in conjunction with the use of the pseudopressure, $m(p)$. Commonly applied techniques include evaluating $c(p)$ at the initial pressure or at the average of the initial reservoir pressure and the bottomhole pressure. Both former techniques are equivalent with assuming a slightly compressible fluid. However, instead of choosing at which pressures to evaluate the compressibility, we parametrize the proxy model (3.3a) with distinct compressibilities c_{f} and c_{tran} in the fracture region and the transition region, respectively, and apply a tuning of the model to achieve a good compromise between the compressibility effects during pressure depletion and build-up. As the pressure in the outer low permeability region remains almost constant, we evaluate $c(p)$ at initial conditions in this region. We further estimate a fracture radius r_{f} and an approximated well length $\tilde{h}$ for describing the varying effective high-conductivity volume in Fig. 3.3, and an average permeability k_{tran} in the intermediate transition region. Together, this defines the vector

$$\theta = \begin{bmatrix} r_{\mathrm{f}} & k_{\mathrm{tran}} & c_{\mathrm{f}} & c_{\mathrm{tran}} & \tilde{h} \end{bmatrix}^T \tag{3.6}$$

of unknown parameters in the proxy model (3.4).

To estimate parameters in the proxy model, we formulate a constrained weighted nonlinear least squares (WNLS) problem. Using this problem for model tuning and system identification as opposed to ordinary LS estimation, provides a means for emphasizing certain dynamics of the system to be modeled. The weighting of the residuals in the WNLS problem can be obtained by prefiltering of the data (Ljung, 1999). A linear prefilter acts as a frequency weighting of the objective function, and can be designed to increase the model accuracy in the frequency range of interest. This interpretation of the prefilter, however, only applies to strictly linear system identification (Spinelli et al., 2005). Moreover, the equivalence of input-output data filtering and prediction error filtering in linear LS system identification (Ljung, 1999), does in general not hold for nonlinear systems (Spinelli et al., 2005). By using the Volterra series representation of nonlinear systems together with the generalized frequency response function, Spinelli et al. (2005) demonstrated that reduced estimation bias in the desired bandwidth can be achieved by filtering of the prediction errors compared with separate filtering of the input/output data set. Consequently, even though the frequency weighting interpretation only strictly holds for linear parameter estimation, we can obtain a similar effect in the nonlinear case by filtering the prediction errors through a properly designed filter. Let

the prefilter $L_{\mathrm{p}}(\gamma)$ be given by a general infinite impulse-response filter (IIR),

$$L_{\mathrm{p}}(\gamma) = \frac{C(\gamma)}{D(\gamma)} = \frac{\sum_{m_1=0}^{M_1} c_{m_1}\gamma^{-m_1}}{1 + \sum_{m_2=1}^{M_2} d_{m_2}\gamma^{-m_2}}, \tag{3.7}$$

where γ^{-1} is the unit delay operator[1]. The filter (3.7) can be included as a recursive digital filter in the WNLS problem. Note that the filter equations simply amount to a set of linear constraints enforced at each timestep, and hence do not increase the complexity of the resulting nonlinear program (NLP). We use a second-order digital low-pass Butterworth filter with a bandwidth $[0, 0.4]$, normalized with respect to the Nyquist frequency. Let

$$\epsilon_k := q_k - q_k^{\mathrm{MFR}} \tag{3.8}$$

be the prediction error, where q^{MFR} is the gas rate from the multi-fracture reference model (MFR). Further, let $\bar{y} = [\bar{y}_1\ \bar{y}_2 ...]^T$ be a predefined vector of *binary* valve-settings defining a shut-in schedule. Using an estimation horizon $\hat{\mathcal{K}}$, we then estimate the proxy model parameters, $\theta \in \mathbb{R}^5$, by solving the NLP,

$$\min_{\theta} \sum_{k\in\hat{\mathcal{K}}} \left(\epsilon_k^{\mathrm{f}}\right)^2 \tag{3.9a}$$

s.t.

$$\epsilon_k^{\mathrm{f}} = L_{\mathrm{p}}(\gamma)\epsilon_k, \qquad \forall k \in \{\mathcal{K} : k \geq \max(M_1, M_2)\} \tag{3.9b}$$

$$\epsilon_k = q_k - q_k^{\mathrm{MFR}}, \qquad \forall k \in \mathcal{K} \tag{3.9c}$$

$$A(\theta)\, m_{k+1} = m_k + B(\theta) q_{k+1}, \qquad \forall k \in \mathcal{K} \setminus K \tag{3.9d}$$

$$m_0 = m^{\mathrm{init}}, \tag{3.9e}$$

$$q_k = \frac{1}{2}\left(\tilde{q}_k + q^{\max} - \sqrt{\left(\tilde{q}_k - q^{\max}\right)^2 + \delta_1^2} + \delta_2\right), \qquad \forall k \in \mathcal{K} \tag{3.9f}$$

$$\tilde{q}_k = \bar{y}_k\, \beta(\theta)\left(m_{k1} - m\left(e^S p_{\mathrm{w}}\right)\right), \qquad \forall k \in \mathcal{K} \tag{3.9g}$$

$$\theta^{\mathrm{lo}} \leq \theta \leq \theta^{\mathrm{up}}, \tag{3.9h}$$

$$c_{\mathrm{tran}} \leq c_{\mathrm{f}}, \tag{3.9i}$$

where (3.9f) is a smooth approximation of the nonsmooth well model (3.4c), with smoothing parameters δ_1 and δ_2 selected from the interval $\left[10^{-2}, 10^{-4}\right]$. The second smoothing parameter $\delta_2 > \delta_1$ is added to prevent small negative rates when the well is shut-in. Observe that whenever $\bar{y}_k = 0$ for some $k \in \mathcal{K}$, then $\tilde{q}_k = 0$ and hence $q_k \approx 0$ since (3.9f) gives the minimum of $\tilde{q}$ and $q^{\max}$. The constraint (3.9i) is based on the property that average compressibility in the reservoir is higher when closer to the wellbore, and serves as a regularization of the parameter estimation problem in addition to the bounds (3.9h) (Thompson and Kramer, 1994; Johansen, 1996). Note that the compact parametric representation (3.9d) of the discretized

[1]The unit delay operator is usually defined as q^{-1} (Ljung, 1999). However, as q is used as notation for the gas rate, we use for notational convenience γ^{-1}as the equivalent operator.

PDE model consists of a set of nonlinear equations used to define the spatial gridding and the transmissibilities in the reservoir model. For further details on the derivation of the proxy model (3.9d), see Knudsen et al. (2012).

For the tuning of the proxy model, we use a high-fidelity multi-fracture reference model constructed with cylindrical geometry similar to Fig. 3.3, with ten equally spaced fractures. The high-fidelity model is implemented in the state-of-the-art simulation software SENSOR (SENSOR, 2011), using logarithmic grid refinements between and outside the fractures, with totally 4200 grid cells. We further extend the high-fidelity reservoir model with a static wellbore model, see e.g. Katz and Lee (1990). Essential characteristic properties of the proxy and the reference models are quoted in Table 3.1. The spatial grid in the proxy model is constructed using $I = 4$ grid blocks, see Knudsen and Foss (2013) for a discussion on the choice of this coarse gridding. We use a $\Delta k = 1$ day fixed timestep for the proxy model, while we use a one hour timestep for the multi-fracture reference model to capture more of the high-frequency dynamics in this model. Note that in practice the gas-rate measurements are likely to be available on a daily basis only. If available gas-rate production data contains several short shut-ins, these data can, after sufficient preprocessing, be directly applied in the tuning of the proxy model. Otherwise, a similar high-fidelity model to the one used in this paper should be constructed, matched and validated against the production data, and then simulated with shut-ins to produce sufficient excitation for the tuning of the proxy model.

The NLP (3.9) is implemented in GAMS (Brooke et al., 2011) and solved with the NLP solver CONOPT v3.15. The results of the tuning of the proxy model is shown in Fig. 3.5, where we have used two shut-ins of five and three days, respectively, to excite the model. The proposed modeling and tuning scheme renders a good match of transients after shut-ins of a couple of days. The match is observed to be tighter in the pseudo-steady state region following the second transient compared with the transient after the first (five-day) shut-in, while a discrepancy is seen in the peak-rate after the second, three-day long shut-in. This discrepancy reflects the trade-offs made in the proxy modeling (3.4) and in the design of the prefilter (3.7): increasing the bandwidth or shifting the bandpass would increase the gain on the fast modes of the system and hence, give a better match of the second peak in Fig. 3.5. A nonzero steady-state gain is necessary, however, to obtain an acceptable match of the model at the time at which the rate crosses q_{gc}, as the rate at 585 days is almost steady-state. Arguably, one could consider using a bandpass filter instead of a low-pass filter to place more emphasis on the transient modes, and thereby better match the peak rates. This will, however, give a higher bias in the rate of the proxy model in the pseudo steady-state regions, and lead to a constant or almost non-decreasing rate after the initial decline of the rate. Some of the mismatch in the curvature of the transients is also due to the constant compressibility assumption, while parts of the mismatch may be compensated for by a higher-order discretization scheme. Note that the parameter estimate will be somewhat sensitive to the NLP starting point since the WNLS (3.9) is a nonconvex NLP. To compensate for this dependency, we have used a multi start-point strategy with different initial choices θ^{init} to provide good starting points for the NLP (3.9). Several of these starting points gave almost identical model match, but slightly different values of θ. For the prediction below, we choose θ^* giving the least

sum-of-squares (3.9a).

The advantage of using a Butterworth filter is the flat magnitude in the bandpass. However, as Butterworth filters are all-pole filters, there will be a significant phase shift at the output of the filter, particularly for high-order filters. Care must therefore be taken with respect to the choice of filter order if the length of the time-series of production rates is short and there are shut-ins at the end of time-series, as parts of the transient may be lost due to the phase lag of the filter. Compared with linear system identification, in which a prefilter tends to cut off model content in frequencies outside the bandpass, the prefilter applied in nonlinear system identification introduces a scattering of the frequency spectrum outside the bandpass. This property may impact the frequency weighting in (3.9) and hence the model match. See (Spinelli et al., 2005) for further details and analysis.

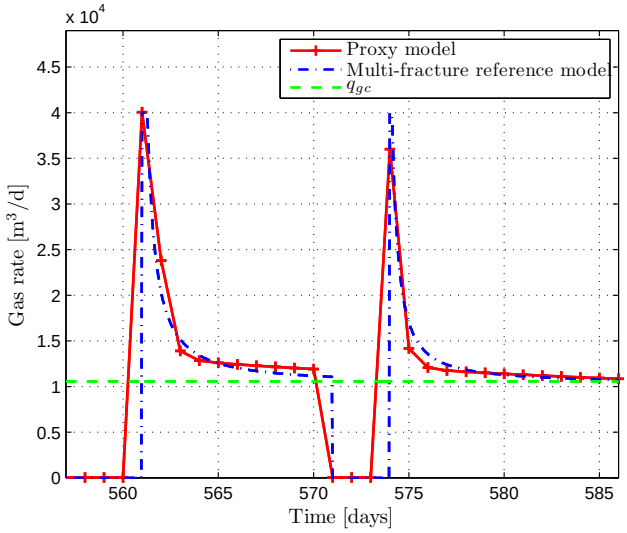


Fig. 3.5: Tuning of the parameters, θ, for the proxy model. The first transient corresponds to a five-day shut-in, while the second shut-in is three days long.

Table 3.1: Given reservoir model parameters. The two parameters below the dashed line only apply to the multi-fracture reference model.

Symbol	Parameter	Value	Unit
ϕ	porosity	6	%
k_{m}	formation permeability	3×10^{-4}	mD
z_{w}	true vertical depth	2300	m
p_{r}	initial reservoir pressure	200	bar
p_{w}	minimum wellhead pressure	6.9	bar
$q^{\max}$	maximum rate	4×10^4	m^3/d
h	true length of horizontal wellbore	492	m
n_{f}	true number of hydraulic fractures	10	-

3.3.1 Updating model parameters

One of the key issues for industrial acceptance of model-based decision support tools is the user's ability to maintain and update the models used (Foss and Jensen, 2011). This is particularly challenging for asset and field-wide optimization similar to the problem in this paper, which includes multiple disaggregated model types such as reservoir, surface and wells models, all having different time scales. Challenges are related to the complexity of the models used in the optimization (Bieker et al., 2007), the level of automation of the model update, and the (possibly excessive) amount of data available. Quite commonly, models are updated relatively infrequently, e.g. on a yearly basis (Foss and Jensen, 2011). Such infrequent model updates introduce a significant time-delay in the feedback decision loop, and clearly deteriorate the accuracy and the performance of the decision support tool (Foss and Jensen, 2011). In this context, there is a clear advantage using model predictive control (MPC) compared with real-time optimization with separate model updating, as model errors and slowly drifting parameters are compensated for by the per-timestep feedback and the estimation of initial conditions. However, the use of MPC in a large, geographically spread petroleum production system introduces several challenges, and requires a high level of automatic control and less human-in-the-loop operations.

Although a cross-validation in Knudsen and Foss (2013) of a similar realization of the proxy model (3.4) showed a good match with the reference model, the accuracy of the proxy model will abate over time. This is a results of the tuning technique (3.9d) in which only a limited window of data and shut-ins is used, as well as a result of using the prefilter. Furthermore, on a long time horizon, the actual flow and pressure characteristics of the reservoir change as a result of local pressure depletion in the fracture network, transition to boundary dominated flow, and changes in the proppant distribution and the near-wellbore skin (Bello and Wattenbarger, 2010). Shale-gas reservoirs may also contain time-varying parameters such as pressure dependent formation permeability (Cipolla et al., 2010). The parameter vector θ should hence be updated using a moving horizon like estimation technique. Using a window length $\hat{\mathcal{K}}_{\text{up}}$ of only the last set of applied shut-ins, we limit the influence of the earlier observed shut-ins and the tuning of the proxy model. For a given optimal θ^*, obtained initially by solving the parameter estimation problem (3.9) shown in Fig. 3.5, we update the proxy model by solving an augmented form of the WNLS problem (3.9),

$$\min_{\theta} \sum_{k \in \hat{\mathcal{K}}_{\text{up}}} \left(\epsilon_k^{\text{f}}\right)^2 + (\theta - \theta^*)^T D_{\text{s}}^T P D_{\text{s}} (\theta - \theta^*) \tag{3.10a}$$

s.t.

$$\text{equation (3.9b) - (3.9i)}, \tag{3.10b}$$

where P is a symmetric positive definite matrix, and D_{s} is a diagonal scaling matrix used to normalize the different parameters in θ. The updating is qualitatively different from ordinary moving horizon based parameter (and state) estimation (MHE) (Rao et al., 2001). In MHE, the quadratic term in (3.10a) corresponds to an arrival cost of new data, where an optimal P can be computed by an Extended

Kalman filter or other techniques. Consequently, P is in this case used to aggregate previously obtained information on the model. The shale-well proxy model, however, is only valid in a limited region (i.e. over a certain time and pressure range). Rather than aggregating information, we want the estimates to be gradually forgotten as new shut-in data becomes available. Consequently, for each update t of the parameter vector θ, we update the matrix P using a type of forgetting factor (Ljung, 1999),

$$P^t = \alpha I P^{t-1}, \tag{3.11}$$

where $\alpha \in (0, 1)$ is a constant, and we use the last optimal parameter estimate

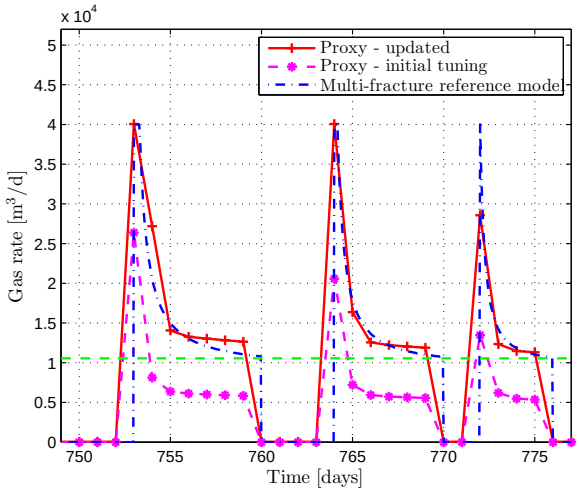


Fig. 3.6: Updating parameters in the proxy model.

as the starting point when solving the NLP (3.10). Fig. 3.6 compares the rate of the proxy model using the initial optimal parameter vector θ^* with the updated proxy model from solving (3.10). The initially tuned proxy preserves the essential curvature in the rate transients, but displays a prominent bias. This bias relates to the total pore volume of the proxy model compared with the actual reservoir (reference model). The volume of the fractured high-conductivity region (cf. Fig. 3.3) is tuned to match short-term pressure transients during shut-ins, and is too small to capture the long-term gas drainage, hence causing a fast pressure depletion of this volume. The outermost grid block, which in a way acts as a source term, is unable to compensate for this pressure depletion due to the very low permeability. The mismatch in the peak rate seen at day 772 in Fig. 3.6 also shows the difficulty encountered when shorter shut-ins are used to update the proxy. For such short shut-ins the dynamics are even faster, and the proxy is unable to match this peak rate without sacrificing accuracy in the match of the other transients.

3.4 Formulation of the optimization problem

This section describes the formulation of the optimization problem for the shale-gas multi-pad structure depicted in Fig. 3.1. All setdefinitions and indices are given in

Table 3.2: Indices and set definitions

Index	Interpretation	Set	Elements
i	spatial reservoir grid block	$\mathcal{I}$	$\{1 \dots I\}$
j	well number	$\mathcal{J}$	$\{1 \dots J\}$
k	discrete time index	$\mathcal{K}$	$\{1 \dots K\}$
l	pad number	$\mathcal{L}$	$\{1 \dots L\}$
n	iteration in Lagrangian scheme	-	$\{1 \dots N\}$
u	cuts in the cutting-plane algorithm	-	$\{1 \dots U\}$
d	terms in the disjunctions	-	$\{1, 2, 11, 12, 13\}$

Table 3.2. We assume that a production plan for K days in terms of a reference rate q^{REF} is given from strategic and contract based management of the field. In the derivation of the Lagrangian relaxation scheme, we will assume that the reference rate is constant over the planning horizon $\mathcal{K}$, while we show at the end of the paper how the scheme can adopt varying reference rates. When shutting in multiple wells at a single location without some scheduling, then in the worst case the entire set of wells at one pad may be shut in at the same time giving a zero total rate, while the peak rate when the wells are subsequently re-opened may be infeasible with respect to the capacity of the surface system. This must clearly be avoided. We want each pad, $l \in \mathcal{L}$, to produce a stable rate to prevent high pressure and rate oscillations in the feed to the compressors, while at the same time producing close to the overall reference rate, q^{REF}, over the given planning horizon. This can be formulated as the problem of minimizing deviations from individual pad reference rates, q_l^{ref}, requiring that the sum of the individual pad reference rates is equal to the field reference rate q^{REF}.

Optimization problems with logical structures and conditions may benefit from starting with a generalized disjunctive programming (GDP) formulation (Raman and Grossmann, 1994) instead of an ad-hoc mixed-integer formulation, as a GDP formulation generally captures more directly the connections between the logical part and the constraints of the problem (Grossmann and Trespalacios, 2013). Based on a nonconvex and nonsmooth GDP formulation using the well model (2.8) explicitly in the disjunction, an MILP reformulation of the problem was derived in Knudsen and Foss (2013). The reformulation is based on a direct big-M type reformulation of the disjunction with a subsequent exact reformulation of the min-function. This formulation was shown to be computationally superior to equivalent reformulations of the disjunction with smooth approximations of the min-function. The min-function, however, is a disjunction itself (Nemhauser and Wolsey, 1988). Consequently, we can rewrite the nonsmooth well-model (2.8) as a disjunction based on the well inflow model and the maximum rate $q^{\max}$. This disjunction is then embedded into the main disjunction deciding if a well is producing or shut-in. We further include logical constraints for requiring a minimum shut-in time, τ_1, and a minimum production time, τ_2, between each succeeding shut-in cycle (Knudsen and Foss, 2013). Minimum production and shut-in times are typically required from an operational point of view, to avoid excessive actuation from too frequent shut-ins, causing wear and tear of the surface equipment. Some wells may

also require a minimum shut-in time to achieve sufficient pressure build-up due to certain properties of the well, or if liquids have accumulated in parts of the well prior to initialization of the cyclic scheme. We assume that the line pressure is kept constant, and that each well operates on a *fixed* wellhead pressure $p_{\mathrm{w},lj}$, with corresponding bottomhole pressure $m_{\mathrm{wf},lj} := m(e^{S_{lj}} p_{\mathrm{w},lj})$. The separation tank at the pad is not included in the system description. This separation will only impact the system by slightly lowering the inlet pressure on the pipeline leaving the pad. By combining the reference rate objectives described in the previous paragraph with the logical formulation for shutting in the wells, we obtain the GDP model

$$Z = \min \sum_{l \in \mathcal{L}} \max_k \left| q_l^{\mathrm{ref}} - q_{lk}^{\mathrm{Pad}} \right| \tag{3.12a}$$

s.t.

$$\sum_{l \in \mathcal{L}} q_l^{\mathrm{ref}} = q^{\mathrm{REF}}, \tag{3.12b}$$

$$q_{lk}^{\mathrm{pad}} = \sum_{j \in \mathcal{J}} q_{jk}, \qquad \forall l \in \mathcal{L}, k \in \mathcal{K} \tag{3.12c}$$

$$A_{lj} m_{ljk+1} = m_{ljk} + B_{lj} q_{ljk+1}, \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K}^{\mathrm{m}} \tag{3.12d}$$

$$m_{lj0} = m_{lj}^{\mathrm{init}}, \qquad \forall l \in \mathcal{L}, j \in \mathcal{J} \tag{3.12e}$$

$$\tilde{q}_{ljk} = \beta_{lj} \left(m_{ljk1} - m_{\mathrm{wf},lj} \right), \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K} \tag{3.12f}$$

$$\begin{bmatrix} Y_{ljk}^1 \\ q_{ljk} \geq q_{\mathrm{gc}} \\ \begin{bmatrix} Y_{ljk}^{11} \\ \tilde{q}_{ljk} \leq q^{\max} - \delta^{\mathrm{e}} \\ q_{ljk} = \tilde{q}_{ljk} \end{bmatrix} \vee \begin{bmatrix} Y_{ljk}^{12} \\ \tilde{q}_{ljk} \geq q^{\max} \\ q_{ljk} = q^{\max} \end{bmatrix} \end{bmatrix} \vee \begin{bmatrix} Y_{ljk}^2 \\ q_{ljk} = 0 \end{bmatrix},$$

$$\forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K} \tag{3.12g}$$

$$Y_{ljk}^1 \veebar Y_{ljk}^2, \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K} \tag{3.12h}$$

$$Y_{ljk}^1 \Leftrightarrow Y_{ljk}^{11} \veebar Y_{ljk}^{12}, \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K} \tag{3.12i}$$

$$\left(Y_{ljk-1}^1 \wedge Y_{ljk}^2 \right) \Rightarrow \bigwedge_{\tau=k+1}^{k+\tau_1-1} Y_{lj\tau}^2 \,,$$

$$\forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \{\mathcal{K} : 1 \leq k \leq K - \tau_1 + 1\} \tag{3.12j}$$

$$\left(Y_{ljk-1}^2 \wedge Y_{ljk}^1 \right) \Rightarrow \bigwedge_{\tau=k+1}^{k+\tau_2-1} Y_{lj\tau}^1 \,,$$

$$\forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \{\mathcal{K} : 1 \leq k \leq K - \tau_2 + 1\} \tag{3.12k}$$

$$Y_{ljk}^1, Y_{ljk}^{11}, Y_{ljk}^{12}, Y_{ljk}^2 \in \{\text{True, False}\}$$

Observe that in (3.12a) we have used the sum of the infinity norm between the per timestep pad rate q_{lk}^{pad} and the corresponding pad reference rate q_l^{ref} as the objective

function. Also note that $\mathcal{K}^{\mathrm{m}}$ in (3.12d) is the set $\mathcal{K}$ shifted one step backwards in time, i.e. $\mathcal{K}^{\mathrm{m}} = \{0, \ldots, K-1\}$. The constant δ^{e} in (3.12g) is included to enforce non-simultaneous satisfaction of both of the inequalities in the embedded disjunction, and is set typically to 0.001 in implementations (Vecchietti and Grossmann, 2000). The above logic-based problem formulation is in nonstandard GDP form (Raman and Grossmann, 1994; Grossmann and Trespalacios, 2013) due to the embedded disjunction (3.12g), but can be systematically transformed into standard GDP form by using the technique described (Vecchietti and Grossmann, 2000; Grossmann and Trespalacios, 2013). This leads to the following logically equivalent representation of the disjunction (3.12g),

$$\begin{bmatrix} Y^1_{ljk} \\ q_{ljk} \geq q_{\mathrm{gc}} \end{bmatrix} \vee \begin{bmatrix} Y^2_{ljk} \\ q_{ljk} = 0 \end{bmatrix}, \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K} \tag{3.13a}$$

$$\begin{bmatrix} Y^{11}_{ljk} \\ \tilde{q}_{ljk} \leq q^{\max} - \delta^{\mathrm{e}} \\ q_{ljk} = \tilde{q}_{ljk} \end{bmatrix} \vee \begin{bmatrix} Y^{12}_{ljk} \\ \tilde{q}_{ljk} \geq q^{\max} \\ q_{ljk} = q^{\max} \end{bmatrix} \vee \left[Y^{13}_{ljk} \right], \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K} \tag{3.13b}$$

$$Y^1_{ljk} \veebar Y^2_{ljk}, \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K} \tag{3.13c}$$

$$Y^{11}_{ljk} \veebar Y^{12}_{ljk} \veebar Y^{13}_{ljk}, \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K} \tag{3.13d}$$

$$Y^1_{ljk} \Leftrightarrow Y^{11}_{ljk} \veebar Y^{12}_{ljk}, \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K} \tag{3.13e}$$

$$Y^{13}_{ljk} \Leftrightarrow Y^2_{ljk}, \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K} \tag{3.13f}$$

The Boolean Y^{13} is introduced to allow the logical state $\neg Y^{11} \wedge \neg Y^{12}$. Transforming (3.12) into the standard form (3.13) allows us to assess different ways of reformulating the disjunctions into algebraic equations. Raman and Grossmann (1994) showed how linear GDPs can be converted to MILPs by introducing binary variables y with a one-to-one correspondence with the Boolean variables Y, and using either a Big-M reformulation (Nemhauser and Wolsey, 1988) or the convex hull description of general linear disjunctions derived by Balas (1985). The polyhedral convex hull description of linear disjunctions is based on disaggregation of the variables in the disjunction and assuring that only one of the terms in the disjunction is active. The convex hull reformulation yields an LP relaxation at least as tight or tighter than big-M reformulations, and hence generally stronger lower bounds and reduced computation time (Grossmann and Trespalacios, 2013; Vecchietti et al., 2003). Still, the convex hull reformulation requires more variables and constraints than big-M reformulations. This may translate into longer computation time for some problems (Vecchietti et al., 2003; Grossmann and Trespalacios, 2013). Based on our computational experience, we observed that a full convex hull reformulation of the disjunctions (3.13a)–(3.13b) resulted in longer computation times than using the convex hull reformulation for (3.13a) and a big-M reformulation of (3.13b), while only a marginal improvement in the LP relaxation was observed when using the full convex hull reformulation. Disaggregation of q_{ljk} in (3.13a) then yields

$$q_{ljk} = q^1_{ljk} + q^2_{ljk}, \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K} \tag{3.14}$$

$$y^1_{ljk} + y^2_{ljk} = 1, \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K} \tag{3.15}$$

$$q^1_{ljk}, q^2_{ljk} \geq 0, \tag{3.16}$$

which by substitution of the disaggregated variables into (3.13a) simplifies to

$$q_{\mathrm{gc}} y^1_{ljk} \leq q_{ljk} \leq y^1_{ljk} q^{\max}, \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K} \tag{3.17}$$

since $q^2_{ljk} = 0$, and $\tilde{q}_{ljk}$ does not appear in this disjunction. The big-M reformulation of (3.13b) is obtained by rewriting the equality constraints as two inequalities and introducing big-M parameters for each term. The logical proposition (3.12j)–(3.12k) for the minimum shut-in and production time can be transformed to linear algebraic inequalities by replacing the implications with disjunctions, converting the resulting disjunction to conjunctive normal form and then replacing the logical relations with its algebraic counterparts, see Raman and Grossmann (1991). This eventually leads to the following primal MILP problem.

$$Z = \min \sum_{l \in \mathcal{L}} f_l \tag{3.18a}$$

s.t.

$$q^{\mathrm{REF}} = \sum_{l \in \mathcal{L}} q^{\mathrm{ref}}_l, \tag{3.18b}$$

$$f_l \geq q^{\mathrm{ref}}_l - q^{\mathrm{pad}}_{lk}, \qquad \forall l \in \mathcal{L}, k \in \mathcal{K} \tag{3.18c}$$

$$f_l \geq -q^{\mathrm{ref}}_l + q^{\mathrm{pad}}_{lk}, \qquad \forall l \in \mathcal{L}, k \in \mathcal{K} \tag{3.18d}$$

$$q^{\mathrm{Pad}}_{lk} = \sum_{j \in \mathcal{J}} q_{jk}, \qquad \forall l \in \mathcal{L}, k \in \mathcal{K} \tag{3.18e}$$

$$A_{lj} m_{ljk+1} = m_{ljk} + B_{lj} q_{ljk+1}, \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K}^{\mathrm{m}} \tag{3.18f}$$

$$m_{lj0} = m^{\mathrm{init}}_{lj}, \qquad \forall l \in \mathcal{L}, j \in \mathcal{J} \tag{3.18g}$$

$$\tilde{q}_{ljk} = \beta_{lj} \left(m_{ljk1} - m_{\mathrm{wf},lj} \right), \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K} \tag{3.18h}$$

$$q_{\mathrm{gc}} y^1_{ljk} \leq q_{ljk}, \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K} \tag{3.18i}$$

$$q_{ljk} \leq q^{\max} y^1_{ljk}, \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K} \tag{3.18j}$$

$$\tilde{q}_{ljk} - q^{\max} + \delta^{\mathrm{e}} \leq M^{11} \left(1 - y^{11}_{ljk} \right), \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K} \tag{3.18k}$$

$$q_{ljk} - \tilde{q}_{ljk} \leq M^{11} \left(1 - y^{11}_{ljk} \right), \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K} \tag{3.18l}$$

$$\tilde{q}_{ljk} - q_{ljk} \leq M^{11} \left(1 - y^{11}_{ljk} \right), \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K} \tag{3.18m}$$

$$q^{\max} - \tilde{q}_{ljk} \leq M^{12} \left(1 - y^{12}_{ljk} \right), \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K} \tag{3.18n}$$

$$q_{ljk} - q^{\max} \leq M^{12} \left(1 - y^{12}_{ljk} \right), \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K} \tag{3.18o}$$

$$q^{\max} - q_{ljk} \leq M^{12} \left(1 - y^{12}_{ljk} \right), \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K} \tag{3.18p}$$

$$y^{11}_{ljk} + y^{12}_{ljk} + y^{13}_{ljk} = 1, \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K} \tag{3.18q}$$

$$y^{11}_{ljk} + y^{12}_{ljk} = y^1_{ljk}, \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K} \tag{3.18r}$$

$$y^1_{ljk} + y^2_{ljk} = 1, \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K} \tag{3.18s}$$

$$y^{13}_{ljk} = y^2_{ljk}, \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \mathcal{K} \tag{3.18t}$$

$$y^1_{ljk-1} + y^2_{ljk} \leq 1 + y^2_{lj\tau}, \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \{\mathcal{K} : 1 \leq k \leq K - \tau_1 + 1\}, \\ \tau \in [k+1, \min\{k + \tau_1 - 1, K\}] \tag{3.18u}$$

$$y^2_{ljk-1} + y^1_{ljk} \leq 1 + y^1_{lj\tau}, \qquad \forall l \in \mathcal{L}, j \in \mathcal{J}, k \in \{\mathcal{K} : 1 \leq k \leq K - \tau_2 + 1\}, \\ \tau \in [k+1, \min\{k + \tau_2 - 1, K\}] \tag{3.18v}$$

Introducing the auxiliary variables f_l with the constraints (3.18c)–(3.18d) exactly reformulates the nonsmooth objective function (3.12a). Note that although the MILP formulation (3.18) can be considerably reduced by substituting dependent binary variables, we leave this reduction and tightening of the problem to the presolve routines in the MILP solver (such as CPLEX and Gurobi).

Similar constraints to those of (2.31m)–(2.31n) for the minimum shut-in and production time for each shut-in cycle have been used for defining minimum up and down time requirements in the hydro-thermal unit commitment (UC) problem in power generation (Takriti et al., 2000), as well as restricted length set-up sequences in single-item lot sizing (Pochet and Wolsey, 2006). Several authors have studied the tightness of different formulations of these constraints in the context of the UC problem (Lee et al., 2004; Rajan and Takriti, 2005). Rajan and Takriti (2005) derives the convex hull polytope of an extended variable formulation of the minimum up and downtime constraints, in which start-up and shut-down costs are present. As the problem (3.12) contains no costs on the shut-ins and start-ups, however, we can omit additional binary variables for modeling start-up and shut-ins transitions, and limit the formulation of the minimum shut-in and production time constraints using only the on/off variables y^1_{ljk} and y^2_{ljk} with the constraints (3.18u)–(3.18v). Note that these constraints are projections of the so-called turn on/off inequalities derived in Rajan and Takriti (2005) onto the space of the y^1_{ljk} and y^2_{ljk} variables (Pochet and Wolsey, 2006, Ch. 11).

3.5 Lagrangian relaxation scheme

Lagrangian relaxation is an efficient decomposition technique for problems with a block-separable structure in the constraint set and few binding or coupling constraints (Guignard, 2003). This structure is often found in optimization problems with some sort of network flow and spatial distribution (Foss et al., 2009; Sagastizábal, 2012). The primal problem (3.18) clearly renders this type of problem structure, as the only constraint that links the L pads together is the constraint (3.18b) requiring that the sum of the individual pad reference rates must be equal the total requested rate q^{REF}. Consequently, by dualizing this constraint, we will obtain a block-separable, decomposable problem with one subproblem for each pad $l \in \mathcal{L}$. Let $\lambda \in \mathbb{R}$ be the Lagrangian multiplier associated with the reference rate equality constraint (3.18b). The Lagrangian relaxation is then given by

$$Z_{\text{LR}}(\lambda) = \min \sum_{l \in \mathcal{L}} f_l + \lambda \left(\sum_{l \in \mathcal{L}} q^{\text{ref}}_l - q^{\text{REF}} \right) \tag{3.19}$$

s.t.

eq. (3.18c)–(2.31n),

hence rendering a spatial decomposition over the set $\mathcal{L}$, such that solving the Lagrangian (3.19) means solving l independent subproblems defined by

$$\begin{aligned} Z_{\mathrm{LR},l}(\lambda) = \min\ & f_l + \lambda q_l^{\mathrm{ref}} \\ \text{s.t.}\ & \\ & \text{eq. (3.18c)–(2.31n) for given } l \in \mathcal{L}. \end{aligned} \tag{3.20}$$

For any $\lambda \in \mathbb{R}$, the Lagrangian (3.19) is a relaxation of the primal problem (3.18), and consequently defines a lower bound on Z (Geoffrion, 1974). The solution of the Lagrangian (3.19) yields a lower bound on Z at least as tight as the LP relaxation (Geoffrion, 1974). The dual variable λ is a marginal cost on the demand constraint (3.18b) of the pads' reference rate q_l^{ref} relative to the given total reference rate (or demand) q^{REF}, by pricing the cost of satisfying an additional unit in total gas demand (Sagastizábal, 2012).

Obtaining the best possible value $Z_{\mathrm{LR}}(\lambda)$ of the Lagrangian relaxation (3.19), that is, the one providing the tightest lower bound on Z, requires finding the optimal multiplier λ. This problem defines the Lagrangian Dual (Geoffrion, 1974),

$$Z_{\mathrm{D}} = \max_{\lambda} Z_{\mathrm{LR}}(\lambda). \tag{3.21}$$

Due to the inherent nonconvexity of the binary variables in the primal MILP (3.18), we may have a nonzero duality gap, i.e. $Z - Z_{\mathrm{D}} > 0$. The Lagrangian $Z_{\mathrm{LR}}(\lambda)$ can be shown to be piecewise linear and a concave function of λ (Nemhauser and Wolsey, 1988), while nondifferentiable at the points where the optimal solution of (3.19) is not unique (Guignard, 2003). The solution of (3.19) will in most cases be primal infeasible in the sense that the dualized reference rate constraint (3.18b) will be violated. Consequently, some technique using the information from solving the Lagrangian relaxation and/or the Lagrangian dual must be developed for generating primal feasible solutions.

3.5.1 Solving the Lagrangian dual

The two most common classes of algorithms for solving the Lagrangian dual are subgradient-type methods and methods based on the cutting-plane approach (Guignard, 2003). A variety of implementations and modifications of algorithms within these two classes exists. Subgradient-type methods are based on using the dualized constraint, i.e. (3.18b) in the Lagrangian relaxation (3.19), as the subgradient for $Z_{\mathrm{LR}}(\lambda)$ in the space of the dual variables. The standard subgradient method (Held et al., 1974) is widely applied for updating Lagrangian multipliers due to its ease of implementation and capability of providing good multiplier updates for many problems. The method may, however, require extensive tuning of the step-size parameters to obtain good practical convergence, and lacks a true termination criteria.

The cutting-plane algorithm for solving the Lagrangian dual (3.21) is based on maximizing a *cutting-plane model* $\tilde{Z}_{\mathrm{LR}}(\lambda)$ of $Z_{\mathrm{LR}}(\lambda)$, given by

$$\tilde{Z}_{\mathrm{LR}}(\lambda) := \min \left\{ \sum_{l \in \mathcal{L}} f_l^u + \lambda \left(\sum_{l \in \mathcal{L}} q_l^{\mathrm{ref},u} - q^{\mathrm{REF}} \right), \quad u = 1, \dots U \right\}, \tag{3.22}$$

i.e., a concave polyhedral upper approximation $\tilde{Z}_{\mathrm{LR}}$ of Z_{LR} (Frangioni, 2005). The solutions $(f_l^u, q_l^{\mathrm{ref},u})$ are obtained by solving the Lagrangian relaxation (3.19) for a given input multiplier λ. Replacing $Z_{\mathrm{LR}}(\lambda)$ in (3.21) with the cutting-plane model (3.22) leads to a maxmin problem, which can be equivalently formulated as the linear program (LP)

$$\tilde{Z}_{\mathrm{D}} = \max_{\lambda,\eta} \ \eta \tag{3.23a}$$

s.t.

$$\eta \leq \sum_{l\in\mathcal{L}} f_l^u + \lambda \left(\sum_{l\in\mathcal{L}} q_l^{\mathrm{ref},u} - q^{\mathrm{REF}} \right), \qquad u = 1, ..., U \tag{3.23b}$$

with $\eta \in \mathbb{R}$. For each iteration n of the primal-dual loop solving (3.19) and (3.23), a new cut is added to (3.23) to improve the cutting-plane model (3.22) of $Z_{\mathrm{LR}}(\lambda)$. Using the cutting-plane algorithm (3.23) for solving the Lagrangian dual (3.21) provides at each step an upper and lower bound on Z_{D} (Frangioni, 2005),

$$\eta^n \geq Z_{\mathrm{D}} \geq Z_{\mathrm{LR}}^n. \tag{3.24}$$

The method terminates when these bounds coincide. The cutting-plane algorithm has finite convergence (Hiriart-Urruty and Lemarechal, 1993, Ch. XII 4.2), but suffers in its basic form from an inherent instability. In early iterations, the cutting-plane model (3.22) of $Z_{\mathrm{LR}}(\lambda)$ is insufficient to provide a bounded problem. Moreover, the sequence λ^n of dual solutions lacks local properties (Frangioni, 2005). These characteristics often lead to oscillatoric performance and slow convergence of the cutting-plane algorithm in practice. The problem of early oscillations and unbounded solutions of (3.23) can be addressed by adding bounds on the dual variables (Hiriart-Urruty and Lemarechal, 1993), and adding known solutions $(\hat{f}_l, \hat{q}_l^{\mathrm{ref}})$ of (3.18) and (3.19) to improve the cutting-plane model $\tilde{Z}_{\mathrm{LR}}(\lambda)$ of $Z_{\mathrm{LR}}(\lambda)$ (Frangioni, 2005). Many rigorous stabilization methods have been suggested to increase the local properties and mitigate the oscillations of the dual solutions from the cutting-plane algorithm, see (Hiriart-Urruty and Lemarechal, 1993, Ch. XV) and Frangioni (2005) for a thorough review. The above approaches for stabilizing the cutting-plane algorithm for solving the Lagrangian dual may be combined and tailored to derive efficient algorithms for updating the Lagrangian multipliers, or possibly combined with a subgradient method (Mouret et al., 2011).

To stabilize the cutting-plane formulation for solving the Lagrangian dual (3.23), we adopt a trust-region scheme from (Hiriart-Urruty and Lemarechal, 1993, Ch. XV 2.1) and Kallehauge et al. (2006). The trust-region approach provides a simple device for coping with instabilities in the cutting-plane algorithm when applied to Lagrangian duals with relatively few dual variables, and is closely related to the boxstep method (Marsten et al., 1975). The algorithm ensures that the next multiplier update λ^{n+1} is never further away than a distance Δ^n (with respect to the 1-norm) from the current *stability center* $\bar{\lambda}$. The algorithm is as follows:

Algorithm 1 : Trust-region cutting-plane algorithm
Initialization: Select an initial trust-region Δ^1, stability center $\bar{\lambda}$, a termination tolerance $\delta^{\mathrm{TR}} \geq 0$, and a descent coefficient $\sigma \in (0,1)$. Set $\lambda^1 = \bar{\lambda}$, $U = 1$, and solve (3.19) to obtain $Z_{\mathrm{LR}}(\lambda^1)$.

Step 1: Solve the LP

$$\max_{\lambda,\eta} \eta \tag{3.25a}$$

s.t.

$$\eta \leq \sum_{l\in\mathcal{L}} f_l^u + \lambda \left(\sum_{l\in\mathcal{L}} q_l^{\mathrm{ref},u} - q^{\mathrm{REF}} \right), \qquad \forall u = 1,\ldots,U \tag{3.25b}$$

$$|\lambda - \bar{\lambda}| \leq \Delta^n, \tag{3.25c}$$

$$\lambda \geq -1, \tag{3.25d}$$

to obtain λ^{n+1} and η^n. Compute

$$\tilde{\rho} := \eta^n - Z_{\mathrm{LR}}(\bar{\lambda}) \geq 0. \tag{3.26}$$

Step 2: If $\tilde{\rho} \leq \delta^{\mathrm{TR}}$, terminate with $Z_{\mathrm{D}} = Z_{\mathrm{LR}}(\bar{\lambda})$.
Step 3: Compute $Z_{\mathrm{LR}}(\lambda^{n+1})$.
Step 4: If

$$Z_{\mathrm{LR}}(\lambda^{n+1}) \geq Z_{\mathrm{LR}}(\bar{\lambda}) + \sigma\tilde{\rho}, \tag{3.27}$$

update the center: $\bar{\lambda} = \lambda^{n+1}$. Otherwise, leave center unchanged.
Step 5: Update trust-region; compute the ratio

$$\rho := \frac{Z_{\mathrm{LR}}(\lambda^{n+1}) - Z_{\mathrm{LR}}(\bar{\lambda})}{\tilde{\rho}}. \tag{3.28}$$

If $\rho = 1$, set $\Delta^{n+1} = 1.5\Delta^n$. If $\rho < 0$, set $\Delta^{n+1} = 0.8\Delta^n$.
Set $U = n+1$ and update the iteration counter $n = n+1$; repeat from *Step 1.*

By evaluating in *Step 4* the actual increase of $Z_{\mathrm{LR}}(\lambda)$ with the *predicted* increase given by $\tilde{\rho}$ in (3.26), an ascent-step (serious step) is only declared if a sufficient increase is obtained relative to the Armijo-like parameter σ. Otherwise, the current iteration is a null-step, while a new cut is added to (3.25) to improve the polyhedral approximation of $Z_{\mathrm{LR}}(\lambda)$. Comparing with the subgradient method, the above scheme guarantees an increase for each serious step. We refer to the above algorithm as a trust-region approach in the sense that we assume the cutting-plane model (3.22) to be a good approximation of $Z_{\mathrm{LR}}(\lambda)$ inside the trust region defined by Δ^n. The criteria in *Step 5* for updating the trust-region may be tuned to the specific problem to improve the performance of the algorithm. Observe that the linear dual of the nonstabilized cutting-plane algorithm (3.25a)–(3.25b) for solving the Lagrangian dual yields the primal master problem in the Dantzig-Wolfe decomposition method (Frangioni, 2005).

The lower bound (3.25d) is added to ensure values of λ that gives bounded Lagrangian subproblems (3.20). The proof of the numerical value of this bound

is given in 3.A. By a similar proof, it can be shown by using Fourier-Motzkin elimination that for a *single* reference rate q^{REF}, any nonnegative value of λ will cause $q_l^{\text{ref}} = 0$, and consequently the optimal solution $f_l = q_l^{\text{pad}} = 0$ for each of the subproblems (3.20). Hence, we can add this a priori solution to (3.25) to improve the initial cutting-plane model of $Z_{\text{LR}}(\lambda)$. Note that in practice, though, it may be sufficient to select an initial small trust region Δ^1 together with a good estimate for λ^1 to prevent optimal values of λ less than -1. This strategy, however, is counter intuitive for trust-region methods.

3.5.2 Generating primal feasible solutions

Obtaining primal feasible solutions in a Lagrangian relaxation scheme normally requires the use of some, possibly problem specific, heuristic. Either or both of the solutions from the Lagrangian relaxation and the Lagrangian dual may be exploited to generate primal feasible solutions, often using some type of greedy algorithm. The solutions from the Lagrangian relaxation often becomes almost primal feasible (with respect to the dualized constraints) with increasing number of iterations n. As Z_{LR} is a lower bound on Z, the solution of the Lagrangian relaxation may in some way aid the search for primal feasible solutions.

Solving the $|\mathcal{L}|$ Lagrangian subproblems generates a schedule (or sequence) of shut-ins for each well on the given pad l together with a reference-rate q_l^{ref}. In other words, the solutions of the subproblems prevent well liquid loading. The shut-in schedule is completely described by the value of the binary variables y_{ljk}^d. This schedule will be affected by the value of the q_l^{ref}, which in turn is a function of the marginal cost λ and the total gas-rate capacity of wells on the pad. That is, how much gas the wells can produce and the level of the peak-rates after shut-ins. To obtain a primal feasible solution and hence an upper bound Z^{UB}, we fix the binary variables y_{ljk}^d to the values obtained from solving the Lagrangian relaxation subproblems (3.20), and solve the primal problem (3.18). This variable fixing corresponds to fixing the well schedule. The primal problem (3.18) reduces to an LP, where the pad reference rates q_l^{ref} now are the only degrees of freedom. The LP solution of this variable fixing returns a set of primal feasible reference rates $\bar{q}_l^{\text{ref}}$ which minimizes (in terms of the sum of the infinity norm) the deviation between the currently fixed production and the reference rates .

In each iteration n, we check if the deviation between the primal feasible reference rates $\bar{q}_l^{\text{ref}}$ obtained by the binary variable-fixing, and the corresponding solution q_l^{ref} from the Lagrangian relaxation, is larger than a certain threshold,

$$\left|q_l^{\text{ref}} - \bar{q}_l^{\text{ref}}\right| > \delta^{\text{ref}}. \tag{3.29}$$

If it is, we perform a local search for each of these pads $l \in \bar{\mathcal{L}} \subseteq \mathcal{L}$ to search for an improved shut-in schedule. Let $\bar{Z}_l = \bar{f}_l$ be the solution from (3.18a) for pad l obtained by the binary variable fixing. The local search is then performed by fixing

the reference rate, $q_l^{\text{ref}} \equiv \bar{q}_l^{\text{ref}}$, and solving the MILP

$$Z_l = \min\ f_l \tag{3.30a}$$

s.t.

$$f_l \leq \bar{Z}_l, \tag{3.30b}$$

$$f_l \geq \bar{q}_l^{\text{ref}} - q_{lk}^{\text{pad}}, \qquad \forall k \in \mathcal{K} \tag{3.30c}$$

$$f_l \geq -\bar{q}_l^{\text{ref}} + q_{lk}^{\text{pad}}, \qquad \forall k \in \mathcal{K} \tag{3.30d}$$

$$\sum_{(l,j,k)\in\left\{(\bar{\mathcal{L}}\times\mathcal{J}\times\mathcal{K}):\,\bar{y}^1_{ljk}=1\right\}} (1 - y^1_{ljk}) \quad + \quad \sum_{(l,j,k)\in\left\{(\bar{\mathcal{L}}\times\mathcal{J}\times\mathcal{K}):\,\bar{y}^2_{ljk}=1\right\}} (1 - y^2_{ljk}) \leq r, \tag{3.30e}$$

eq. (3.18e)–(2.31n) for given $l \in \bar{\mathcal{L}}$.

The constraint (3.30e) is a local branching type constraint (Fischetti and Lodi, 2003), where $\bar{y}^1_{ljk}, \bar{y}^2_{ljk}$ is the shut-in schedule obtained from the solving the Lagrangian relaxation, and the parameter r is a neighborhood radius (or Hamming distance). This constraint allows at most r switchings of wells from on to off and from off to on, given by the first and the second term, respectively. The radius r must be selected sufficiently large for improved solutions to exist, while at the same time being small enough such that the local search can be performed efficiently. We implement a strategy where we initialize $r = \lceil (1/3)\,|\mathcal{K}|\,|\mathcal{J}| \rceil$, and increase r by a factor of 2 at most five times if (3.30) is infeasible, and we allocate a maximum CPU time between 15-90 seconds for each local search, depending on the problem size.

The complete Lagrangian relaxation scheme, comprising the Lagrangian relaxation, the primal heuristic, Lagrangian dual and the termination criteria, is outlined in Fig. 3.7. If the loop in the Lagrangian scheme terminates with a nonzero duality gap, we perform an extended but similar local search like (3.30) around the *best found* feasible solution. In particular, we remove the local branching constraint (3.30e) to explore a larger search space and perform a time-limited search for each subproblem $l \in \mathcal{L}$, in which we now fix q_l^{ref} to its corresponding argument of the best found feasible solution Z^{UB}. In contrast, the local search (3.30) inside the primal heuristic part in Fig. 3.7 is performed around the currently recovered primal feasible solution. Note that we preserve primal feasibility when performing both of these local search methods. If the solution improves the upper bound, we update the best feasible solution and terminate the search.

The proposed model-based optimization scheme is depicted in Fig. 3.8, showing the connections between the different modules, the equation numbers for the respective optimization problems, and the software used to implement and solve these problems. The scheduling of the well shut-ins is in practice likely to be performed in a receding horizon fashion, see for instance the textbook on model predictive control by Rawlings and Mayne (2009). This is illustrated in Fig. 3.8 by the upper dashed feedback loop, in which a new shut-in schedule is computed by the Lagrangian relaxation scheme with a certain interval. This re-optimization, and how often it is performed, is not further detailed in this paper. Note though that the updating of the proxy model described in section 3.1, is normally performed less frequently than the re-optimization of the shut-in schedule.

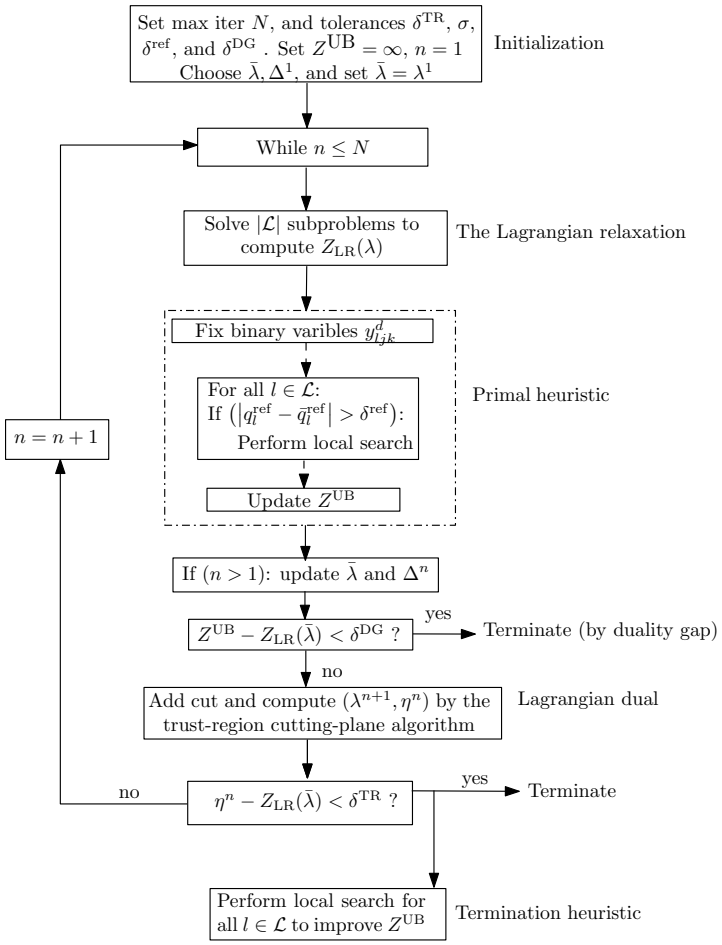


Fig. 3.7: Schematic description of the Lagrangian relaxation scheme.

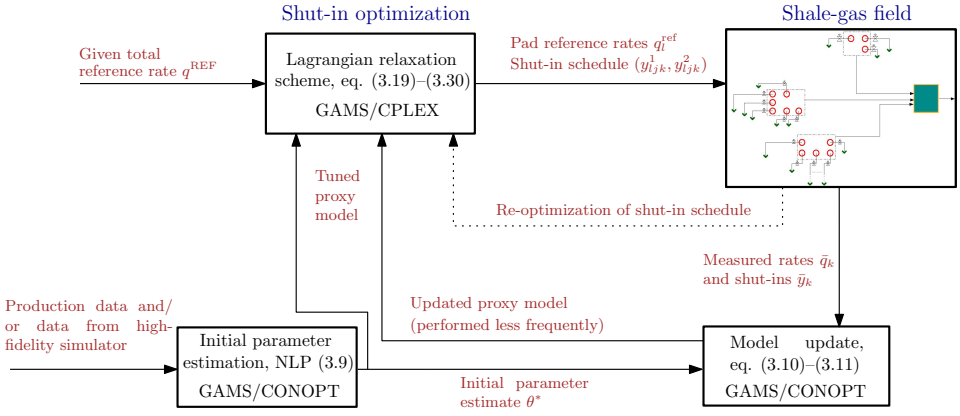


Fig. 3.8: Illustration of the computational modules and their interconnections with the actual field.

3.6 Computational Results

The performance of the Lagrangian relaxation scheme is assessed by solving a series of test sets of the shale-gas multi-pad system with the structure shown in Fig. 3.1. The sets are generated with different number of pads $|\mathcal{L}|$, while for simplicity we have used six wells per pad l. The proxy models in the numerical examples are realized and tuned against the multi-fracture reference model as described in Section 3.3, using three different formation permeabilities $k_{\mathrm{m}} \in \left[1 \times 10^{-4}, 3 \times 10^{-4}, 5 \times 10^{-4}\ \mathrm{mD}\right]$. When using equal fracture geometry and initial pressure, the value of the formation permeability is decisive for the time when the gas rate first crosses the critical rate q_{gc}, hence giving somewhat different performance of the wells. We assume equal permeability for each well on a given pad, but we will assume that each well has been operated at different lengths of time, hence giving different initial pseudopressures m_{lj}^{init}. We use a one-day fixed timestep, and set $\tau_1 = \tau_2 = 2$ days.

The Lagrangian scheme outlined in Fig. 3.7 is implemented in GAMS (Brooke et al., 2011), using IBM CPLEX v12.3 to solve the LPs and the MILPs. The problems are solved using a Dell laptop with Intel I7 quad-core CPU and 8GB of RAM, using deterministic parallel mode with up to 8 threads. The tolerances in the trust-region algorithm are set to $\delta^{\mathrm{TR}} = 10^{-6}$, $\sigma = 0.01$ and $\delta^{\mathrm{ref}} = 0.01$, while we set the duality gap tolerance to $\delta^{\mathrm{DG}} = 0.01$ (cf. Fig. 3.7) and the maximum number of iterations $N = 15$. The trust-region is initially set to $\Delta^1 = 0.3$, while $\bar{\lambda}$ is initially set to -0.5, the arithmetic mean of the lower bound (3.25d) and value of the Lagrange multiplier of (3.18b) obtained by the LP relaxation of the primal MILP (3.18). For the MILP Lagrangian relaxation subproblems (3.20), we use a relative gap less than 1% or an absolute gap less than 10^{-4} as stopping criteria, and allocate a maximum CPU time of four hours for each of the subproblems. The absolute-gap criteria is added as the LP relaxation of the subproblems tends to be zero or a very small value, since we are minimizing the deviation between a reference rate and the actual rate. Branching priorities are added to all of the MILPs, in which y^1 are given higher priority than y^2 and y^{13}, which again have higher priority than y^{11} and y^{12}. The size of the problems are shown in Table 3.3. Note that the size of all of the MILPs are significantly reduced by the presolve routines in CPLEX. All reported times are real (clock) time, and the reported relative duality gaps (in %) are defined as (Pochet and Wolsey, 2006, Ch. 3),

$$\mathrm{DG} := 100[\%] \times \frac{|Z^{\mathrm{UB}} - Z_{\mathrm{LB}}|}{|Z^{\mathrm{UB}}|}, \tag{3.31}$$

where $Z_{\mathrm{LB}} = Z_{\mathrm{D}}$ for the Lagrangian scheme, while Z_{LB} is the best LP relaxation bound in the fullspace solution. The results of the Lagrangian scheme is in Table 3.4 compared with the fullspace solution of the primal MILP model, that is, solving (3.18) by a direct MILP approach. For the purpose of this comparison, we allocate a solution time for the fullspace approach approximately equal to the time spent by the Lagrangian scheme.

Table 3.4 shows characteristics of the results of applying the Lagrangian relaxation scheme to sets with different number of pads and planning horizons. All of the sets in Table 3.4, except for the sets with $|\mathcal{L}| = 6$ and $K = 14$ and 21,

Table 3.3: Problem size of test instances in Table 3.4.

$\|\mathcal{L}\|$	K	Binary var	Cont. var	Constraints
3	7	720	895	2666
3	14	1350	1672	5103
3	21	1980	2449	7580
6	7	1140	1789	5330
6	14	2700	3343	10244
6	21	3960	4897	15158
9	7	2160	2686	7994
9	14	4050	5014	15365
9	21	5940	7345	22736

Table 3.4: Computational results of the Lagrangian relaxation scheme and a fullspace approach.

			Lagrangian scheme					Fullspace			
$\|\mathcal{L}\|$	K	q^{REF} †	DG [%]	Time [min]	#iter	Z^{UB}†	Z_{D}†	DG [%]	Time [min]	Z^{UB}†	Best Z_{LB}†
3	7	16.0	22.9	1	7	0.41	0.32	100	2	0.50	0
3	14	16.0	11.9	20	6	0.49	0.43	100	20	1.60	0
3	21	16.0	11.7	151	5	0.52	0.46	100	84*	4.37	0
6	7	43.2	2.4	2	7	2.09	2.04	90.4	2	2.57	0.25
6	14	43.2	0.0	33	3	3.59	3.59	64.6	33	4.30	1.52
6	21	43.2	0.0	580	5	4.13	4.13	63.3	580	6.67	2.45
9	7	53.0	5.4	3	7	0.91	0.86	100	3	4.83	0
9	14	53.0	14.6	43	9	1.55	1.32	100	43	9.42	0
9	21	53.0	13.4	643	9	1.73	1.50	100	107*	18.1	0

† Results Z (lower and upper bounds) and the reference rate q^{REF} are given in 10^4 m^3/d.

* Fullspace problems eventually ran out of memory due to size of nodefile in the order ∼10GB.

respectively, are terminated with the criteria $\tilde{\rho} \leq \delta^{\mathrm{TR}}$ in step 2 of the trust-region cutting-plane algorithm for solving the Lagrangian dual (3.21). None of the sets reach the maximum number of iterations. Several of the sets terminate with duality gaps greater than 10%. Hence, if only evaluated by the size of the duality gap, the performance of the Lagrangian relaxation scheme applied to these sets may be considered as only average. The final duality gaps are not necessarily worse for longer compared with short planning horizons K. For planning horizons longer than three weeks, the time required to solve each of the MILP subproblems (3.20) to global optimality becomes computationally prohibitive. The difficulty in solving the MILP (3.18) becomes evident when comparing the solution characteristics of the Lagrangian relaxation scheme with solving (3.18) as one fullspace problem. The Lagrangian relaxation scheme consistently finds significantly improved lower and upper bounds compared with the fullspace solution. Solving (3.18) by a fullspace branch-and-cut algorithm also highlights the difficulty in improving the lower bound; six out of nine sets in Table 3.4 are terminated with a 100% duality gap when solved with the direct approach, meaning that the best lower bound remains close to zero (cf. (3.31)) throughout the enumeration.

The duality gap by termination of the Lagrangian relaxation scheme depends on the given reference rate q^{REF}. This is shown in Fig. 3.9 for the set consisting of three pads and a one week planning horizon. For low values of q^{REF}, the gas-rate capacity of the multi-pad system is sufficient to meet the reference rate, seen by the

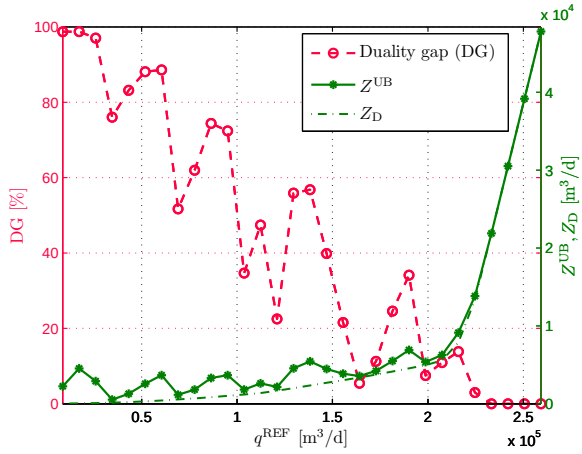


Fig. 3.9: The duality gap DG, the best primal feasible solution Z^{UB} (i.e. sum of maximum deviations from q_l^{ref}) and the lower bound Z_{D} as a function of different values of given q^{REF}, using the configuration of $(|\mathcal{L}|, K)$ as in the first instance in Table 3.4.

low value of Z^{UB}, and hence a low total deviation between the produced pad rates and the corresponding pad reference rates. The duality gap, however, is large for low values of q^{REF}, and decreases for higher values of q^{REF}. The lower bound Z_{D} is seen in Fig. 3.9 to be close to zero for low values of q^{REF}, in which Z^{UB} dominates (3.31), causing large duality gaps. In a certain range of values for q^{REF}, it is seen that the size of the duality gap compared with the value of Z^{UB} is balanced. For high values of q^{REF}, the duality gap converges to zero while the optimal solution value increases rapidly, leading to unacceptable deviations between the produced rates and the reference rates.

The trust-region cutting-plane algorithm works well for solving the Lagrangian dual of the problem. All sets terminate with an optimal dual variable $\bar{\lambda}$ after 4-9 iterations. The method is also observed to be fairly robust with respect to initial choices of $\bar{\lambda}$ and Δ^1. Still, choosing a higher value of the initial trust region and a lower initial value of $\bar{\lambda}$ (closer to -1) leads to slightly more iterations and to increased oscillations in the sequence of solutions λ^n. In this case the trust-region becomes ineffective. The use of the local search heuristic (3.30) mainly improved the upper bound Z^{UB} whenever the binary-fixing heuristic found an improved primal feasible solution. In these cases, however, the local search often reduced the upper bound by as much as 50% to 200% of its value from the previous iteration.

The practical performance from the solution of (3.18) is illustrated in Fig. 3.10 showing the total rate for the set in Table 3.4 with 9 pads and a three week planning horizon (i.e. the last instance). The pad reference rate and the total produced rate is shown for three of the nine pads in Fig. 3.11. The total gas rate for the pads is shown to tightly follow the given total reference rate q^{REF} over the planning horizon, while each of the pads rates follow the optimal pad reference rates with only small fluctuations as seen in Fig. 3.11. Consequently, any high and low peak rates are avoided, both in the total rate, and for the rates from the individual pads. In this context, the actual optimized result by the Lagrangian relaxation scheme

clearly renders a good feasible solution despite the duality gap being more than 13%. Comparing the subfigures in Fig. 3.11 also highlights a qualitative difference in the performance of the gas rates q_{lk}^{pad} from the pads; the wells with the highest permeability, i.e. the wells on pad 5, preserve a higher rate after shut-ins. Hence, fewer shut-ins are needed, and the total rate from the pad is more stable as seen in Fig. 3.11c.

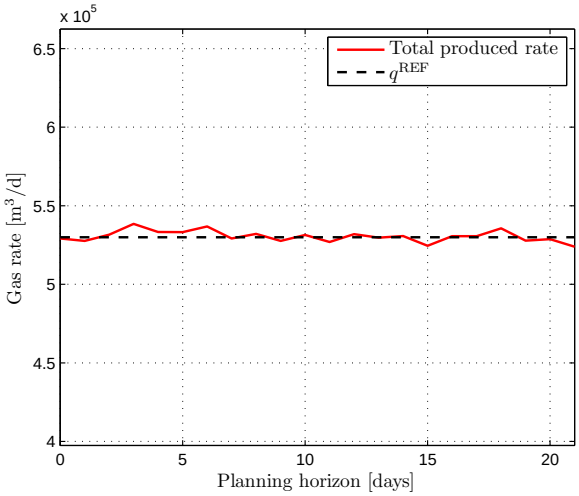
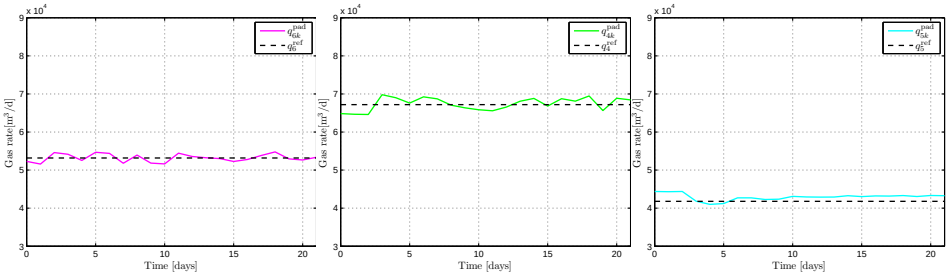


Fig. 3.10: Total rate from $|\mathcal{L}| = 9$ pads with a three week planning horizon solved with the Lagrangian relaxation scheme.



(a) Pad 6, $k_{\text{m}} = 1 \times 10^{-4}$ mD. **(b)** Pad 4, $k_{\text{m}} = 3 \times 10^{-4}$ mD. **(c)** Pad 5, $k_{\text{m}} = 5 \times 10^{-4}$ mD.

Fig. 3.11: Pad reference rates and produced rates for three of the pads in the last instance in Table 3.4.

3.7 Discussion

The main contribution in this paper is the development of a decomposable Lagrangian relaxation scheme from a fullspace GDP formulation for field-wide shut-in scheduling in large shale-gas systems. The scheme was observed to consistently solve

the Lagrangian dual and find feasible solutions through a heuristic using the solutions from the Lagrangian relaxation. The Lagrangian scheme clearly outperformed a fullspace approach, both in terms of the duality gap and the lower and upper bounds. The schemes' ability to generate small duality gaps, however, is observed to depend somewhat on the problem configuration. Nevertheless, the solutions obtained are still considered as good feasible solutions. If, however, small duality gaps are desirable (or required), one may consider using a strategy based on the Augmented Lagrangian. Under mild assumptions, Augmented Lagrangians ensure zero duality gaps (Hiriart-Urruty and Lemarechal, 1993, Ch. XII 5.2), but unfortunately the augmented quadratic term destroys separability of the dual function. Some approximation is therefore required to obtain separability of the relaxation, see Li and Ierapetritou (2012); Ruszczynski (1995). Another possible strategy for reducing the duality gaps would be to integrate the Lagrangian relaxation (3.19) in a branch-and-bound enumeration scheme (i.e., a branch-and-price framework).

The computational challenges experienced with improving the lower bound is related to two aspects of the problem formulation. The first is the strength of the lower bound provided by the Lagrangian relaxation, which may depend strongly on the chosen relaxation (Guignard, 2003). If the variables in the linking constraints also appear in several of the other constraints in the subproblems, then applying Lagrangian *decomposition* (Guignard and Kim, 1987) is generally preferable, in the sense that the lower bound in Lagrangian decomposition is as least as tight as any other Lagrangian relaxation. The complexity, on the other hand, in solving the scheduling problem for the shale-gas multi-pad system is inherent, as the problem involves both dynamics, combinatorial decisions and large-scale systems. The scheduling is based on switching wells on-off to track a reference rate. As we assume that each well has a separate independent closed-loop control of the wellhead pressure, meeting the reference rate *exactly* will almost certainly be impossible.

There are several possible extensions and modifications of the proposed Lagrangian relaxation scheme to reduce the computation time of the Lagrangian (3.19). One is to consider decomposition per wells, using a type of bi-level decomposition scheme in which the current Lagrangian relaxation is retained while the subproblems (3.20) are further decomposed into one problem per well. However, the reformulation (3.18c)–(3.18d) of the infinity norm can in this case no longer be applied, requiring either a special purpose algorithm for ℓ_1 or ℓ_∞-type problems or a different objective function. Temporal decomposition of the primal MILP (3.18) may also be considered, however, leading to an enormous increase in the number of dual variables. Finally, note that a simple acceleration scheme for the given Lagrangian relaxation scheme is parallelization of the subproblems using grid-computing.

The proposed Lagrangian relaxation scheme is easily extended to handle a time-varying reference rate q_k^{REF}. This is illustrated in Fig. 3.12 for a system with three pads and a three week planning horizon. The Lagrangian relaxation scheme becomes computationally more demanding with a time-varying reference rate, and gives K_c+1 dual variables where $K_c < K$ is the number of changes in the reference rate q_k^{REF}. If the number of dual variables is close to the length of the planning horizon, using a more sophisticated stabilization of the cutting-plane algorithm than the trust-region approach may reduce the number of iterations n and hence

improve the performance of the scheme, see (Hiriart-Urruty and Lemarechal, 1993, Ch. XV).

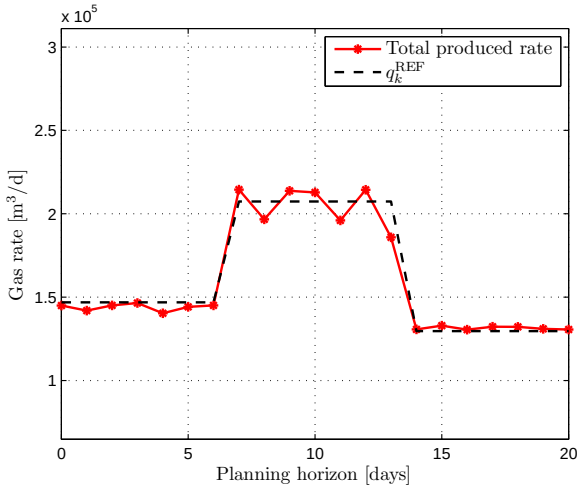


Fig. 3.12: Example on time-varying given reference rate q_k^{REF}

The paper also includes a systematic, frequency selective tuning procedure for the shale well proxy model, together with a simple scheme for updating the parameters in the model. By properly designing the filter for the prediction errors, it is possible to select which of the dynamics in shale-gas wells the proxy model should emphasize. The simplicity of the proxy model still limits to some extent the range of dynamics the model can cover. The tuning procedure further consists of solving a nonconvex NLP, possibly leading to suboptimal parameters in the proxy model. This can be compensated for by frequently updating the model parameters using the updating scheme described in Section 3.3.1.

Scheduling shut-ins of shale-gas systems relative to a reference rate rather than maximizing the production may be beneficial both from an operational as well as an economic and energy perspective. Operationally it leads to more stable rates as seen in Fig. 3.10 and 3.11, with the potential of significantly reducing high and low peaks compared to a standard approach with fixed shut-in times (Knudsen and Foss, 2013). The economic incentive is related to the current low sales prices of natural gas, a result of increased availability mainly due to overproduction of shale gas. Gas prices may vary substantially with seasonal demands, and produced natural gas may be stored and sold at times with higher prices. As mentioned, the loss in cumulative production by performing short shut-ins may be kept very low for shale-gas wells due to the fracture networks acting as a recharge and storage system. Consequently, by optimizing well shut-ins with respect to varying demands as illustrated in Fig. 3.12, the wells may be scheduled to account for seasonal variations in demands and gas prices. The high availability of shale-gas, together with the shut-in characteristics and properties described and optimized in this paper, makes it possible to utilize shale-gas systems as a buffer of natural gas in an energy and power planning context.

3.8 Conclusions

In this work, a novel Lagrangian relaxation based scheme for shut-in scheduling in shale-gas multi-pad systems has been presented. The shut-in scheduling is integrated in field-wide production planning based on demands in gas-rate, showing how production objectives and constraints at different levels in the large-scale system can be included in the formulation. The decomposable scheme is scalable in the number of pads, and shown through computational testing to outperform a fullspace approach for all sizable problems, although the efficiency of the scheme is observed to decrease for long planning horizons. Hence, the proposed method is indeed interesting for shale-gas systems where the number of wells run into the hundreds and beyond.

3.A Proof of bound for dual variable

For proving the bound $\lambda \geq -1$, we can consider the following simplified Lagrangian relaxation subproblem (3.20):

$$\min\ f + \lambda q^{\text{ref}} \tag{3.32a}$$

s.t.

$$f \geq q^{\text{ref}} - q^{\text{pad}}, \tag{3.32b}$$

$$f \geq -q^{\text{ref}} + q^{\text{pad}}, \tag{3.32c}$$

$$q^{\text{Pad}} \leq \bar{U}, \tag{3.32d}$$

$$f, q^{\text{ref}}, q^{\text{Pad}} \geq 0, \tag{3.32e}$$

where we for simplicity have omitted the indices k and l. The variables f, q^{ref} are block separable from the remaining variables defining the feasible set of (3.20), while the total rate q^{Pad} is defined by a polytope $\bar{\mathcal{Q}}$ spanned by (3.18e)–(2.31n) for given $l \in \mathcal{L}$, i.e. a bounded polyhedron (Nemhauser and Wolsey, 1988). Consequently, we can deduce an upper bound $\bar{U}$ on q^{Pad}. Consider the general LP, $Z_{\text{LP}} = \min_x\{c^T x : Ax \geq b,\ x \geq 0\}$, and the corresponding symmetric dual LP, $Z_{\text{LPDual}} = \max_u\{b^T u : A^T u \leq c,\ u \geq 0\}$. By weak duality, if the primal LP is unbounded then the dual LP is infeasible (Nemhauser and Wolsey, 1988). Hence, by considering the feasible set $A^T u \leq c,\ u \geq 0$ for the symmetric dual LP of (3.32) parametrized by λ, we obtain the linear inequalities,

$$u_1 + u_2 \leq 1, \tag{3.33a}$$

$$-u_1 + u_2 \leq \lambda, \tag{3.33b}$$

$$u_1 - u_2 - u_3 \leq 0, \tag{3.33c}$$

$$-u_1 \leq 0, \tag{3.33d}$$

$$-u_2 \leq 0, \tag{3.33e}$$

$$-u_3 \leq 0. \tag{3.33f}$$

Feasibility requirements of the above inequalities can be derived by Fourier-Motzkin elimination, see for instance Martin (1998). By successively eliminating u_1 and u_2

by pairing the inequalities (3.33), we obtain the inequalities

$$-u_3 \leq \frac{1}{2}(1+\lambda), \tag{3.34a}$$
$$0 \leq \frac{1}{2}(1+\lambda), \tag{3.34b}$$
$$-u_3 \leq 1, \tag{3.34c}$$
$$0 \leq 1, \tag{3.34d}$$
$$-u_3 \leq 0. \tag{3.34e}$$

The bound $\lambda \geq -1$ then immediately follows from (3.34b).

Chapter 4

Shale-gas Scheduling for Natural-gas Supply in Electric Power Production

This chapter consists of the paper Knudsen et al. (2014c):

Knudsen, B. R., Whitson, C. H., and Foss, B. (2014c). Shale-gas scheduling for natural-gas supply in electric power production. *Energy* (accepted).

Abstract

This paper describes a novel integration of shale-gas supply in geographical proximity to natural-gas power production. Shale-gas reservoirs hold special properties that make them particularly suited for intermittent shut-in based production schemes. The proposed scheme argues that shale-gas reservoirs can be used to shift storage of gas used for meeting varying demands, from separate underground storage units operated by local distribution companies to the gas producers themselves. Based on this property, we present an economical attractive option for generating companies to increase their use of firm gas-supply contracts to the natural-gas power plants in order to secure a sufficient gas supply. The shale-well scheduling is formulated as profit-maximization model for well operators, in which we seek to include their main operational challenges, while preserving an economic incentive for the operators to adopt the proposed scheme. The resulting large-scale mixed integer linear program is solved by a Lagrangian relaxation scheme, with a receding horizon strategy implemented to handle operational uncertainties. We present the proposed optimization framework by illustrative case studies. The numerical results show a significant economic potential for the shale-well operators, and a viable approach for generating companies to secure a firm gas supply for meeting varying seasonal electricity demands.

4.1 Introduction

The use of natural gas for energy production in the US has increased significantly over the last decade. Natural gas constituted 30% of the total electricity genera-

tion in the US in 2012 EIA (2014), from which the use of natural-gas power plants (NGPPs) has increased by 35% from 2005 to 2012. One of the main factors for this increase has been the access to large volumes of "new" gas recovered from unconventional resources such as shale (ultra-tight) and other tight-gas formations (EIA, 2013). About 45 GW of natural gas-fired generation capacity is expected to be developed over the next ten years (NERC, 2011), while the US Energy Information Administration (EIA) projects the share of natural gas in total electricity generation to reach 35% by 2040 (EIA, 2014). This continued increase in use of natural-gas for electric power generation in the US is both related to the displacement of coal-fired power plants to meet CO_2 reduction targets (Macmillan et al., 2013), and a sustainable future supply from giant shale-gas recoverable reserves ($2.1 \times 10^{10}\text{m}^3$ (EIA, 2013)).

The remarkable speed and scale of the US shale-gas developments has reduced the country's dependency on gas import, and thereby improved its security of supply. This development has however resulted in an abundance of natural gas in the domestic US market (MIT Energy Initiative, 2011; EIA, 2013), and consequently caused a substantial decrease in natural-gas price. The combination of low fuel prices, high efficiencies of modern combined-cycle power plants (Kehlhofer, 2009) and significantly lower emission levels compared to coal-fired power plants (MIT Energy Initiative, 2011) has favored natural gas for electricity generation, both for serving baseloads and for satisfying demands during peak periods (Chen and Baldick, 2007). The gas demand in the electric power sector is very sensitive to the relative price difference between gas and coal; after a consistent increase until 2012, the total gas consumption by the electric power sector was reduced in 2013 compared to 2012, mainly caused by an increase in gas price (EIA, 2014). The dependency of power-plant gas usage on gas price is mutual for producers and consumers: While continued low gas price is essential for making natural gas a preferred fuel in electric power generation, it has also dramatically reduced the profitability of many dry-gas fields (Hughes, 2013; Kaiser, 2012; MIT Energy Initiative, 2011). Shale-gas operators' interest in dry-gas fields have been on a decline, causing a significant reduction in the rig-count for many dry-gas fields (Liu et al., 2013), and a shift in focus to condensate-rich shale-gas fields. Wells that do not return the required capital expenditure within the first two-three years are often considered unprofitable and thus abandoned. In some situations, operators also choose to shut-in wells to wait for higher gas prices (Helman, 2012; Reuters, 2012), or to postpone the start-up of completed wells (Redden, 2012) as the initial peak rate yields a major and decisive part of the profit of shale-gas wells. As such, there is a mutual interest both from the US shale-gas industry and from the electric power sector to explore ways of better utilizing the abundance of dry natural-gas caused by the shale-gas revolution.

Natural gas is normally contracted and sold directly from producers to local distribution companies (LDCs) as illustrated in Fig. 4.1, paying a transportation fee to a shipper or a gas pipeline company (Avery et al., 1992). The LDC distributes the gas to end-users in the different sectors, including electric power generation, residential and commercial heating, transportation and a variety of other industrial customers. Natural gas is traded between producers and LDCs, which may be located in the proximity of each other or distanced far apart; in the latter case

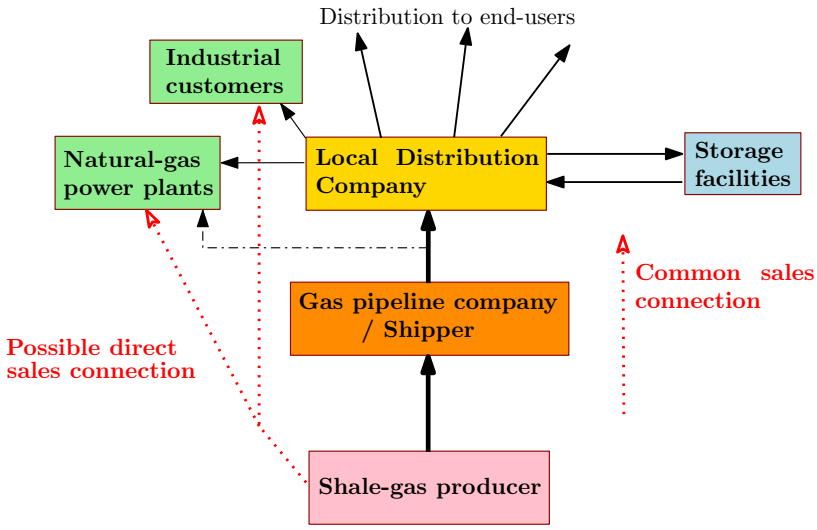


Fig. 4.1: Participants in the value chain between gas producers and natural-gas power plants.

increasing the compression cost for the pipeline company and hence the transportation fee. The seasonal demand for gas in the aforementioned sectors varies and depends on factors such as the price for gas compared to alternative fuels, temperature variations, and economy in the industrial and residential sectors (FERC, 2012; Guldmann and Wang, 1999). However, comparing in Fig. 4.2 the year-to-year seasonal demands, the different sectors exhibit a relatively predictable trend in natural-gas demand. As an example, NGPPs tend to consume more natural gas during the summer months to meet air conditioning loads (FERC, 2012). To facilitate these varying gas demands, the LDCs exploit extensively underground storage using salt caverns and depleted reservoirs. The end users pay additional fees to the LDCs for this storage service, a fee which includes: transportation to and from the storage facility, injection well compression, withdrawal and capacity charges (Avery et al., 1992; FERC, 2012), and losses of injected gas.

Power plants and industrial customers may contract their gas supply directly with the producer to save costs, in particular the storage fees paid to the LDCs (Avery et al., 1992; FERC, 2012; Liu et al., 2009). The entry of shale-gas in the US natural-gas value chain both changes and increases the possibility for direct contracting and distribution of gas between producers and large end-users. As shale-gas is a land-based resource and often located much closer to end-users in the industrial and electric power sector than conventional natural-gas resources and LNG terminals, it may reduce the distance gas needs to be transported. This lowers fuel consumption by the midstream gas compressors, eventually leading to lower transportation fees for the end-users. A further special property of shale and other fractured tight-formation gas-wells is that they can be shut-in for short periods with a minimal loss of recovery (Whitson et al., 2012; Knudsen and Foss, 2013). This property implies that shale-gas reservoirs may essentially be utilized for storing natural gas, with the same purpose as LDCs make use of separate storage

facilities, to meet varying prices and demands. Using shale-gas wells directly for scheduling natural-gas production and supply with respect to varying seasonal demands means that the storage in Fig. 4.1 is moved from the LDC connection back to the origin of the producer. Large shale-gas fields, consisting of many wells, are as such particularly well suited for supplying large natural-gas end-users as NGPPs. We elaborate on this in Section 4.2.1.

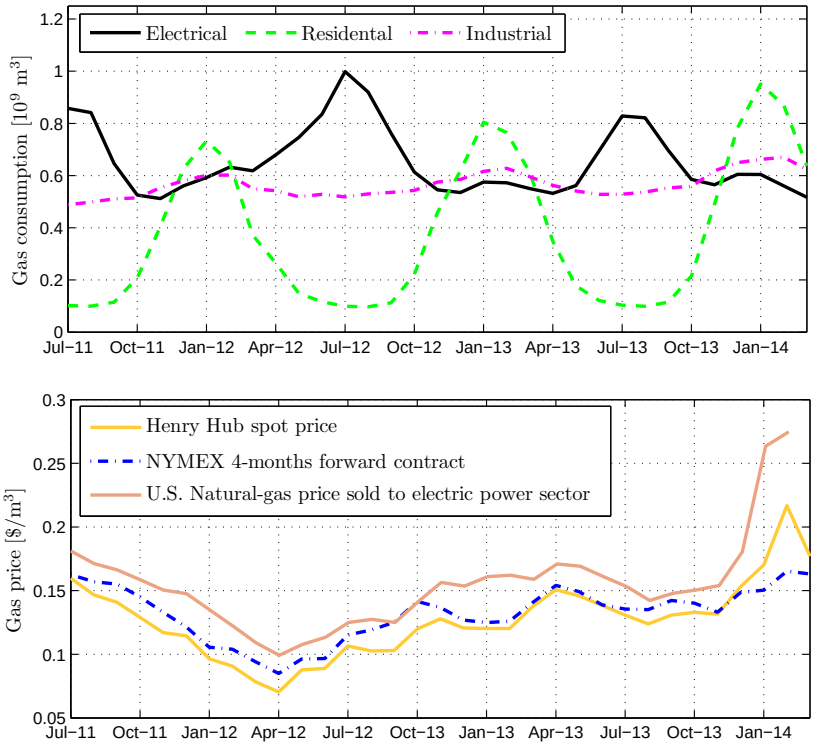


Fig. 4.2: Variations in gas price and seasonal demands in the US (EIA, 2013b,a).

Electric utility companies (EUCs) or generating companies (GENCOs) that own and operates one or more NGPPs need to ensure contract portfolios with low-cost fuel supplies to be competitive in the electricity generation market. Contracts are generally distinguished as *firm* or *interruptible* (Avery et al., 1992; FERC, 2012). EUCs and GENCOs have traditionally preferred to use interruptible contracts with gas suppliers Liu et al. (2009); ISO New England (2008), providing the power plant with inexpensive gas which can be interrupted by both parties at short notice. The preference of interruptible contracts is also related to predominant use of natural gas for satisfying peak demands Chen and Baldick (2007). In contrast, a *firm* contract entitles the customer with a high priority in which the gas is to be supplied with no interruptions. NGPPs holding only interruptible contracts will thus be susceptible to gas curtailment during periods of peak demand, pipeline congestion, or in the event of a gas-producer reaching its maximum recovery rates (IFC International, 2013). In contrast, customers holding firm supply contracts,

together with customers in the non-electrical sectors will retain the highest priority of the available gas supply (Munoz et al., 2003; Guldmann and Wang, 1999). Note that both firm and interruptible contracts are also applied for the transportation service provided by the shipper (Avery et al., 1992; FERC, 2012). The increased use of natural-gas in electricity generation leads to challenges and new requirements for securing the gas supply to power plants (MIT Energy Initiative, 2011; ISO New England, 2008). Several places in the US, the infrastructure of the natural-gas value chain has not yet been developed to support the high volumes of natural gas produced from the vast shale resources (NERC, 2011), causing low excess capacity in the transmission and distribution pipelines and hence creating a higher risk of interruptions in the supply to customers holding non-firm contracts. This poses a significant challenge in the context of replacing coal-fired power plants with NGPPs as coal, in contrast to natural gas, can be stored on-site at power plants. Moreover, this is an important element for overall system reliability with the increased use of intermittent renewable energy sources (MIT Energy Initiative, 2011).

The reliability and security of natural-gas supplies to NGPPs may be improved by increasing the use of *firm* contracts (ISO New England, 2008), both for the transportation service, and for the supply of gas from the producer to the LDCs or directly to end-users. This will further increase the reliability of natural-gas based electricity generation. Increasing the share of firm contracts for gas supply to NGPPs is, however, clearly a cost-benefit question as well as a security issue, and is related to the security-constrained thermal unit commitment (UC) problem (Liu et al., 2009; Li et al., 2008; Fu et al., 2005; Partovi et al., 2011), as well as optimization of the natural-gas supply mix for EUCs/GENCOs and LDCs (Chen and Baldick, 2007; Guldmann and Wang, 1999; Avery et al., 1992). To render an attractive opportunity for increasing the use of direct firm supply and transportation contracts between shale-gas producers and NGPPs, it is necessary to create an economic incentive for shale-gas operators to increase their requisite scheduling efforts to meet the firm demand rates. Furthermore, pricing of the firm contracts has to be at a level below the average firm contract option provided by LDCs.

In this paper, we propose a novel scheme for scheduling shale-gas wells that may enable NGPPs to establish firm contracts directly with shale-gas producers, thereby circumventing LDC firm contracts. This is achieved by exploiting shale-gas wells and their near well region as a storage facility by scheduling well shut-ins according to a carefully designed optimization procedure. We will consider daily scheduling of wells in large shale-gas fields to account for varying seasonal demands as shown in Fig. 4.2. The scheduling problem is formulated as a large-scale PDE-constrained generalized disjunctive program (GDP) (Raman and Grossmann, 1994), in which each reservoir and well unit is efficiently modeled using a reduced-order proxy model. Following a reformulation of the GDP to a mixed integer linear program (MILP), we solve the scheduling problem using Lagrangian relaxation, applying a proximal bundle method (Kiwiel, 1990) to solve the Lagrangian dual, while we construct a novel binary-fixing heuristic for recovering primal feasibility from the solution of the Lagrangian. To accommodate uncertainties in well operations, spot price forecasts and prediction of well rates, we implement a receding horizon control strategy (Rawlings and Mayne, 2009) in which the well schedule is reoptimized daily during the planning horizon.

The reminder of the paper is organized as follows. Section 4.2 contains a review of the well and reservoir proxy models, with a subsequent analysis of the special shut-in and gas storage possibility of shale-gas wells. In Section 4.3 we present the shale-well system, and the operational constraints that need to be included in the optimization problem for scheduling a firm supply to the NGPP. Section 4.4 describes the Lagrangian relaxation based decomposition scheme, while Section 4.5 describes the use of receding horizon optimization to robustly meet the demand rates. In Section 4.6 we demonstrate the efficiency of our proposed scheme using several case studies, with a subsequent discussion of the results in Section 4.7. Concluding remarks end the paper in Section 4.8.

4.2 Shale well modeling

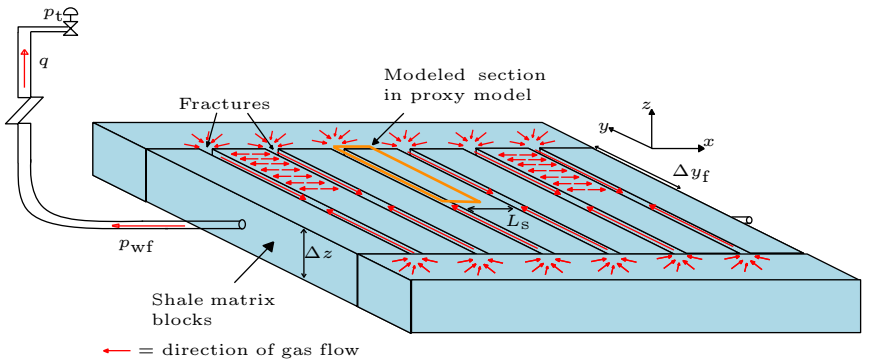


Fig. 4.3: Illustration of reservoir and proxy model.

Hydraulically fractured shale and tight-formation gas reservoirs consist of tight, low permeable rock matrix blocks intersected by highly conductive fractures as illustrated in Fig. 4.3. These systems are mainly modeled either as a so-called dual-porosity system, see e.g. Bello and Wattenbarger (2010); Nobakht et al. (2012), or as fully PDE-discretized single-porosity dual-permeability models (Cipolla et al., 2010). The former, idealized modeling scheme is widely used to derive static long-term production forecasting tools and assumes steady-state operations, while the latter scheme leads to complex, numerically demanding models that take into account transient effects dictated by the differential flow equations, transients that are important to the modeling of short-term shut-in periods as promoted in this paper. A reduced-order shale well and reservoir proxy model was derived in Knudsen et al. (2012); Knudsen and Foss (2013), using first-principal physics of the storage and transport mechanisms in the well and the reservoir. The proxy model was developed for efficiently optimizing short cyclic well shut-ins to prevent co-produced liquids to accumulate in the wellbore, and shown through a tuning scheme to give a good transient fit. Knudsen et al. (2014b) developed a similar but slightly modified shale-well proxy model based on the Cartesian geometry shown in Fig. 4.3. By extending the static pressure model of the wellbore with a quadratic friction term (Katz and Lee, 1990) and using a similar formulation of the frequency-selective

tuning scheme described in Knudsen et al. (2014a), the modified proxy model was shown to further improve the transient fit and in particular the steady-state fit compared to the model used in Knudsen and Foss (2013).

Consider again the idealized shale-well geometry shown in Fig. 4.3. By assuming equal spacing and distribution of the fractures, and that the fractures have infinite conductivity, i.e. with no pressure loss, and penetrate the entire organic-rich formation, a reduced-order proxy model can be constructed by only considering a quarter section of the matrix-fracture system as illustrated by the orange frame in Fig. 4.3 (Knudsen et al., 2014b). The dominating direction of the gas-flow in the shale-matrix is orthogonal to the fractures (Bello and Wattenbarger, 2010; Nobakht et al., 2012), that is, in the x-direction shown in Fig. 4.3. The proxy is hence constructed as a one-dimensional flow model, using a single layer, a spatially dependent permeability $k(x)$ and an integral transformation from pressure p to *pseudopressure* $m(p)$ (Al-Hussainy et al., 1966),

$$m(p) := 2\int_{p_\mathrm{b}}^{p} \frac{p'}{\mu(p')Z(p')}dp', \tag{4.1}$$

where $\mu(p)$ is the gas viscosity, $Z(p)$ is the gas compressibility factor and p_b is a low base pressure, rendering the semi-linear initial-boundary value problem (IBVP) (Knudsen et al., 2012, 2014b)

$$\phi\mu c\frac{\partial m}{\partial t} = \frac{\partial}{\partial x}\left(k(x)\frac{\partial m}{\partial x}\right), \tag{4.2a}$$

$$\left.\frac{\partial m}{\partial x}\right|_0 = q\frac{2Tp_\mathrm{sc}}{T_\mathrm{sc}\Delta z\Delta y_\mathrm{f}k_\mathrm{f}}, \tag{4.2b}$$

$$\left.\frac{\partial m}{\partial x}\right|_{\frac{L_\mathrm{s}}{2}} = 0, \tag{4.2c}$$

$$m(x,0) = m^\mathrm{init}. \tag{4.2d}$$

In (4.2), ϕ is porosity, $c(p)$ is total compressibility, q is the gas rate at standard conditions, 1 bar and 15.6°C, T is temperature and p is pressure. Spatial references are shown in Fig. 4.3. An I-dimensional spatial discretization of (2.6) is constructed using central difference approximations, while time discretization applies the backward Euler approximation using a timestep Δk. This leads to the discretized reservoir proxy model

$$Am_{k+1} = m_k + Bq_{k+1}, \tag{4.3a}$$

$$m_0 = m^\mathrm{init}, \tag{4.3b}$$

where k is the discrete time index. The gas rate flowing from the reservoir into the wellbore is given by

$$q_k = \beta\left(m_{k1} - m_{\mathrm{wf},k}\right), \tag{4.4}$$

where m_{k1} is the pseudopressure in the gridblock adjacent to the fracture, m_wf is the bottomhole pseudopressure and β is a constant. For a given tubinghead pressure

p_{t}, the gas rate the well can deliver in a timestep k is found by the intersection of (4.4) and the static tubing-model (Katz and Lee, 1990),

$$p_{\mathrm{t},k}^2 = \frac{1}{C_{\mathrm{t}}^2} q_k^2 + e^{-S} p_{\mathrm{wf},k}^2, \tag{4.5a}$$

$$m_{\mathrm{wf},k} := \tilde{a}_1 p_{\mathrm{wf},k}^2 + \tilde{a}_2, \tag{4.5b}$$

where $(\tilde{a}_1, \tilde{a}_2)$ in (4.5b) are regression parameters used to convert the square of bottomhole pressure p_{wf} to m_{wf} (Knudsen et al., 2014b). The first term on the right-hand side of (4.5a) corresponds to the pressuredrop caused by the tubing friction, while the term e^S yields the hydrostatic head of the gas column. C_{t} is a tubing specific constant, see Katz and Lee (1990). Note that p_{t} serves as the boundary condition of the aggregated single well and reservoir proxy model.

4.2.1 Shale reservoirs as gas storage

Shale-gas reservoirs have as mentioned in Section 4.1 the special property that production may be shut in for short periods without losing significant long-term recovery. That is, the magnitude of the subsequent peak rate after a shut-in largely recovers for the temporary loss in production during shut-in, hence avoiding any significant reduction in net present value (NPV). This property is in contrast with conventional high-permeability gas reservoirs. For these reservoirs, considering a fixed life-span of the well, the temporary loss in production during a shut-in would not be recovered before *after* the end of the prediction horizon, thus reducing the NPV of the well. Applying shut-ins of shale-gas wells was initially developed and studied by Whitson et al. (2012) as a cyclic production scheme for eliminating so-called liquid loading in gas-wells (Turner et al., 1969). The basis for the shut-in property of shale-gas wells is the high pressure gradients between the low-permeable shale matrix and the interconnected network of fractures, causing the shale matrix to act merely as a source term feeding the fractures with gas. During shut-ins, these pressure gradients will cause the gas to continue to flow into the fractures, recharging the fractures with gas and thereby increasing the near-wellbore pressure. Once the well is reopened, the gas recharged in the fractures will cause a high peak in the gas rate (Knudsen and Foss, 2013). The volume of the gas recovered after the shut-in compared to if the well is continuously producing, depends on the length of the shut-in and the formation permeability, the latter being the main parameter controlling the ability of the gas to flow through the tight formation (Cipolla et al., 2010).

In Fig. 4.4, we demonstrate the effects of applying intermittent shut-ins on shale-gas wells by simulating a high-fidelity shale reservoir model for different shut-in lengths and different formation permeabilities. The figure shows the total recovery after a fixed 10 year simulation horizon, normalized against the cumulative rate obtained if the well is continuously produced, i.e. with no shut-ins, displayed as a function of the formation permeability and the % of the 10 years operation time the well is shut-in. For each simulated formation permeability, k_{m}, 0% shut-in time hence corresponds to continuous production, giving a normalized recovery of 1 as shown in the upper left corner of Fig. 4.4. Each time the gas rate hits the

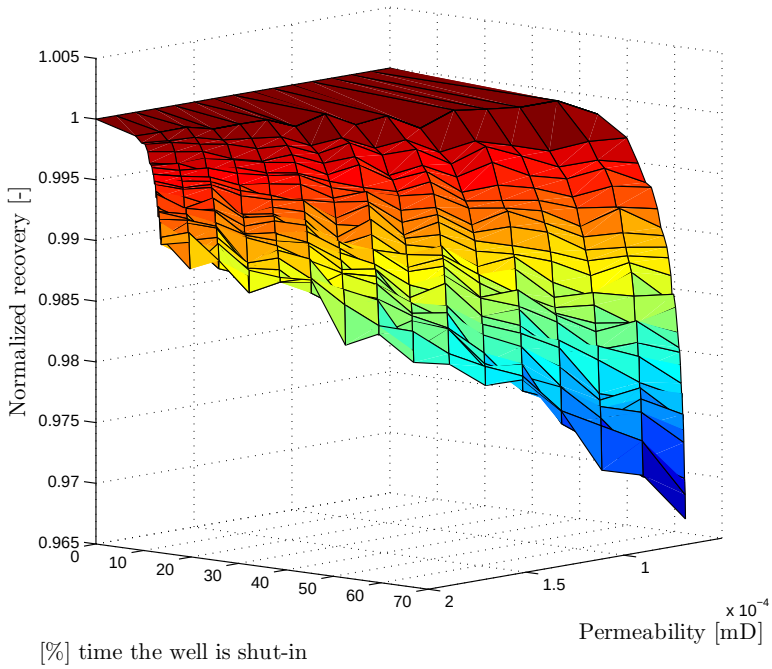


Fig. 4.4: Normalized relative recoveries as a function of the formation permeability k_{m} and the % time of total operation time the well is shut-in. Each total recovery rate is withdrawn after running the simulation with a fixed 10 year horizon, and normalized against the maximum recovery within this horizon obtained by applying no shut-ins.

lower (critical) rate, q_{gc}, a rate needed to ensure continuous removal of co-produced liquids Turner et al. (1969); Coleman et al. (1991); Lea and Nickens (2004), we shut in the well for a predefined, minimum time, and reopen the well once the pressure is sufficiently built-up to regain a flowrate above q_{gc}. Observe that a relative recovery of 1 is as such only a theoretical value, as co-produced liquids and possible liquid loading in practice always will require shut-ins throughout the life of a well. The simulations are performed using a finely gridded realization of the model shown in Fig. 4.3 implemented in the state-of-the-art reservoir simulation software SENSOR (SENSOR, 2011), assuming 10 equally spaced fractures, 200 bar initial pressure and using the correlation in Coleman et al. (1991) for computing q_{gc} .

For this given shale-well realization, it is clear from Fig. 4.4 that low formation permeabilities permit the well to be shut-in a high percentage of the total operation time while still recovering close to 100% of the maximum recovery obtained by producing the well continuously. Low permeabilities cause a long shut-in time compared to the length of the subsequent production period, giving more than 50 % total shut-in time for the lower range of the simulated formation permeabilities. Note that a lower value of the given shut-in rate q_{gc} would as such reduce the total % time the well is shut-in of the fixed 10 year time horizon. Notwithstanding, it is evident that fractured shale-gas wells with very low permeabilities are particularly suitable to intermittent shut-in schemes. If the shut-ins are scheduled optimally

with respect to varying demands and prices, this property may then translate into increased profit for the operators. This utilization of shale-gas reservoirs is equivalent to a type of gas storage, since very little recovery is lost. It is important to note that above a certain percentage total shut-in time the recovery suddenly drops, eventually leading to loss in NPV and as such reducing the economic viability of such shut-in schemes.

4.3 Formulation of the shale-well scheduling problem

To create an incentive for shale-gas operators to schedule the well-production to meet seasonal varying demands from NGPPs, the alternative production scheme must be economically attractive and feasible with respect to daily field operations. Shale-gas fields normally consist of a large number of geographically distributed wells, and may hence, although land-based and relatively easily accessible, be operationally challenging. The formulation of the shale-well scheduling problem to meet firm demand rates must therefore integrate both the main operational constraints and a compact description of the surface-system. In the following section, we present our multi-well, integrated shale-well scheduling problem.

Fig. 4.5 shows the topology of a surface gathering system consisting of distributed shale-gas wells and a compression unit connected to a transmission pipeline. Each well $j \in \mathcal{J}$ is equipped with its own wellhead choke, and we assume we can remotely control the associated tubinghead pressure p_{t}, cf. Fig. 4.3. Quite commonly, each well is connected to the gathering system with two flowlines, one connected upstream of the compressor, and one bypassing the compressor. The well flows are routed to one of the flowlines. The compression unit is either operated by a midstream company or by the well operator, in both cases leading to a compression cost compared to if the flow is bypassed the compressors, but also increased well deliverability as the well can be operated at a lower pressure p_{t}. Wells are typically produced initially to the high-pressure line bypassing the compressor, and then routed to the low-pressure line when the rate in a well drops. Some wells may be installed with one flowline only, as indicated for some of the wells in Fig. 4.5. Any co-produced liquids are separated from the gas by small shared or individual separation tanks at the wellhead. This separation will only impact the system by slightly lowering the gas pressure, and is hence omitted in the system description.

Natural gas producers will always seek to operate gas-fields in a way that maximizes their total profit. Marketing groups with shale-well operators have the choice of selling gas in both the spot market and in the forward market, to LDCs or to large customers both in the electric and non-electric sectors. Consequently, to maximize profit from season to season with varying demands, cf. Fig. 4.2, the operators may use a sales portfolio with some gas sold as firm supply, i.e. in the forward market, and some sold in the spot market. Selling the gas through a firm contract with a generating company will secure the operator with sales at a certain price, while it also commits the operator to deliver the gas to avoid penalties or the need to buy gas from other producers to meet the contractual obligations, hampering the operator from profiting from a sudden upswing in the gas spot price. Given a firm natural-gas demand d_k^{NGPP} from an NGPP, we hence enable the shale-gas producer

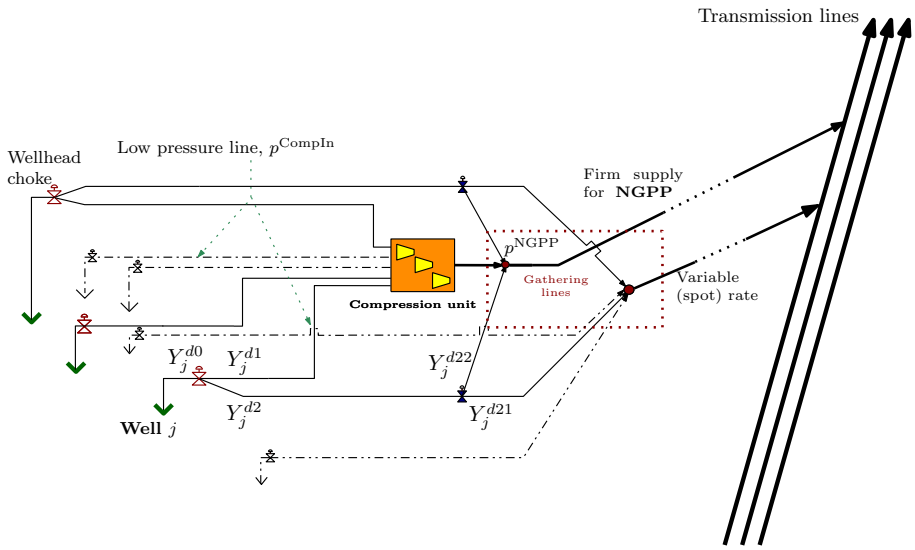


Fig. 4.5: System description of a shale-gas production facility

both to commit supply to the NGPP while selling any excess gas produced in a spot market. The objective function then reads

Table 4.1: Indices and set definitions

Index	Interpretation	Set	Elements
i	spatial reservoir grid block	$\mathcal{I}$	$\{1 \dots I\}$
j	well number	$\mathcal{J}$	$\{1 \dots J\}$
k	discrete time index	$\mathcal{K}$	$\{1 \dots K\}$
d	terms in the disjunctions	$\mathcal{D}$	$\{d0, d1, d2, d11, d21, d22\}$
n	iteration in Lagrangian scheme	-	$\{1 \dots N\}$
u	cuts in the cutting-plane model	-	$\{1 \dots U\}$

$$\max\ C^{\text{oc}} \sum_{j \in \mathcal{J}} \sum_{k \in \mathcal{K}} \left(R_{jk}\ -\ \sigma^{\text{u}}\, \delta q_k^{\text{u}}\right) \Delta k, \tag{4.6}$$

with the firm supply to the NGPP given by the constraint,

$$q_k^{\text{Comp}} + q_k^{\text{HighpresNGPP}} + \delta q_k^{\text{u}} = d_k^{\text{NGPP}}, \qquad \forall k \in \mathcal{K}, \tag{4.7}$$

$$\delta q_k^{\text{u}} \geq 0, \qquad \forall k \in \mathcal{K}. \tag{4.8}$$

In the objective (4.6), R_{jk} is the per well revenue in US\$/day, depending on the gas sales price, the line scheduling and associated operational costs, while $C^{\text{oc}} = (1 - RO)(1 - T)$ are costs in terms of a percentage royalty rate RO and tax T. Moreover, σ^{u} is a penalty in US\$/m^3 for each unit of underproduction δq_k^{u} with respect to the gas-rate demand d_k^{NGPP} given in m^3/day. In (4.7), q_k^{Comp} is the total compressed rate per timestep, i.e. the total outlet rate from the compressor,

while $q_k^{\text{HighpresNGPP}}$ is the sum of the rate from well-flows that are bypassed the compressor and used to meet the firm demand d_k^{NGPP} as illustrated in Fig. 4.5. We assume by this formulation that the primary goal for the producer is to meet the firm NGPP demand d_k^{NGPP}, in the sense that all the gas scheduled through the compressor is used to meet this demand. We regard lease operating costs as capital expenditures and therefore leave this term out of C^{oc}, while manpower costs often are quite small for dry-gas wells (MIT Energy Initiative, 2011, App. D) and therefore also omitted in (4.6).

The compression unit has an upper load capacity $q_{\text{tot}}^{\text{up}}$ limiting how much gas it can compress,

$$q_k^{\text{Comp}} \leq q_{\text{tot}}^{\text{up}}, \qquad \forall k \in \mathcal{K}. \tag{4.9}$$

New wells are normally frequently drilled and added to the shared surface gathering and compression system. Consequently, when the number of wells grows, the total production may eventually exceed this upper load capacity of the compression unit, requiring wells to be scheduled to the high pressure line bypassing the compressor.

In order to efficiently model the disjunctive well flows shown in Fig. 4.5, we introduce Boolean variables Y_{jk} for each routing decision and use generalized disjunctive programming (GDP) (Raman and Grossmann, 1994; Grossmann and Trespalacios, 2013) to formulate the scheduling problem. Using disjunctions to model the routing and the shut-ins of the wells increases the level of flexibility in the model formulation by rendering a formulation that more directly captures the connections between the logical part and the constraints of the problem compared to an ad-hoc mixed-integer formulation (Grossmann and Trespalacios, 2013). Consider in Fig. 4.5 the lower-most well with the indicated Boolean variables Y^d. At the first valve, there are three Booleans Y^{d0}, Y^{d1} and Y^{d2}, modeling the logical decisions corresponding to the well being shut-in, the well being routed upstream of the compression unit and the well being bypassed the compressor, respectively. If the well-flow is routed directly downstream of the compressor, a subsequent valve routes the flow either to be sold in the spot market or to supplement the compressed rate to meet the firm demand d_k^{NGPP}. This decision is modeled by the Booleans Y^{d21}, Y^{d22}, where Y^{d22} as shown in Fig. 4.5 corresponds to the well flow being routed to supply the NGPP, while $Y^{d21} = \text{True}$ means that the gas is routed and sold in the spot market. Note that the Booleans in both the first and second routing valves are exclusive.

The gas must be delivered to the NGPP above a minimum pressure. A contract of firm gas supply between a shale-gas producer and an NGPP also requires contracting firm transportation service with the shipper (NERC, 2011), to ensure that a certain pipeline capacity is allocated for the gas exchange. The transportation contract with the shipper however also commits the shipper to deliver the gas to the end-user at certain specifications (Tomasgard et al., 2007). The shipper will hence require the gas from the shale-gas producer to be delivered at a minimum pressure p^{NGPP}, based on the pressure required by the NGPP, while it is the shipper that ensures finally delivery of the gas at this pressure. Downstream of the compression unit the gas pressure will therefore have to be at or above p^{NGPP}. Since any pressure loss in the shale-gas gathering system is small and hence negligible, p^{NGPP} will be the lower bound on the tubinghead pressure p_{t} if the well flow

is bypassed the compressor. If, on the other hand, the well flow is routed through the compression unit, p_{t} will be bounded below by a minimum compressor inlet pressure p^{CompIn}. The wells will typically operate at a pressure close to p^{CompIn}. Consequently, with small inlet pressure variations and the assumption of constant downstream pressure, the fuel consumption will be close to constant (Guo et al., 2007), and is hence modeled as a simple fixed-cost term C^{C}. Finally, each well must produce at a rate above the aforementioned critical gas-rate q_{gc} in order to prevent so-called liquid loading (Turner et al., 1969; Coleman et al., 1991; Whitson et al., 2012; Knudsen and Foss, 2013), a state of the well causing erratic and unstable rates, and which is one of the main operational concerns for gas-well operators (Lea and Nickens, 2004). The critical rate is correlated with the wellhead pressure Turner et al. (1969), and will hence be higher when the well is scheduled to the high pressure line. Including these operational constraints with the well model (4.3)–(4.5), we formulate the scheduling of each well by the disjunction

$$A_j m_{jk+1} = m_{jk} + B_j q_{jk+1}, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K}^{\mathrm{m}}, \tag{4.10a}$$

$$m_{j0} = m_j^{\mathrm{init}}, \qquad \forall j \in \mathcal{J}, \tag{4.10b}$$

$$p^2_{\mathrm{wf},jk} = e^S \left(p^2_{\mathrm{t},jk} + \frac{1}{C^2_{\mathrm{t},j}} q^2_{jk} \right), \qquad \forall j \in \mathcal{J}, k \in \mathcal{K}, \tag{4.10c}$$

for all $j \in \mathcal{J}, k \in \mathcal{K}$: (4.10d)

$$\begin{bmatrix} Y_{jk}^{d0} \\ q_{jk} = 0 \\ m_{jk1} = \tilde{a}_1 p^2_{\mathrm{wf},jk} + \tilde{a}_2 \\ p_{\mathrm{t},jk} \le p_t^{\mathrm{up}} \\ R_{jk} = 0 \end{bmatrix} \veebar \begin{bmatrix} Y_{jk}^{d1} \\ q_{jk} = \beta \left(m_{jk1} - \tilde{a}_1 p^2_{\mathrm{wf},jk} - \tilde{a}_2 \right) \\ p_{\mathrm{t},jk} \ge p^{\mathrm{CompIn}} \\ q_{jk} \ge q_{\mathrm{gc}}^{\mathrm{low}} \\ R_{jk} = G^{\mathrm{NGPP}} \left(1 - C^{\mathrm{C}}\right) q_{jk} \end{bmatrix} \veebar \begin{bmatrix} Y_{jk}^{d2} \\ q_{jk} = \beta \left(m_{jk1} - \tilde{a}_1 p^2_{\mathrm{wf},jk} - \tilde{a}_2 \right) \\ p_{\mathrm{t},jk} \ge p^{\mathrm{NGPP}} \\ q_{jk} \ge q_{\mathrm{gc}}^{\mathrm{high}} \\ \begin{bmatrix} Y_{jk}^{d21} \\ R_{jk} = \hat{G}_k^{\mathrm{spot}} q_{jk} \end{bmatrix} \vee \begin{bmatrix} Y_{jk}^{d22} \\ R_{jk} = G_k^{\mathrm{NGPP}} q_{jk} \end{bmatrix} \end{bmatrix},$$

$$Y_{jk}^{d2} \Leftrightarrow Y_{jk}^{d21} \veebar Y_{jk}^{d22}, \tag{4.10e}$$

$$\Omega(Y_{jk}^{d0}, Y_{jk}^{d1}, Y_{jk}^{d2}, Y_{jk}^{d12}, Y_{jk}^{d22}) = \mathrm{True}, \tag{4.10f}$$

$$Y_{jk}^{d} \in \{\mathrm{True}, \mathrm{False}\}.$$

In (4.10d), $\vee$ and $\veebar$ is the OR and exclusive OR operator, respectively. The logic proposition (10e) ensures that exactly one of Y^{d21} and Y^{d22} is True if $Y^{d2} =$ True. The symbolic equation (4.10f) comprised by Ω represents possible logical relations between the Boolean variables, i.e. for defining minimum production and shut-in times. p_t^{up} is an upper bound on the allowed tubinghead pressure, while the set $\mathcal{K}^{\mathrm{m}}$ for the implicit time reservoir model (4.10a) is shifted one step of $\mathcal{K}$, i.e. $\mathcal{K}^{\mathrm{m}} = \{0, \ldots, K-1\}$. By scheduling the well gas-flows to supply the NGPP, the operator receives the preset contracted gas price G_k^{NGPP}. If, on the other hand, the well is scheduled to sell the gas in the spot market, the operator only has a certain *estimate* $\hat{G}_k^{\mathrm{spot}}$ of the future gas spot price. This is further addressed in Section 4.5 below. In (4.10), we have substituted the mapping (4.5b) into the inflow (4.4), thereby eliminating m_{wf}.

The nonlinear GDP problem (4.6)–(4.10) can be converted to a mixed integer program by reformulating the linear disjunctions (4.10d) to algebraic constraints,

using either a reduced-size big-M reformulation (Nemhauser and Wolsey, 1988) or the increased-size, tight convex hull description of linear disjunctions (Balas, 1985). The latter convex-hull reformulation is based on disaggregation of variables in the disjunction, introducing binary variables y^d with one-to-one correspondence with the Boolean variables Y^d, and assuring that only one of the terms in the disjunction is active. The particular convex hull reformulation of (4.10d) is shown in Appendix 4.A, yielding a full mixed integer reformulation of (4.6)–(4.10). A direct result of applying the convex hull compared to a big-M reformulation is that each routing option is assigned a distinct *disaggregated gas-rate* variable, q^d_{jk}, for each $d \in \mathcal{D}$, where the set $\mathcal{D}$ is given in Table 4.1. Consequently, we can describe the aggregated total-rates q^{Comp}_k and $q^{\text{HighpresNGPP}}_k$ in (4.7) in terms of q^d_{jk}, thus avoiding bilinear products of y^d_{jk} and q_{jk}. To further improve the mixed integer formulation, we can avoid disaggregated variables R_{jk} by directly transforming the revenue parameter in the objective function (4.6) (Grossmann and Trespalacios, 2013),

$$Z = \max\ C^{\text{oc}} \sum_{j\in\mathcal{J}} \sum_{k\in\mathcal{K}} \sum_{d\in\mathcal{D}} \left(R^d_{jk} q^d_{jk} - \sigma^{\text{u}}\, \delta q^{\text{u}}_k\right) \Delta k. \tag{4.11}$$

By substituting R^d_{jk} with the terms given in Appendix 4.A, the objective function is formulated as

$$\begin{aligned} Z = \max\ C^{\text{oc}} \sum_{j\in\mathcal{J}} \sum_{k\in\mathcal{K}} \Big(& G^{\text{NGPP}} \left(\left(1 - C^{\text{C}}\right) q^{d1}_{jk} + q^{d22}_{jk}\right) \\ & + G^{\text{spot}} q^{d21}_{jk} - \sigma^{\text{u}}\, \delta q^{\text{u}}_k \Big) \Delta k, \end{aligned} \tag{4.12}$$

while the two common constraints (4.7) and (4.9) expressed in terms the disaggregated flow variables q^d_{jk} yield

$$\sum_{j\in\mathcal{J}} \left(q^{d1}_{jk} + q^{d22}_{jk}\right) + \delta q^{\text{u}}_k = d^{\text{NGPP}}_k, \qquad \forall k \in \mathcal{K}, \tag{4.13a}$$

$$\sum_{j\in\mathcal{J}} q^{d1}_{jk} \le q^{\text{up}}_{\text{tot}}, \qquad \forall k \in \mathcal{K}. \tag{4.13b}$$

The nonlinearities in (4.10) are due to univariate polynomials of $p_{\text{wf}}, p_{\text{t}}$ and q. As p_{wf} appears only as a mapping from $p^2_{\text{wf},jk}$, while p_{t} appears similarly in a quadratic term in (4.10c) in addition to simple bounds in the disjunction (4.10d), we can reduce the nonlinearity in (4.10) by defining new variables as the *square* of its original variables,

$$\bar{p}_{\text{wf}} := p^2_{\text{wf}}, \tag{4.14a}$$

$$\bar{p}_{\text{t}} := p^2_{\text{t}}, \tag{4.14b}$$

and modify all simple bounds on p_{t} accordingly. Note that all the pressures are defined as non-negative variables. By these substitutions of variables, the only remaining nonlinearities are the q^2_{jk} terms in (4.10c). To obtain a complete MILP model, we hence formulate q^2_{jk} as a piecewise linear function using special-order sets of type 2 (SOS2) (Beale and Tomlin, 1970), see Appendix 4.B.

Finally, we include constraints for defining minimum up and down times between each succeeding well shut-in cycle (Knudsen and Foss, 2013), as well as minimum stay times on the flowlines downstream of the second valves in Fig. 4.5, both comprised by the general symbolic equations Ω in (4.10f). Minimum up and down times are typically required from an operational point of view, to avoid excessive actuation from too frequent switchings, causing wear and tear of the surface equipment. If liquids have accumulated in parts of the well prior to a shut-in, then it may also be necessary to impose a minimum shut-in time to achieve sufficient pressure build-up before the well is reopened. Denoting the minimum down (shut-in) time τ_1 and the minimum up (production) time τ_2, the logical propositions for limiting the frequency of well shut-ins are

$$\left(\neg Y^{d0}_{jk-1} \wedge Y^{d0}_{jk}\right) \Rightarrow \bigwedge_{\tau=k+1}^{k+\tau_1-1} Y^{d0}_{j\tau} \,, \qquad j \in \mathcal{J}, k \in \{\mathcal{K} : 1 < k \leq K - \tau_1 + 1\} \,, \quad (4.15)$$

$$\left(Y^{d0}_{jk-1} \wedge \neg Y^{d0}_{jk}\right) \Rightarrow \bigwedge_{\tau=k+1}^{k+\tau_2-1} \neg Y^{d0}_{j\tau} \,, \quad j \in \mathcal{J}, k \in \{\mathcal{K} : 1 < k \leq K - \tau_2 + 1\} \,. \quad (4.16)$$

Using the techniques in Raman and Grossmann (1991), these expressions are converted to the linear constraints

$$y^{d0}_{jk} - y^{d0}_{jk-1} \leq y^{d0}_{j\tau}, \qquad \forall j \in \mathcal{J}, k \in \{\mathcal{K} : 1 < k \leq K - \tau_1 + 1\} \,, \; \tau \in [k+1, \min\{k + \tau_1 - 1, K\}] \quad (4.17a)$$

$$y^{d0}_{jk-1} - y^{d0}_{jk} \leq 1 - y^{d0}_{j\tau}, \qquad \forall j \in \mathcal{J}, k \in \{\mathcal{K} : 1 < k \leq K - \tau_2 + 1\} \,, \; \tau \in [k+1, \min\{k + \tau_2 - 1, K\}] \quad (4.17b)$$

Equivalent constraints are also imposed for the minimum stay times for the second switching valves, replacing y^{d0}_{jk} in (4.17) with y^{d21}_{jk}. Note that by imposing the switching constraints (4.17) on y^{d0}_{jk} and y^{d21}_{jk}, respectively, we also impose the same conditions on the remaining binaries from the algebraic reformulation of (4.10d) due to the SOS type 1 constraints (4.32p) and (4.34f), allowing only one binary in each subset $\{d0, d1, d2\}$ and $\{d21, d22, \neg d2\}$ of $\mathcal{D}$ to be nonzero (Beale and Tomlin, 1970). Observe that the additional term $\neg d2$ arises from reformulation of the embedded disjunction in (4.10d) as a separate disjunction to allow the state $\neg d21 \wedge \neg d22$, see Appendix 4.A. It should be commented that Rajan and Takriti (2005) derived the convex hull polytope of an extended variable formulation of the minimum up and downtime constraints. However, as we do not include switching costs in our model, we can retain a reduced problem size by only using the constraints (4.17).

4.4 Decomposition by Lagrangian relaxation

A large number of shale-gas wells is needed to supply the NGPP with high, sustainable natural-gas rates. This leads to a large-scale primal MILP, which problem size impedes a direct solution by a fullspace method. A preferred solution approach

is therefore to solve the field-wide shale-well scheduling problem by a decomposition scheme. Problems with few linking constraints and a block-separable structure can often be efficiently solved using Lagrangian relaxation (LR) (Guignard, 2003). By dualizing the linking constraints with Lagrangian multipliers, the relaxation changes the primal problem to separable form, and, being a relaxation, renders an easier way of computing an upper bound on the optimal objective value (for maximization problems). Solving a large-scale mixed integer program by Lagrangian relaxation is an iterative technique that consists of solving the Lagrangian for given input multipliers to compute an upper bound (in terms of a maximization objective), updating the multipliers, and recovering primal feasibility from the solution of the Lagrangian.

Let $\lambda_k \in \mathbb{R}$ and $\pi_k \geq 0$ be Lagrangian multipliers associated with the constraints (4.13a) and (4.13b), respectively. Furthermore, let $g_j\left(m_{jk}, p_{jk}, q_{jk}, y^d_{jk}\right) \leq 0$ for $j = 1 \dots J$ denote the set of constraints (4.3), (4.35)–(4.17) and (4.32)–(4.34), defining the separate scheduling for each well $j \in \mathcal{J}$. Dualizing the $2 \times |\mathcal{K}|$ constraints (4.13) yields the Lagrangian

$$\begin{aligned} Z_{\mathrm{LR}}(\lambda, \pi) = \max\ & C^{\mathrm{oc}} \sum_{j\in\mathcal{J}} \sum_{k\in\mathcal{K}} \Big(G^{\mathrm{NGPP}} \left(\left(1 - C^{\mathrm{C}}\right) q^{d1}_{jk} + q^{d22}_{jk} \right) \\ & + \hat{G}^{\mathrm{spot}} q^{d21}_{jk} - \sigma^{\mathrm{u}}\, \delta q^{\mathrm{u}}_k \Big) \Delta k \\ & + \sum_{k\in\mathcal{K}} \pi_k \Big(q^{\mathrm{up}}_{\mathrm{tot}} - \sum_{j\in\mathcal{J}} q^{d1}_{jk} \Big) \\ & + \lambda_k \Big(\sum_{j\in\mathcal{J}} \left(q^{d1}_{jk} + q^{d22}_{jk} \right) + \delta q^{\mathrm{u}}_k - d^{\mathrm{NGPP}}_k \Big), \end{aligned} \tag{4.18}$$

which can be solved as $|\mathcal{J}|$ independent subproblems given by the MILP

$$\begin{aligned} Z_{\mathrm{LR},j}(\lambda, \pi) = \max\ & C^{\mathrm{oc}} \sum_{k\in\mathcal{K}} \Big(G^{\mathrm{NGPP}} \big(\left(1 - C^{\mathrm{C}}\right) q^{d1}_{jk} + q^{d22}_{jk} \big) + \hat{G}^{\mathrm{spot}} q^{d21}_{jk} \Big) \Delta k \\ & + \left(\lambda_k - \pi_k\right) q^{d1}_{jk} + \lambda_k q^{d22}_{jk} \\ \text{s.t.}\quad & g_j\left(m_{jk}, p_{jk}, q_{jk}, y^d_{jk}\right) \leq 0, \quad \text{for given } j \in \mathcal{J},\ \forall k \in \mathcal{K}, \\ & y^d_{jk} \in \{0,1\}, \qquad\qquad \forall j \in \mathcal{J}, k \in \mathcal{K}, d \in \mathcal{D}, \end{aligned} \tag{4.19}$$

for any given set of feasible input multipliers (λ, π). The variable δq^{u}_k for total underproduction with respect to the demand d^{NGPP}_k is not included in the subproblems (4.19). This may be somewhat detrimental for the quality of the upper bound computed by the Lagrangian. Notwithstanding, we can compute δq^{u}_k from a separate subproblem,

$$Z^{\delta}_{\mathrm{LR}}(\lambda) = \max_{\delta q^{\mathrm{u}}} \sum_{k\in\mathcal{K}} \left(\lambda_k - C^{\mathrm{oc}} \sigma^{\mathrm{u}} \Delta k\right) \delta q^{\mathrm{u}}_k \tag{4.20a}$$

$$\text{s.t.}\quad 0 \leq \delta q^{\mathrm{u}}_k \leq d^{\mathrm{NGPP}}_k, \qquad \forall k \in \mathcal{K}, \tag{4.20b}$$

where the upper bound on δq_k^{u} follows from the observation that the total underproduction cannot be larger than the demand d_k^{NGPP}. Consequently, the upper bound on Z is given by,

$$Z \leq Z_{\mathrm{LR}}(\lambda, \pi) = \sum_{j \in \mathcal{J}} Z_{\mathrm{LR},j} + Z_{\mathrm{LR}}^{\delta}. \tag{4.21}$$

To compute the least upper bound on Z, we need to solve the Lagrangian dual,

$$Z_{\mathrm{D}} = \min_{\lambda,\pi} Z_{\mathrm{LR}}(\lambda, \pi). \tag{4.22}$$

The Lagrangian $Z_{\mathrm{LR}}(\lambda, \pi)$ can be shown to be piecewise linear and a convex function of (λ, π) (Nemhauser and Wolsey, 1988), while nondifferentiable at the points where the optimal solution is not unique (Guignard, 2003). Bundle methods, comprising trust-region and proximal bundle methods (Hiriart-Urruty and Lemarechal, 1993, Ch. XV), are robust methods for solving the Lagrangian dual that efficiently address the issue of nondifferentiability. For a given set of input multipliers (λ^n, π^n), let $Z_{\mathrm{LR}}^n = Z_{\mathrm{LR}}(\lambda^n, \pi^n)$ denote the solution of (4.18) computed in iteration n of the Lagrangian scheme. At this dual point (λ^n, π^n), it can be shown that $f(\lambda^n, \pi^n) := [c^n, h^n]$, where

$$c^n = c(\lambda^n) = \sum_{j \in \mathcal{J}} \left(q_{jk}^{d1,n} + q_{jk}^{d22,n} \right) + \delta q_k^{\mathrm{u},n} - d_k^{\mathrm{NGPP}}, \tag{4.23a}$$

$$h^n = h(\pi^n) = q_{\mathrm{tot}}^{\mathrm{up}} - \sum_{j \in \mathcal{J}} q_{jk}^{d1,n}, \tag{4.23b}$$

defines a subgradient of $Z_{\mathrm{LR}}(\lambda, \pi)$ (Hiriart-Urruty and Lemarechal, 1993, Ch. XII). For different input multipliers (λ^u, π^u), we iteratively aggregate corresponding subgradients and solutions of the Lagrangian in a *bundle* $\mathbb{B} := \{(\lambda^u, \pi^u),\ Z_{\mathrm{LR}}^u,\ [c^u, h^u],\ u = 1, ..., U\}$, which is used to construct a *cutting-plane model*

$$\tilde{Z}_{\mathrm{LR}}(\lambda, \pi) := \max \left\{ Z_{\mathrm{LR}}^u + c(\lambda^u)\big(\lambda - \lambda^u\big) + h(\pi^u)\big(\pi - \pi^u\big), u = 1, \dots U \right\}. \tag{4.24}$$

The model (4.24) is a convex lower approximation of the Lagrangian $Z_{\mathrm{LR}}(\lambda, \pi)$, where each cut u defines a linearization of the nondifferentiable function $Z_{\mathrm{LR}}(\lambda, \pi)$ at the point (λ^u, π^u). Proximal bundle methods for computing optimal multipliers (λ, π) are based on solving the Lagrangian dual by a regularized form of the cutting-plane model (4.24). The regularization is constructed around a *prox* (or stability) center $(\bar{\lambda}, \bar{\pi})$, in which we solve the following quadratic program (QP) to find new multiplier updates (Kiwiel, 1990):

$$\min_{v,\lambda,\pi} \ v + \frac{1}{2}\gamma^n \left(||\lambda - \bar{\lambda}||^2 + ||\pi - \bar{\pi}||^2 \right) \tag{4.25a}$$

s.t.

$$v \geq c(\lambda^u)\big(\lambda - \bar{\lambda}\big) + h(\pi^u)\big(\pi - \bar{\pi}\big) - \bar{\alpha}^u, \qquad \forall u = 1...n \tag{4.25b}$$

where $||\cdot||$ is the Euclidean norm and $\gamma > 0$ is a scalar penalizing deviations from the prox center. The parameter $\bar{\alpha}^u \geq 0$ is the linearization error between the

Lagrangian and the u'th cut in the model (4.24), evaluated at the current prox center $(\bar{\lambda}, \bar{\pi})$,

$$\bar{\alpha}^u := Z_{\mathrm{LR}}(\bar{\lambda}, \bar{\pi}) - Z^u_{\mathrm{LR}} - c(\lambda^u)\big(\bar{\lambda} - \lambda^u\big) - h(\pi^u)\big(\bar{\pi} - \pi^u\big). \tag{4.26}$$

When solving (4.25), v is the predicted (nominal) decent from the current prox center (Kiwiel, 1990),

$$v^n = \tilde{Z}_{\mathrm{LR}}(\lambda^{n+1}, \pi^{n+1}) - Z_{\mathrm{LR}}(\bar{\lambda}, \bar{\pi}). \tag{4.27}$$

After solving the Lagrangian (4.18) for the new multipliers $(\lambda^{n+1}, \pi^{n+1})$, the prox center is updated if $Z_{\mathrm{LR}}(\lambda^{n+1}, \pi^{n+1})$ is sufficiently decreased relative to v^n, according to the rule

```
if:    Z_LR(λ^{n+1}, π^{n+1}) ≤ Z_LR(λ̄, π̄) + δ^S v^n,
       ⇒  λ̄ = λ^{n+1}, π̄ = π^{n+1}                    (serious step)
else:
       λ̄ = λ^n, π̄ = π^n                               (null step)
```

for a predefined descent coefficient $\delta^{\mathrm{S}} \in (0, 0.5)$. The algorithm terminates with optimal multipliers $(\bar{\lambda}, \bar{\pi})$ if $v^n \geq -\delta^{\mathrm{B}}$ for some small positive *tolerance* δ^{B}. A weak point of proximal bundle methods is that the numerical efficiency may rely strongly on the choice of updating strategy for γ^n. In this paper we implement the safeguarded quadratic interpolation strategy from Kiwiel (1990), a commonly applied technique for updating the weight γ^n on the prox center. The initial value, γ^1, is set equal to the norm of the first subgradient.

4.4.1 A novel primal recovery heuristic

The solution obtained by solving the Lagrangian (4.18) is generally infeasible with respect to the dualized constraints (4.13). Recovering primal feasible solutions from solving the Lagrangian hence normally requires the use of some heuristic procedure. Commonly applied primal recovery heuristics in Lagrangian relaxation include constructing problem specific heuristics with priority lists, e.g. (Borghetti et al., 2003), augmenting the Lagrangian with a strong convexification term (Dubost et al., 2005; Sagastizábal, 2012), and partial (Gollmer et al., 2000; Takriti and Birge, 2000) or full fixing of the set of binary variables to its current Lagrangian solution (Knudsen et al., 2014a).

In this paper, we develop a fixing-based Lagrangian heuristic inspired by the Feasibility Pump (Fischetti et al., 2005) heuristic and the branch-and-bound improvement heuristic presented in (Danna et al., 2004). While the heuristics in these two references are generic MILP construction and improvement heuristics, respectively, our approach uses the solution of the Lagrangian relaxation as a *basis* for the heuristic, and seeks to combine the projection strategy in Fischetti et al. (2005) and the fixing strategy in Danna et al. (2004). The idea of the heuristic is to recover a primal feasible solution in the proximity of the current solution of the Lagrangian by fixing only a subset of the binary variables. Given the current Lagrangian solution $y^{d,n}_{jk}$, the initial step of the heuristic is to find a point $\tilde{y}^d_{jk}$ close to $y^{d,n}_{jk}$ that

yields primal feasibility, but not necessarily integer feasibility. Denoting y^n as the vectorized representation of $y^{d,n}_{jk}$, we obtain this point $\tilde{y}^n$ by solving the linear program (LP)

$$\tilde{y}^n = \arg\min \, ||y - y^n||_1 + \rho^n \sum_{k\in\mathcal{K}} \delta q^{\mathrm{u}}_k \tag{4.28a}$$

s.t.

$$\sum_{j\in\mathcal{J}} \left(q^{d1}_{jk} + q^{d22}_{jk}\right) + \delta q^{\mathrm{u}}_k = d^{\mathrm{NGPP}}_k, \qquad \forall k \in \mathcal{K}, \tag{4.28b}$$

$$\sum_{j\in\mathcal{J}} q^{d1}_{jk} \leq q^{\mathrm{up}}_{\mathrm{tot}}, \qquad \forall k \in \mathcal{K}, \tag{4.28c}$$

$$g_j\left(m_{jk}, p_{jk}, q_{jk}, y^d_{jk}\right) \leq 0, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K}, \tag{4.28d}$$

$$y^d_{jk} \in [0,1], \qquad \forall j \in \mathcal{J}, k \in \mathcal{K}, d \in \mathcal{D}. \tag{4.28e}$$

The LP (4.28) is a relaxation of the fullspace primal problem with the original objective (4.12) replaced by the augmented distance function (4.28a). The last term in (4.28a), $\rho^n \delta q^{\mathrm{u}}_k$, where ρ^n is a nonnegative penalty parameter, is added to prevent recovery of primal solutions with high underproduction δq^{u}_k, as this would lead to poor solution quality. We implement an adaptive updating strategy for ρ,

$$\rho^{n+1} = \min\left(\psi\rho^n, \rho^{\max}\right), \qquad \psi > 0 \tag{4.29}$$

increasing the penalty on δq^{u}_k with increasing number of LR iterations n, in order recover solutions with low underproduction. Note that the ℓ_1-norm in (4.28a) is easily reformulated as a linear function, see Fischetti et al. (2005).

Unless the solution of (4.28) yields $||\tilde{y}^n - y^n||_1 = 0$, which would correspond to a primal feasible solution to the original MILP, we obtain a point $\tilde{y}^d_{jk}$ with some of the binaries being fractional in order to generate a feasible solution to (4.28). To generate a *primal integer feasible* solution rendering a lower bound Z_{LB} on Z, that is, a solution which satisfies both the integrality requirement of y^d_{jk} as well as the common constraints (4.28b) and (4.28c), we first fix all y^d_{jk} that returns a binary solution value $\tilde{y}^d_{jk}$ when solving the LP (4.28). The argument for this variable fixing follows the intuition that binaries y^d_{jk} that retain an integer value after the ℓ_1-projection of the Lagrangian solution $y^{d,n}_{jk}$ onto the LP feasible region (4.28b)–(4.28e), are likely to have the same integer value also in a primal integer feasible solution. Fixing these binary variables reduces the size of the solution space significantly, allowing a primal feasible solution to be obtained by solving the reduced-size fullspace MILP problem.

The suggested primal recovery approach is general and hence applicable to a wide range of mixed integer problems solved by Lagrangian relaxation. Initial testing revealed that as much as 85-95% of the binary variables could be fixed for many LR iterations. However, for large instances of the shale-well scheduling problem in Section 4.3, the branching effort required due to the SOS2 approximations (4.35) of the nonlinear tubing model caused prohibitive long computation times also for the reduced-sized fullspace MILP. To overcome this issue, we apply the following

problem-specific heuristic for the piecewise linear approximations: First, we replace the condition $\theta^b_{jk} \in$ SOS2 with the equivalent condition imposed by the constraints (4.36) and the additional binary variables y^b_{jk} associated with the $|\mathcal{B}^{\mathrm{p}}| - 1$ linear functions, see 4.B. By post-calculating y^b_{jk} based on the solution values of θ^b_{jk} from the Lagrangian, we then fix these binaries in the same way as the binaries y^d_{jk} by integrating them in (4.28). If this extended binary fixing causes integer infeasibility for the MILP, we relax the binaries $y^{\tilde{b}}_{jk}$ which are fixed to 1, together with the two adjacent binaries $y^{\tilde{b}-1}_{jk}$ and $y^{\tilde{b}+1}_{jk}$ to allow q_{jk} to vary on a wider range of possible values.

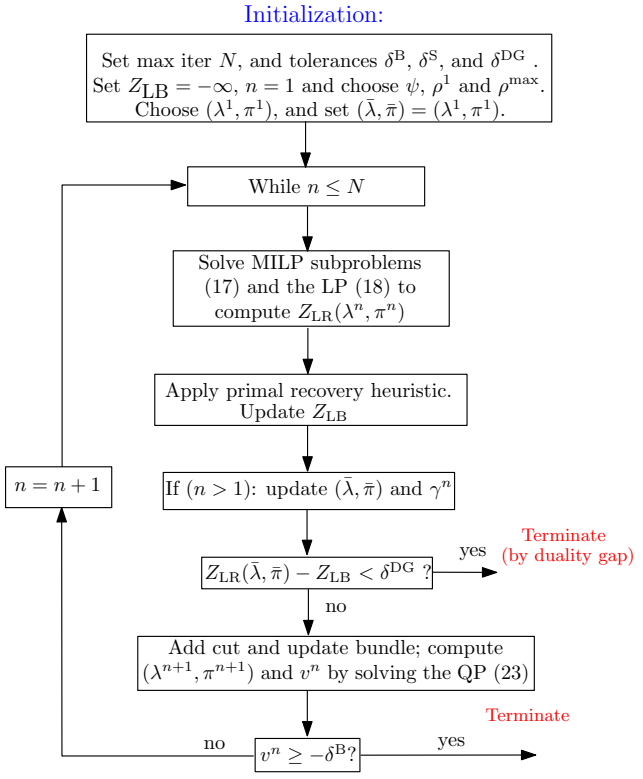


Fig. 4.6: Schematic description of the proposed Lagrangian relaxation scheme.

The complete Lagrangian relaxation scheme is shown in Fig. 4.6, comprising the solution of the Lagrangian, the primal recovery heuristic and the proximal bundle method for updating the Lagrangian multipliers.

4.5 Receding horizon optimization

Solving the above field-wide shale-well scheduling problem involves several sources of uncertainty. These uncertainties include parameters in the well and reservoir model (4.3)–(4.5), uncertainties in the spot price estimate $\hat{G}^{\mathrm{spot}}$, operational un-

certainties and disturbances such as a sudden drop in the rate for some wells, unplanned shut-ins or required maintenance, and changes in the line pressures. Moreover, the overall system may change during the planning horizon $\mathcal{K}$ if new wells are tied to the gathering system. Some control scheme is hence necessary to ensure reliability of the natural-gas supply to the NGPP.

To handle operational disturbances and uncertainties, as well as a varying gas spot price G^{spot}, we implement a receding horizon scheme (Rawlings and Mayne, 2009; Mayne et al., 2000), also referred to as model predictive control (MPC). At the current control time t_k, we compute optimal production settings on a prediction horizon K, while we only implement the *first* optimal control input, i.e. $p_{\text{t},j1}$ and y^d_{j1} for all $j \in \mathcal{J}$ and $d \in \mathcal{D}$. At the next time control time, t_{k+1}, we reoptimize the system on a receding horizon $K+1$. Between the control inputs, we collect system information and measurements, hence introducing feedback in the system. At each iteration t_k of the receding horizon scheme, we use the optimal solution of the Lagrangian from the previous iteration, t_{k-1}, as a starting point for the MILP subproblems (4.19), together with $(\bar{\lambda}, \bar{\pi})$ and γ from iteration t_{k-1} as initial guess for the multipliers (λ^1, π^1) and the initial penalty parameter in the bundle method (4.25).

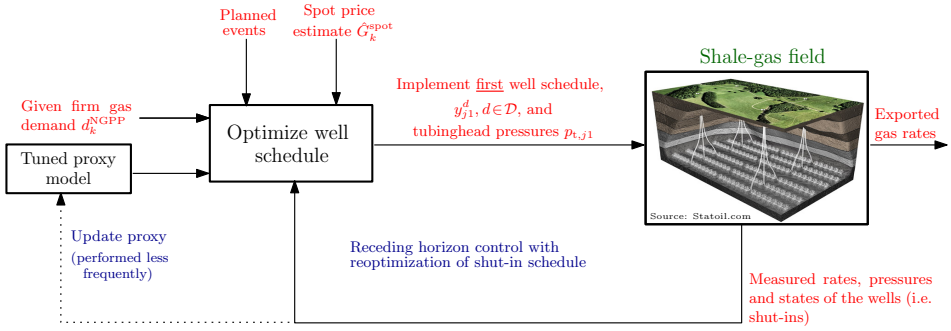


Fig. 4.7: Modules in the receding horizon scheme.

The implemented receding horizon scheme is illustrated in Fig. 4.7. At a given control time t_k, we assume that a spot price estimate $\hat{G}^{\text{spot}}_{t_k}$ for the prediction horizon K is given as a system input, provided by a third-party. If the reoptimization of the well schedule is performed on a daily basis, then each evening after closing hours of the gas trading, the gas spot price for the day ahead delivery is collected (FERC, 2012), together with an updated gas-price estimate $\hat{G}^{\text{spot}}_{t_{k+1}}$. Note that the gas price $\hat{G}^{\text{spot}}_1$ for the first timestep in the optimization, i.e. for the first day ahead, will be the actual settled gas price. The updating of the spot price will hence correct the price estimate $\hat{G}^{\text{spot}}_2$ used in the previous computed well schedule at iteration t_{k-1}. This set-up of the gas spot price introduces a trade-off between trusting the future price estimate $\hat{G}^{\text{spot}}_k$ compared to the known one day ahead price. To accommodate this trade-off when computing the optimal well schedule,

we introduce a discount factor,

$$\hat{G}_k^{\text{spotDisc}} = \frac{\hat{G}_k^{\text{spot}}}{(1+\xi)^{(k\times\Delta k)}}. \tag{4.30}$$

The discount rate ξ in (4.30) represents *risk* more than depreciation of income: Assigning a high value to ξ would reduce the risk of losing profit if the price should decrease. In contrast, a low discount factor corresponds to trusting the spot price estimate $\hat{G}_k^{\text{spot}}$, in which the gas production to the spot market may be choked to wait for an estimated higher future spot price.

For each well, we collect available measurements such as the gas rate and tubinghead pressure. In particular, we collect the state of each well, i.e. if the well is producing or shut-in. Wells may experience sudden, unplanned shut-ins due to several reasons, including mismatch in actual and predicted rate (e.g. the gas rate drops below q_{gc}) and failure of surface equipment. These events amount to operational uncertainties. Wells may further have to be shut-in for certain periods due to planned maintenance. Either of these events are handled through the feedback in the receding horizon scheme, by fixing binaries y_{jk}^{d0} to 1 for the next k' timesteps for those wells that must be shut-in.

To obtain closed-loop behavior in practice by a receding horizon control scheme, there are several other aspects that must be considered. Unless a measurement of the complete state vector is available, then some estimation technique, typically a Kalman filter or a moving horizon estimation technique (MHE) (Rao et al., 2001), must be applied to estimate the initial condition for iteration t_{k+1}. For the shale-well problem, this amounts to estimating m_j^{init}. In the current implementation, however, we omit for simplicity the state estimator, and hence apply the pseudopressure predicted by the proxy to update $m_{jt_{k+1}}^{\text{init}}$. Furthermore, an updating strategy for the proxy model should be applied, using any collected measurements of the rates and pressures. This can be performed using either the simple updating strategy described in Knudsen et al. (2014a), or by some MHE based technique (Rao et al., 2001).

In conventional stabilizing tracking MPC (Mayne et al., 2000), the change in control input is normally penalized to avoid excessive actuation causing wear and tear of the equipment. The proposed receding horizon scheme, however, resembles economic MPC (EMPC) (Amrit et al., 2011), in which adding such a penalization term would deteriorate the economic interpretation of the objective (4.6). The constraints (4.17) added to limit the switching frequency of the wells, however, serve a similar purpose as a penalization term on the change in control input. To enforce these constraints also in the shift from one receding horizon iteration to the next, we implement a *move-blocking* strategy Cagienard et al. (2007). Before each reoptimization, we check if the well has changed its state, either from on to off or vice versa, and if so, we *fix* y_{jk}^{d0} to its current value for the next $\tau_1 - 1$ or $\tau_2 - 1$ timesteps. The same type of variable fixing is applied to y_{jk}^{d21} to avoid excessive changes in the state of the second valve in Fig. 4.5. Note that by fixing either y_{jk}^{d0} or y_{jk}^{d21}, we also fix the variables $(y_{jk}^{d1}, y_{jk}^{d2})$, $(y_{jk}^{d22}, y_{jk}^{\neg d2})$, respectively, due to the SOS1 constraints (4.32p) and (4.34f) associated with the reformulation of the disjunctions (4.10d).

4.6 Case studies

In this section we demonstrate the proposed shale-well scheduling scheme through numerical case studies. To construct a time series of firm gas demands d_k^{NGPP} from an NGPP, we use as basis the gas demand in the US electric power sector between mid June and mid September as shown in Fig. 4.2. This period falls during the time of year with normally the highest gas demand in the electric power sector. The gas prices used in the model, G^{NGPP} and G^{spot}, are shown in Fig. 4.8, where we use the Henry Hub spot price from June to September 2013, cf. Fig. 4.2, to define the spot price G_k^{spot} (EIA, 2013b). The price G_k^{NGPP} for firm gas supply is more difficult to quantify, as it will depend on the price negotiated and contract set up by the marketing groups of the gas well and NGPP operators. The volumetric rates for the firm gas supply to the NGPP may typically be contracted many months ahead of actual delivery, using expected seasonal variations, e.g. as shown in Fig. 4.2. Hence, to define a reasonable firm price level for G_k^{NGPP}, we use the NYMEX future average price for four-months ahead contracts (EIA, 2013b). Hence, the G^{NGPP} June price shown in Fig. 4.8 is the NYMEX price for four-months ahead contracts from March 2013, while the G^{NGPP} July price is the corresponding NYMEX forward price from April, etc. To construct a nominal gas price used for the input spot price estimate $\hat{G}^{\text{spot}}$, we simply use the Henry Hub spot price as mean and add Gaussian noise with 2% standard deviation, see Fig. 4.8. Note that the price G_k^{NGPP} for the firm supply do not include the transportation fee paid to the shipper; The final price paid *by* the EUCs and GENCOs for the firm supply to their NGPPs will hence be higher than the value of G_k^{NGPP} shown in Fig. 4.8.

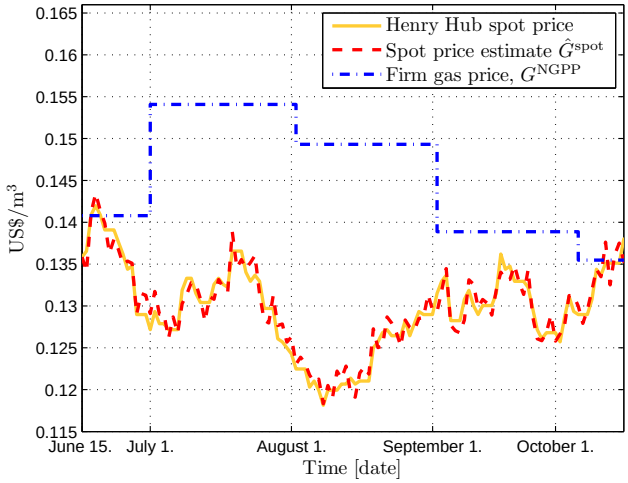


Fig. 4.8: The Henry Hub spot price (EIA, 2013b), the nominal spot price estimate $\hat{G}^{\text{spot}}$, and the gas price G^{NGPP} for the firm delivery to the natural-gas power plant. Observe that the prices corresponds to the NYMEX and Henry Hub spot price from 2013 shown in Fig. 4.2.

The Lagrangian scheme outlined in Fig. 4.6 is implemented in GAMS (Brooke et al., 2011) and connected with `Matlab` through the GDX interface for implementation of the receding horizon scheme. All the QPs and the MILPs are solved using

IBM CPLEX v12.3. The computations are performed on a Dell laptop with Intel I7 quad-core CPU and 8GB of RAM, using deterministic parallel mode with up to 8 threads. For the bundle method, we set $\delta^{\mathrm{B}} = 0.05$, $\delta^{\mathrm{S}} = 0.1$ and the maximum number of iterations to $N = 50$. The duality gap tolerance δ^{DG} for both the MILP subproblems and for the Lagrangian scheme is set to 2%. The parameters in (4.29) for the primal recovery technique is set to $\rho^1 = 5, \rho^{\max} = 10$ and $\psi = 1.2$.

When generating the initial conditions for the case studies, we follow the same procedure as in Knudsen et al. (2014a). In particular, we apply reservoir properties corresponding to typical values in the Marcellus formation (EIA, 2011; Cipolla et al., 2010), and we realize the proxy models using three different formation permeabilities in the range $k_{\mathrm{m}} \in \left[5 \times 10^{-5} - 2 \times 10^{-4}\ \mathrm{mD}\right]$. Each reservoir proxy model is constructed with $I = 4$ grid blocks, and tuned against a high-fidelity reference model implemented in SENSOR (SENSOR, 2011), using a tuning technique similar to Knudsen et al. (2014a). From these realizations, we set m_j^{init} by evaluating the pseudopressures for the different wells at randomly chosen points in time ranging from 200 to 600 days after start-up of the wells. The piecewise linearization of the nonlinear tubing model (4.5a) is constructed using $|\mathcal{B}^{\mathrm{p}}| = 5$ breakpoints. We assume a royalty rate RO and tax T of 38.5% and 12.5%, respectively (MIT Energy Initiative, 2011, App. D).

We use a four week prediction horizon $\mathcal{K}$ for the receding horizon (RH) scheme, where we use a $\Delta k = 1$ day timestep for the first week of the horizon, and a $\Delta k = 3$ days timestep for the last three weeks. We only enforce the switching constraints (4.17) for the first week of the horizon, using $\tau_1 = \tau_2 = 2$ days. The reasoning to enforce only these constraints for the first part of the prediction horizon, is that only the *first* timestep of each iteration of the receding horizon optimization is implemented, and, by using $\Delta k = 3$ for the last three weeks, the timestep is longer than the minimum required up- and downtime for the wells. The schedule is reoptimized by the receding horizon scheme on a daily basis. The operational disturbances are simulated using a Binomial probability distribution with a probability of 5% that a well experience a sudden, unplanned shut-in. Consequently, if a well has an unplanned shut-in, we fix y^{d0} to 1 in the first timestep of the subsequent receding horizon iteration. For all cases we apply a 20% discount factor ξ for the three last weeks of the prediction horizon K. Adding this discount factor serves two objectives: the estimated gas price $\hat{G}^{\mathrm{spot}}$ becomes significantly more uncertain beyond the first week of the prediction horizon, and as such, choking the production due to an estimated possible increase in the gas price would be risky for the operator. Additionally, it reduces the weights on the gas-rates q_{jk}^{d21} in (4.12) for the three last weeks of the prediction horizon, as these terms end up with a high weight due to $\Delta k = 3$ for this part of $\mathcal{K}$.

The output power P^{NGPP} generated by a natural-gas combined-cycle power plant can be described by the relation (Munoz et al., 2003)

$$P^{\mathrm{NGPP}} = LHV\, \eta(q^{\mathrm{NGPP}})\, q^{\mathrm{NGPP}} \tag{4.31}$$

where LHV is the low heating value for natural gas, 35.07 $\mathrm{MW/m^3/s}$ (Munoz et al., 2003), and $\eta(q^{\mathrm{NGPP}})$ is the combined cycle efficiency, dependent on the gas supply rate q^{NGPP} to the power plant. Using the same procedure as in (Munoz-Estrada

et al., 2004), we will use (4.31) with a given desired, possibly time-varying, output power P^{NGPP} and a given combined-cycle efficiency $\eta = 53.8\%$ (Arvesen et al., 2013) to compute the firm gas-demand rates d_k^{NGPP}.

Table 4.2: Problem size for the MILPs in the case studies.

	$\lvert\mathcal{J}\rvert$	Binary var	SOS2 var	Cont. var	Constraints
Case 1	10	860	700	2995	5409
Case 2	60	5160	4200	17895	32309
Case 3	105	9030	7350	31305	56519

We use three different case studies in order to demonstrate and analyze different aspects of the proposed shale-well scheduling scheme. In case 1, we use a small generator as a means of motivating the receding horizon scheme compared to the open-loop solution. Moreover, we analyze the computational performance of the proposed Lagrangian relaxation scheme. In case 2, we evaluate the solutions for different discount factors ξ, using as example a 300 MW NGPP, which is supplied with gas from a medium sized shale-gas field with 60 wells. This example also addresses the sensitivity of the scheduling with respect to the spot price estimate $\hat{G}^{\text{spot}}$. Finally, in case 3 we demonstrate the application of the proposed scheduling scheme for a shale-well field with more than hundred producing wells, supplying an NGPP with a 400 MW generator running at a high load-factor and having high demands on the reliability of the supply. Problem sizes are shown in Table 4.2.

In case 2 and 3, we will assess the potential economic benefit for the shale-well operator from applying the proposed scheduling scheme, by comparing the result with the following *base case*: Each well is assumed to continue producing gas to the flowline it were initially producing to, and we assume an ideal production scenario where no shut-ins are necessary, i.e. we ignore the imposed minimum rate q_{gc} in (4.10d) to prevent liquid loading. The latter simplification can be justified by Fig. 4.4, in the sense that almost all production lost during a shut-in is immediately recovered by the subsequent peak-rate from reopening the well. We further ignore costs from operating the compressor, and we assume that all the gas produced is sold to an LDC at the spot price G^{spot}.

4.6.1 Case 1: Test case with a 43 MW generator and 10 producing wells

In the first case, we consider a small-scale combined-cycle NGPP running a generator with maximum 43 MW output capacity, e.g. the ALSTOM GTX100 gas turbine. The NGPP is supplied with gas from a shale-gas field with 10 producing wells. We assume that the EUC/GENCO persistently seeks full load of its generator in order to maximize the efficiency, causing a constant demand d_k^{NGPP} as shown with black dots in Fig. 4.9, and we use a 8 weeks time horizon for the example.

Fig. 4.9 compares the solution of the receding horizon scheme with the solution obtained by applying the *open-loop* solution to the shale-gas field. In the open-loop solution, we assume that the operator for the first four weeks applies the initial solution from the LR scheme computed on June 15, and then re-optimize the scheme four weeks later, i.e. on July 13, in order to generate the optimal well

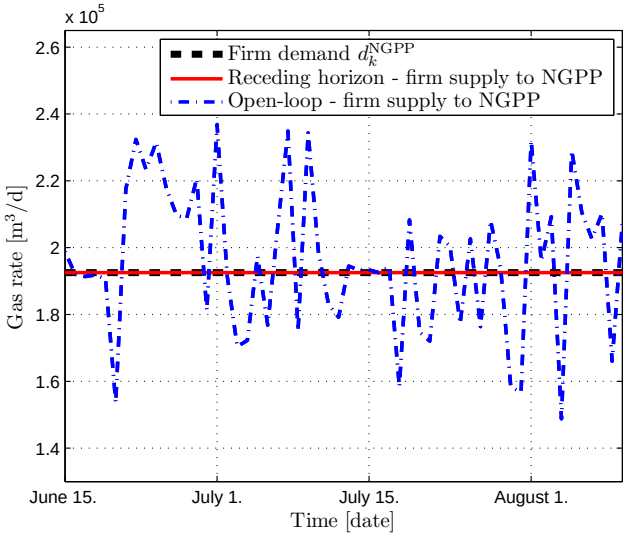


Fig. 4.9: Comparison of the receding horizon scheme's and the open-loop scheme's ability to track the demand rate d_k^{NGPP}.

schedule for the next four weeks. We assume that the first unplanned well shut-in occurs 5 days after initialization of the optimal well schedule. In Fig. 4.9, it can be seen that the open-loop solution gives satisfactory tracking of d_k^{NGPP} for the first few days when there are no disturbances, while the tracking performance of the open-loop scheme is substantially deteriorated in the event of an unplanned shut-in. The large oscillations seen in the supply are due to the peak rates occurring after shut-ins of shale-gas wells (Knudsen and Foss, 2013). The deviations between d^{NGPP} and that actual gas supply to the NGPP is in this case unacceptably high. In contrast, the solution of the receding horizon scheme meets in this case the demand d_k^{NGPP} exactly over the entire time horizon. This is achieved by the *feedback* introduced through receding horizon optimization, in which the scheme in each iteration recomputes tubinghead pressures $p_{\mathrm{t},jk}$ and well schedules y_{jk}^d to compensate for unplanned shut-ins and thereby meet the demand. The receding horizon scheme is clearly superior to the open-loop scheme for handling unplanned events and disturbances, and actually a necessity in order to tightly meet the demand d_k^{NGPP} as time progress.

Solving the Lagrangian subproblems (4.19) for the 10-well case required on average 0.5 seconds, ranging from 0.03 to 4.9 seconds. The convex hull reformulation (4.32)–(4.34) of the GDP formulation (4.10d) for the routing decisions for each well, renders as such a tight MILP formulation, requiring limited computation time for solving each of the Lagrangian subproblems. To demonstrate the performance of the primal recovery fixing heuristic in Section 4.4.1, we show in Fig. 4.10 the progress in improvement of the lower bound Z_{LB} and the % of the binary variables fixed for the eighth RH iteration in case 1. Note that the % number of binary variables fixed includes the additional binaries y_{jk}^b imposed for the piecewise linear approximations during the primal recovery, cf (4.36). By reusing the optimal solution from previous

RH iteration t_{k-1} as starting point for the Lagrangian scheme, the primal recovery heuristic initially fixes a high number of variables since the provided starting point is almost primal feasible. The value of the Lagrangian tends to increase slightly in the next iterations as cuts need to be generated and added to the bundle $\mathbb{B}$ in order to refine the polyhedral approximation $\tilde{Z}_{\text{LR}}$ in (4.24) and hence produce good multiplier updates $(\lambda^{n+1}, \pi^{n+1})$. During these iterations the number of fixed binary variables is seen to drop, as the solution of the Lagrangian is further from being primal feasible. After this initial drop, the number of fixed binary variables is seen to remain at a level between 85-95%, and gradually increase with the number of LR iterations n, during which the lower bound Z_{LB} is seen to be consistently improved.

The same progress of the primal recovery heuristic as shown in Fig. 4.10 is also observed for the majority of the remaining receding horizon iterations. Out of all the Lagrangian relaxation iterations applied in case 1, the initial fixing procedure in the primal recovery heuristic fixed averagely 86% of the binary variables. The fixing of binaries y^b_{jk} associated with the piecewise linearization is, however, somewhat troublesome; in 41% of the total number of LR iterations in case 1, the initial binary fixing caused the primal MILP to be integer infeasible. However, after the relaxation of the SOS associated binaries y^b_{jk} as described in Section 4.4.1, leaving averagely 67% of the binaries fixed, then only 0.2% of the LR iterations were not able to return a primal feasible solution. Consequently, with an 1.1% average duality gap at the termination of each RH iteration, the proposed primal recovery heuristic and Lagrangian scheme is able to efficiently produce solutions of high quality.

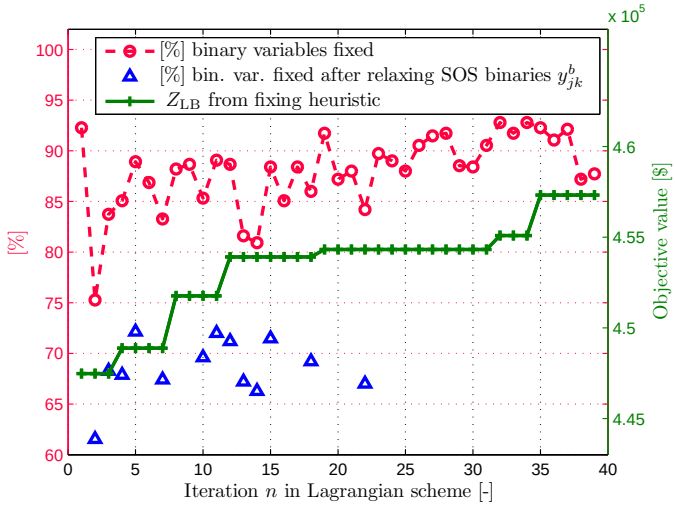


Fig. 4.10: The percentage of variables fixed and the improvement of the lower bound for the Lagrangian scheme in RH iteration number 8, case 1. The final duality gap in this RH iteration is 1.0%.

4.6.2 Case 2: A 60 wells shale-gas field supplying a 300 MW NGPP

In case 2, we consider a medium-sized combined cycle NGPP with a maximum power output of 300 MW (Arvesen et al., 2013). The demand d_k^{NGPP} is set relative to the gas demand in the electric power sector shown in Fig. 4.2. Many NGPPs purchase their natural gas using a portfolio with both firm and interruptible supply (NERC, 2011). As such, we assume in this case that the NGPP seeks to secure a certain gas supply based on an expected electric power demand. In order to reduce the risk of having to pay for surplus gas due to a lower electric power demand than forecasted, we assume that the NGPP requests between 48-70% of the maximum output of 300 MW, purchasing any remaining needed gas supply from an LDC. These assumptions lead to the monthly varying, firm demand rate d_k^{NGPP} shown with black strokes in Fig. 4.11. The penalty σ^{u} for the underproduction δq_k^{u} is set equal to five times G_1^{NGPP}.

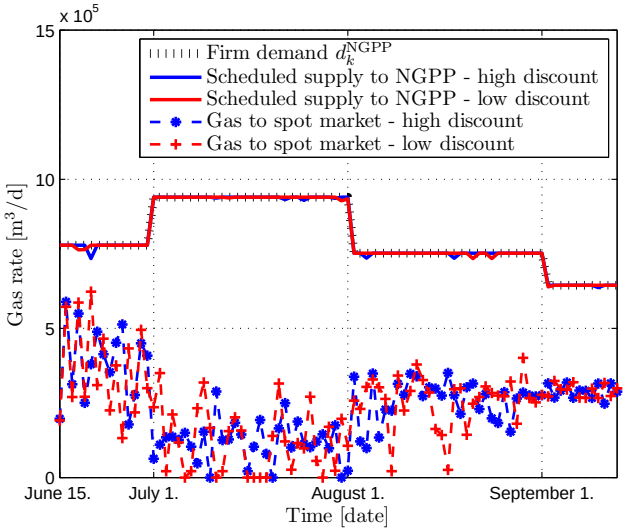


Fig. 4.11: Produced rates for case 2 supplying a 300 MW power plant with natural gas.

Table 4.3: Average optimization results for the Lagrangian relaxation based receding horizon scheme in case 2.

	Average		Median	
	High discount ξ	Low discount ξ	High discount ξ	Low discount ξ
Duality gap [%]	2.0	1.9	1.8	1.8
# LR iterations	12.8	13.5	11	11
Time per RH iterations [min]	15.4	16.4	12.3	12.7
v^n at termination	-0.38	-0.34	-0.28	-0.29
Total computation time [hours]	23.1	24.5	-	-
Total revenue [10^6 US$]	6.72	6.74	-	-

In Fig. 4.11 we compare the results of using two different discount factors ξ as defined in (4.30). The results in the blue line apply a $\xi = 4\%$ discount factor

for the first week, and subsequently a $\xi = 20\%$ discount factor for the last three weeks of the prediction horizon K, while the red line applies no discount to the first week and equally a $\xi = 20\%$ for the rest of the horizon. In both cases, the scheduled firm supply to the NGPP is seen to tightly follow d_k^{NGPP}, however, the case with the lowest discount factor (red line) gives a slightly better tracking. The small observable discrepancies, corresponding to nonzero values of δq_1^{u}, results from tuning the optimization scheme and not the capacity of the shale-gas wells, since the field is producing gas to the spot market at the dates where the produced firm supply differs from d_k^{NGPP}. Comparing the scheduled rates in Fig. 4.11 with the gas price in Fig. 4.8, it can also be observed that the highest discrepancies between d_k^{NGPP} and the scheduled supply to the NGPP occur early in the horizon during which the spot price G_k^{spot} is slightly higher than G_k^{NGPP}. Using the same figures to compare the spot price with the gas sold to the spot market, it can be seen that the production follows the trend of the spot price, causing a higher spot rate during early July and late August when the price is high compared to the low-point in mid August. Although both spot production rates are erratic, the blue line corresponding to the highest discount factor is observed to exhibit a spot rate that is slightly less erratic than the spot rate with the low discount factor.

The solution statistics of the Lagrangian relaxation based receding horizon scheme are shown in Table 4.3 for the two cases with different discount factors. The results for the two cases are quite equal, both spending around 15 minutes on each iteration of the receding horizon scheme and terminating with around 2% duality gap. As for case 1 in Section 4.6.1, the first iterations of the receding horizon scheme requires significantly more LR iterations, and terminates either by the condition $v^n \geq -\delta^{\text{B}}$ or by the maximum number of iterations N. The same naturally also holds for iterations with a high number of unplanned well shut-ins, since the optimization problems in these cases have a poorer starting point from iteration t_{k-1}. Comparing the optimal solution values, we see that the case with the lowest discount factor actually leads to a marginally higher revenue of 0.27%. In contrast, the base case described above would generate a revenue of 5.94×10^6 US$, more than 11% less than the revenue obtained by either of the two scheduling results shown in Fig. 4.11.

In Fig. 4.12, we compare the result of the scheduling using the nominal gas price estimate $\hat{G}_k^{\text{spot}}$ shown in Fig. 4.8, and a scenario in which the spot price is consistently *overestimated* with up to 10%. Both cases apply the low discount ξ described above. No clear, significant difference is observed in the scheduled rates when comparing the results of using the two different spot price estimates. Both cases exhibit high variations in the gas produced to the spot market, caused both by the gas availability as well as the variability in the spot price, while both cases produce a supply to the NGPP with only small discrepancies from d_k^{NGPP}. Consistently overestimating the spot price gives a more aggressive production strategy similar to the case above with the high discount factor, and actually caused an 1.6% increase in the revenue obtained by the *spot sales* compared to the case with the nominal price estimate. The updating of the day ahead spot price by the receding horizon scheme does as such compensate for a poor quality of the spot estimate. As for the case with the different discount factors, the revenue obtained by either of the schedules shown in Fig. 4.12 is about 11% better than the base case.

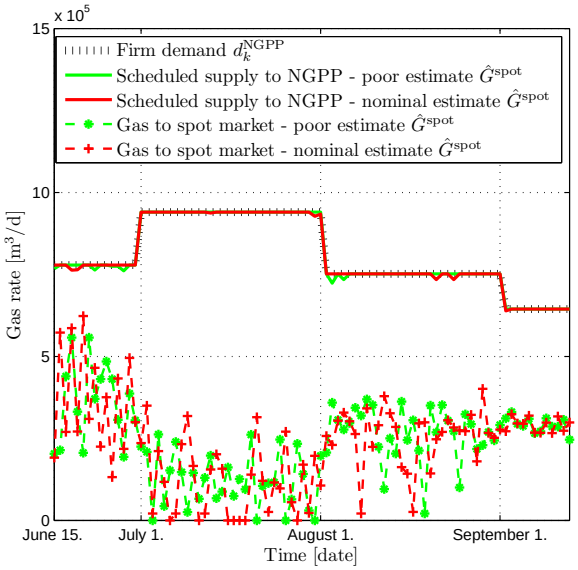


Fig. 4.12: Comparison of scheduled total-rates with the nominal spot price estimate $\hat{G}_k^{\text{spot}}$ shown in Fig. 4.8, and a corresponding poor, too optimistic price estimate.

4.6.3 Case 3: 105 shale-gas wells supplying a 400 MW NGPP

In the last case study, we consider a field consisting of 105 shale-gas wells supplying a combined-cycle NGPP running a 400 MW generator. We assume that the operator of the power plant seeks to run the generator close to its maximum output capacity, particularly during August when the electricity demand peaks, cf. Fig. 4.2, and thus desires a high and reliable firm supply d_k^{NGPP}. Using (4.31), we therefore compute d_k^{NGPP} from respectively 85, 95 and 80% of the generator's given maximum power output, LHV and assumed efficiency η, creating the firm demand d_k^{NGPP} shown with black strokes in Fig. 4.13.

We compare in Fig. 4.13 the spot rate and the scheduled supply to the NGPP using two different values for the penalty σ^{u} for the underproduction δq_k^{u} with respect to d_k^{NGPP}. The low value of σ^{u} is set to 5 times G_1^{NGPP} as in the other examples, while the high value of σ^{u} is set to 10 times G_1^{NGPP}. The scheduled supply to the NGPP can be seen to tightly follow d_k^{NGPP} in both cases, however, with slightly larger deviations in the case with the lowest value of σ^{u}. Although the value of σ^{u} would generally be based on contractual agreements, this example shows that the value of σ^{u} may also be used as tuning parameter and hence increased in order to reduce possible discrepancies between the supplied rate to the NGPP and d_k^{NGPP}. This is particularly relevant if there is nonzero underproduction δq_k^{u}, while surplus gas at the same time is being sold to the spot market. The increased value of σ^{u} to put higher emphasis on meeting d_k^{NGPP} resulted in a 1.9% reduction in the revenue of the gas sold to the spot market compared to the case with the low value of σ^{u}. No significant difference is observed in the pattern of the gas produced to the spot market for the two cases, while both rates are less erratic than in case

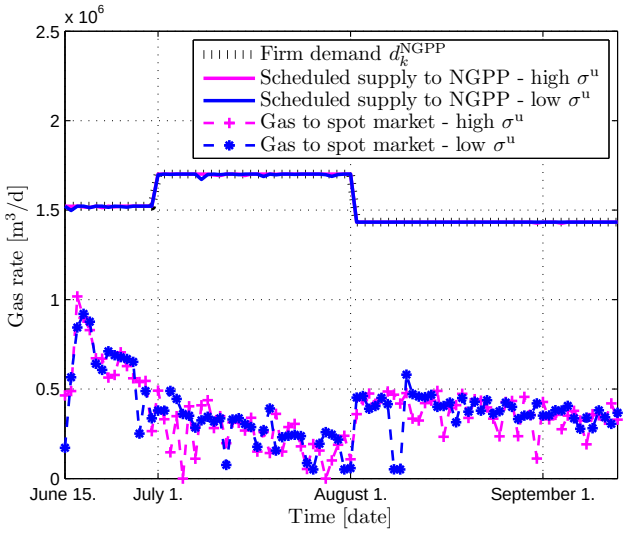


Fig. 4.13: The scheduled rates for the case with 105 shale wells supplying an NGPP running a 400 MW generator with high load.

2, cf. Fig 4.11 and 4.12. The spot production rates follow to some extent the main variations of the gas price in Fig. 4.8, with lower total rates produced to the spot market during the periods of highest demand from the NGPP. There is, however, no distinct coherence between the spot price and the gas produced to the spot market as shown in Fig. 4.14. For well operators, the correlation between price and production as illustrated in Fig. 4.14 would optimally be a linear increase in spot production as a function of the gas price.

Comparing the results of the scheduling in Fig. 4.13 with the base case, the revenue for the shale-well operator is increased with approximately 12% for both cases. The increase in profit is due to a higher sales price for the firm supply to the NGPP, but also due to a 6.5% increase in the cumulative production caused by a higher number of wells producing to the low pressure line than in the base case. However, if we compare the results with a modified base case were all wells are constantly producing to the low-pressure line, the profit by the scheduling scheme is still 9.6% higher. Table 4.4 provides sensitivities of the increase in profit with respect to the value of G^{NGPP}, in which we post-calculate the profit Z in (4.6) with G^{NGPP} replaced by the NYMEX future price for 1 to 3 months ahead delivery, respectively (EIA, 2013b). Recall that G^{NGPP} in the cases studies were set equal to the NYMEX future price for 4 months ahead delivery. The table shows the % increase in profit by the proposed shut-in based scheduling scheme compared to the base case described in the end of Section 4.6, and compared to selling all the gas produced directly to an LDC, receiving the spot price G^{spot} for all the gas. We include the scheduling for both cases with different values of σ^{u}. For all the three lower values of G^{NGPP}, the scheduling scheme still gives a higher profit, both in comparison with the base and when compared to selling all the gas produced directly to an LDC. However, the profit for the operator obviously shrinks when

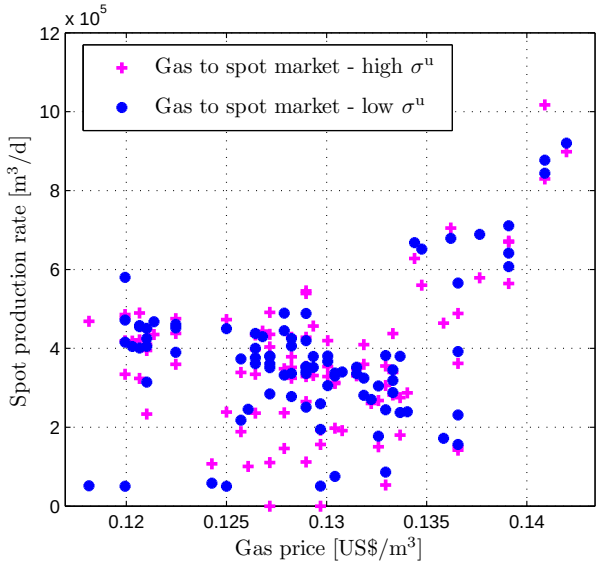


Fig. 4.14: The total spot rate production as a function of the gas spot price for case 3.

the lead time for the sales decreases.

Table 4.4: The % increased profit in case 3 for different values of the firm price G_k^{NGPP} for the gas sold to the NGPP. The profit by the proposed scheduling scheme is compared to the base case, and to selling all the gas obtained by the shut-in scheduling directly to an LDC.

	$\sigma^{\text{u}} = 5 \cdot G_1^{\text{NGPP}}$		$\sigma^{\text{u}} = 10 \cdot G_1^{\text{NGPP}}$	
Increase in profit compared to:	Base case	Spot sales only	Base case	Spot sales only
G^{NGPP} = NYMEX - 4 months ahead delivery	12.2%	11.7%	12.0%	11.8%
G^{NGPP} = NYMEX - 3 months ahead delivery	10.2%	9.7%	10.0%	9.8%
G^{NGPP} = NYMEX - 2 months ahead delivery	5.9%	5.3%	5.6%	5.4%
G^{NGPP} = NYMEX - 1 months ahead delivery	2.2%	1.7%	1.9%	1.7%

In Fig. 4.15 we show the duality gap and number of iterations required by the Lagrangian relaxation scheme for each of the 90 iterations of the receding horizon scheme in case 3. The figure shows that the proposed Lagrangian relaxation scheme consistently finds solutions with small duality gaps, except for the first iterations in which the maximum number of iterations or the tolerance δ^{B} in the Bundle method terminates the LR algorithm. Each iteration of the RH scheme requires on average $n = 10$ iterations of the LR scheme. The reduction in duality gaps and number of iterations from the initial iterations, indicate that the LR scheme benefits significantly from using both the solution as well as the optimal multipliers $(\bar{\lambda}, \bar{\pi})$ from RH iteration t_{k-1} as a starting point for the scheme. Possible remedies for reducing the duality gap in the first iterations may hence be to provide a better starting point and initial guess for (λ^1, π^1), or to increase N and reduce δ^{B} to allow more iterations. Furthermore, the primal recovery heuristic may enter a loop where the same set of primal feasible solutions are recovered, similar to the problem of cycling in the Feasibility Pump (Fischetti et al., 2005). A possible extension of the

proposed recovery heuristic is therefore to implement a perturbation technique, or to reset the value of the penalty parameter ρ^n, to explore new areas of the solution space. As for case 1, the solution of each LR subproblem takes on average 0.5 seconds, while each LR iteration requires averagely 25 minutes; the computation time for each LR iteration can be significantly reduced by solving the Lagrangian subproblems in parallel. However, when the number of wells runs into the hundreds and beyond, the resulting large problem size causes the primal recovery technique to lose its efficiency; if the heuristic is able to fix only 60-70% of the binaries, the remaining number of binaries, is still high, cf. Table 4.2, causing the heuristic to require a large share of the total computation time.

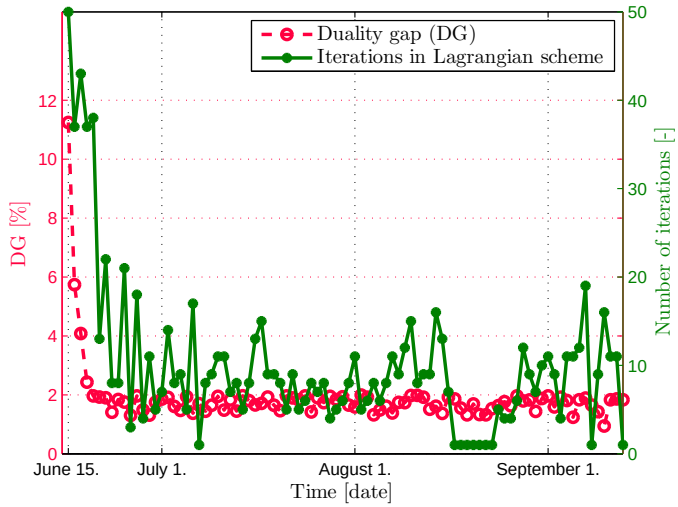


Fig. 4.15: The duality gap and number LR iterations n for each iteration of the receding horizon scheme in case 3 with the low value of σ^{u}.

4.7 Discussion

The goal of this paper is to demonstrate possible benefits for both well operators and for generating companies by scheduling of shale-gas production relative to varying electric power demands, and to develop and test an efficient optimization scheme to facilitate this strategy. The proposed scheme is based on the gas storage properties of shale-gas reservoirs illustrated in Fig. 4.4. The actual % time a well may be shut-in over a given time horizon, e.g. 10-15 years, without reducing the cumulative recovery with more than a few percentage, naturally contains some uncertainty due to unmodeled well dynamics and simplifications in the reservoir fracture modeling. The governing dynamics and driving forces of the gas storage capabilities in shale reservoirs are, however, still captured by the high-fidelity, yet simplified, reference model shown in Fig. 4.3 as well as the proxy model (2.6)–(4.5). As such, we argue that shale-gas reservoirs are particularly suited for intermittent production schemes with respect to varying gas demands.

A necessity for the proposed scheduling scheme to be adopted by the shale-gas industry is a potential increase in profit for the operators. Consequently, the scheduling formulation is formulated with the objective of increasing the profit for the shale-well operator, and shown through the case studies to have a potential to increase the profit by around 10% with the *given* economic and system assumptions. These numbers serve the purpose of illustrating the possible increase in profit for the well operator, and are clearly subject to the firm gas price G^{NGPP} and the contractual arrangement between the marketing group of the shale-gas company, the EUC or GENCO operating the NGPP, and the shipper transporting the gas to the NGPP.

An implication of applying the shale-well shut-in scheme and connecting the gas sales directly from the shale-gas producer to the NGPP, is that the gas storage for meeting variations in demand is essentially moved from the LDC to the gas producer. For the operator of the NGPPs, the economical potential lies both in the possibility of omitting a price-raising link, i.e. the LDC, in the purchase of the gas, as well as increased reliability of the natural-gas supply due to the higher potential of using firm supply contracts. The latter may as such translate to higher reliability in the power production for the NGPP, and hence increased profit. Note that the suggested scheme is also applicable to scheduling of natural-gas supply for the hourly unit commitment (UC) problem, see e.g. Liu et al. (2009). This would however alter the time-scale of the shale-well scheduling, and as such possibly reduce the frequency of shut-ins. Finally, the suggested scheduling scheme relates to the principle of producing only the amount of gas that is secured sales at a given price. Hence, in a broader perspective, the suggested scheme may be utilized for complementing the increased use of intermittent renewable production, as a means of securing acceptable electricity generation reliability in the context of variations in demands, production and errors in forecasts.

A thorough discussion on the optimization scheme developed to solve the shale-well scheduling is left out of the paper. We refer the reader to references such as Grossmann and Trespalacios (2013); Vecchietti and Grossmann (2000); Knudsen et al. (2014a) for details on the disjunctive modeling, Guignard (2003); Frangioni (2005) for a general description of Lagrangian relaxation, to Kiwiel (1990) and (Hiriart-Urruty and Lemarechal, 1993, Ch. XV) for analysis on the convergence properties of the proximal bundle method for solving the Lagrangian dual, and to Rawlings and Mayne (2009); Amrit et al. (2011) for further description of receding horizon optimization. We finish by briefly mentioning the resemblance between the problem formulation for the shale-well scheduling and the deterministic profit-based UC problems. Rather than maximizing the profit from the dispatch of hydro-thermal power units, we similarly schedule pressures and up/down times for shale-gas wells to maximize the profit for the producer. Similarly, the power plants are scheduled subject to an electricity demand and a spinning reserve, i.e. reserve capacity, in the UC problem, while the shale-gas wells are scheduled subject to the given gas demand (4.13a) and the compressor capacity (4.13b). The primal recovery technique described in Section 4.4.1 is hence applicable and may have merits in applications of Lagrangian relaxation for solving variations of the UC problem. Finally, we note that the scheme may benefit from being extended with a stochastic programming approach to further address uncertainties in future gas

prices and demands, and uncertainties in future total gas-supply capacity of the shale-gas field, see e.g. Sahinidis (2004).

4.8 Conclusion

This paper argues that shale-gas reservoirs hold special properties that can be utilized by a proper scheduling scheme to produce reliable gas supplies to natural-gas power plants. The numerical results show that increased profit can be obtained for shale-well operators by extending and improving the well scheduling. The proposed scheme challenges the current practice of producing each well at their maximum irrespective of seasonal demands and price variations, and may, on a longer perspective, mitigate the currently seen abundance of natural-gas supply in the U.S. market.

4.A Reformulation of disjunctions to algebraic constraints

The polyhedral convex hull description of linear disjunctions (Balas, 1985) is obtained by disaggregating variables in the disjunction, and associating each term $d \in \mathcal{D}$ in the disjunction with a binary variable y^d. Each of the original variables in the constraints are then replaced by the associated disaggregated variable for the given term, and constraints are added to ensure that only one of the terms and one of each set of disaggregated variables are active.

In order to reformulate (4.10d), the embedded disjunction in (4.10d) must first be formulated as a separate disjunction (Vecchietti and Grossmann, 2000; Knudsen et al., 2014a). By applying this reformulation, and substituting the auxiliary variables (4.14) for the pressure-squared terms, the convex hull reformulation of the main disjunction (4.10d) yields the linear constraints

$$q = q^{d1} + q^{d2}, \tag{4.32a}$$
$$\bar{p} = \bar{p}^{d0} + \bar{p}^{d1} + \bar{p}^{d2}, \tag{4.32b}$$
$$p_{\mathrm{t}} = p_{\mathrm{t}}^{d0} + p_{\mathrm{t}}^{d1} + p_{\mathrm{t}}^{d2}, \tag{4.32c}$$
$$m_1 = m_1^{d0} + m_1^{d1} + m_1^{d2}, \tag{4.32d}$$
$$R = R^{d1} + R^{d2}, \tag{4.32e}$$
$$R^{d1} = G^{\mathrm{NGPP}} \left(1 - C^{\mathrm{C}}\right) q^{d1}, \tag{4.32f}$$
$$m_1^{d0} = \tilde{a}_1 \bar{p}^{d0} + y^{d0} \tilde{a}_2, \tag{4.32g}$$
$$q^{d1} = \beta \left(m_1^{d1} - \tilde{a}_1 \bar{p}^{d1} - y^{d1} \tilde{a}_2, \right), \tag{4.32h}$$
$$q^{d1} \geq y^{d1} q_{\mathrm{gc}}^{\mathrm{low}}, \tag{4.32i}$$
$$p_{\mathrm{t}}^{d1} \geq y^{d1} \bar{p}^{\mathrm{CompIn}}, \tag{4.32j}$$
$$q^{d2} = \beta \left(m_1^{d2} - \tilde{a}_1 \bar{p}^{d2} - y^{d2} \tilde{a}_2, \right), \tag{4.32k}$$
$$p_{\mathrm{t}}^{d2} \geq y^{d2} \bar{p}^{\mathrm{NGPP}}, \tag{4.32l}$$
$$q^{d2} \geq y^{d2} q_{\mathrm{gc}}^{\mathrm{high}}, \tag{4.32m}$$

$$m_1^d \le U^m y^d, \quad \bar{p}^d \le U^{\bar{p}} y^d, \quad p_{\mathrm{t}}^d \le U^{p_{\mathrm{t}}} y^d, \qquad d \in \{d0, d1, d2\}, \tag{4.32n}$$

$$q^d \le U^q y^d, \qquad d \in \{d1, d2\}, \tag{4.32o}$$

$$y^{d0} + y^{d1} + y^{d2} = 1. \tag{4.32p}$$

For notational convenience, we have left out the subscripts jk; the equations and variables above are still imposed for all wells $j \in \mathcal{J}$ and timesteps $k \in \mathcal{K}$. All of the disaggregated variables are defined as non-negative variables, and will hence be zero if $y^d = 0$ for any $d \in \mathcal{D}$, due to the upper bounds (4.32n) and (4.32o). The constants $U^m, U^{\bar{p}}, U^{p_{\mathrm{t}}}$ and U^q are upper bounds on the respective disaggregated variables in (4.32). The inner embedded disjunction in (4.10d) must be expanded by an additional term with associated Boolean $Y^{\neg d2}$ to allow the state $\neg Y^{d21} \wedge \neg Y^{d22}$, yielding the disjunction

$$\begin{bmatrix} Y_{jk}^{d21} \\ R_{jk} = \hat{G}_k^{\mathrm{spot}} q_{jk} \end{bmatrix} \veebar \begin{bmatrix} Y_{jk}^{d22} \\ R_{jk} = G_k^{\mathrm{NGPP}} q_{jk} \end{bmatrix} \veebar \begin{bmatrix} Y_{jk}^{\neg d2} \\ q_{jk} \le U^q \end{bmatrix} \tag{4.33}$$

Using the same technique as in (4.32) with disaggregation of variables, the embedded disjunction in (4.10d) is reformulated by its polyhedral convex hull description,

$$q = q^{d21} + q^{d22} + q^{\neg d2}, \tag{4.34a}$$

$$R^{d2} = R^{d21} + R^{d22}, \tag{4.34b}$$

$$R^{d21} = G^{\mathrm{spot}} q^{d21}, \tag{4.34c}$$

$$R^{d22} = G^{\mathrm{NGPP}} q^{d22}, \tag{4.34d}$$

$$q^d \le U^q y^d, \qquad d \in \{d21, d22, \neg d2\}, \tag{4.34e}$$

$$y^{d21} + y^{d22} + y^{\neg d2} = 1, \tag{4.34f}$$

$$y^{d2} = y^{d21} + y^{d22}, \tag{4.34g}$$

$$y^{\neg d2} = y^{d0} + y^{d1}. \tag{4.34h}$$

Note that $y^{\neg d2}$ may in principle be eliminated by using (4.34f). Any dependent binary variables are, however, automatically detected and removed by presolve routines in MILP solvers such as CPLEX and Gurobi. The logical propositions (4.10e) are transformed to a linear algebraic form using the techniques described in Raman and Grossmann (1991).

4.B Piecewise linear approximations

By substituting p_{t}^2 and p_{wf}^2 with $\bar{p}_{\mathrm{t}}$ and $\bar{p}_{\mathrm{wf}}$ in (4.10c), respectively, we can apply a piecewise linear approximation of the remaining nonlinearity q^2 using SOS2 sets (Beale and Tomlin, 1970), such that the nonlinear tubing model (4.10c) is

reformulated as

$$q_{jk} = \sum_{b \in \mathcal{B}^{\mathrm{p}}} \theta^b_{jk} q^b_{jk}, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K}, \tag{4.35a}$$

$$\sum_{b \in \mathcal{B}^{\mathrm{p}}} \theta^b_{jk} = 1, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K}, \tag{4.35b}$$

$$\bar{p}_{\mathrm{wf},jk} = e^S \left(\bar{p}_{\mathrm{t},jk} + \frac{1}{C_{\mathrm{t}}^2} \sum_{b \in \mathcal{B}^{\mathrm{p}}} \left(q^b_{jk} \right)^2 \right), \qquad \forall j \in \mathcal{J}, k \in \mathcal{K}, \tag{4.35c}$$

$$\theta^b_{jk} \in \mathrm{SOS2}, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K}, b \in \mathcal{B}^{\mathrm{p}}, \tag{4.35d}$$

where q^b_{jk} are preselected values of q_{jk} at the breakpoints $b \in \mathcal{B}^{\mathrm{p}}$. Most MILP solvers allow a direct definition of θ^b_{jk} as SOS2 variables, in which the solver imposes the condition that at most two variables in the set can be nonzero, and that these must be consecutive variables in the piecewise linear approximation. The same condition can be imposed by introducing $|\mathcal{B}^{\mathrm{p}}| - 1$ binary variables y^b and the SOS2 constraints

$$\theta^1_{jk} \leq y^1_{jk}, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K}, \tag{4.36a}$$

$$\theta^b_{jk} \leq y^b_{jk} + y^{b-1}_{jk}, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K}, b = 2 \dots B^{\mathrm{p}} - 1, \tag{4.36b}$$

$$\theta^{B^{\mathrm{p}}}_{jk} \leq y^{B^{\mathrm{p}}-1}_{jk}, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K}, \tag{4.36c}$$

$$\sum_{b=1}^{B^{\mathrm{p}}-1} y^b_{jk} = 1, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K}. \tag{4.36d}$$

Note that these latter constraints are only implemented in the problem-specific primal recovery heuristic described in the end of Section 4.4.1, to allow a higher number of binary variables to be fixed and hence reduce the required branching effort to obtain a primal feasible solution. When solving the Lagrangian subproblems (4.19), we hence define θ^b_{jk} directly as SOS2, cf. (4.35d), in order to utilize special branching rules for SOS2 variables implemented in the MILP solver.

Finally, we comment that one may consider using an approach based on piecewise McCormick envelopes or multiparametric disaggregation (Kolodziej et al., 2013) as an alternative to piecewise linear approximations of the quadratic terms.

Chapter 5

Towards an Objective Feasibility Pump for Convex MINLPs

This chapter consists of the paper Sharma et al. (2014):

Sharma, S., Knudsen, B. R., and Grimstad, B. (2014). Towards an objective feasibility pump for convex MINLPs. *Manuscript submitted for publication (Computational Optimization and Applications).*

Abstract

This paper describes a heuristic algorithm for finding good feasible solutions of mixed-integer nonlinear programs (MINLPs). The algorithm we propose is a modification of the Feasibility Pump (FP) heuristic, in which we aim at balancing the two goals of quickly obtaining a feasible solution and preserving quality of the solution with respect to the original objective. The effectiveness and merits of the proposed algorithm are assessed by evaluation of extensive computational results from a set of 146 convex MINLP test problems. We also demonstrate applicability of the proposed heuristic on a complex shale-well scheduling problem, and show how a set of user-defined parameters may be selected in order for the user to choose whether low computation time or high solution quality should be emphasized.

5.1 Introduction

Mixed-integer nonlinear programs (MINLPs) are a class of optimization problems consisting of both continuous and discrete variables, and with nonlinear functions in the constraints and/or in the objective function. A large number of optimization problems can be formulated as MINLPs, including scheduling of natural gas and petroleum production systems (Foss et al., 2009), dispatch of hydro-thermal units in electric power generation (Sagastizábal, 2012) and optimization of nuclear reactor reloading (Quist et al., 2001). A comprehensive review of real-world applications of MINLPs can be found in Belotti et al. (2013, Chap. 2). MINLPs belong to the class of NP-hard problems (Vavasis, 1991), and are as such computationally

challenging problems to solve, particularly for large-scale industrial problems. In this paper we will devote our attention to MINLPs which are convex, meaning that the optimization problem obtained by discarding the integrality condition on the discrete variables is a convex nonlinear program (NLP).

Several algorithms for convex MINLPs have been developed during the past three decades, such as: Outer Approximation (Duran and Grossmann, 1986; Fletcher and Leyffer, 1994), LP/NLP based branch-and-bound (Quesada and Grossmann, 1992) and nonlinear branch-and-bound (Gupta and Ravindran, 1985). Common for these algorithms is that they are proven to converge to the global optimum for convex MINLPs. For large-scale problems, however, these exact algorithms often require a prohibitively long computation time, thus limiting their practical applicability. Although heuristics for MINLPs have not yet reached the level of sophistication and widespread acceptance as they have for MILPs (Lodi, 2010), there are indications that they are likely to garner more interest in the coming years due to a steady rise in application areas (Bonami and Gonçalves, 2010; Berthold, 2013). In this paper we present a modification of the Feasibility Pump (FP) heuristic (Fischetti et al., 2005). The FP belongs to the class of primal heuristics (Berthold, 2006) – heuristics designed with the goal of quickly computing primal feasible solutions. The potential benefits of being able to quickly compute feasible solutions are twofold: Firstly, any feasible solution provides an upper bound (for minimization problems) on the optimal solution. Within exact MINLP solvers, an efficient primal heuristics may hence reduce the size of the search space, and thereby accelerate the search for the optimum. Secondly, primal heuristics are generally quicker than exact methods when it comes to finding feasible solutions, hence being particularly useful as standalone algorithms for solving optimization problems within decision support tools with solution-time requirements (Knudsen et al., 2014a). For these applications, finding good feasible solutions in short computation times is more important than finding the globally optimal solution.

The Feasibility Pump was originally introduced as a heuristic for finding feasible solutions to difficult MILPs (Fischetti et al., 2005). The main idea of this heuristic is to solve for two sequences of points, in which one sequence consists of points which satisfy the constraints, while the other sequence consists of points which satisfy the integrality condition. If the two sequences converge to the same point, then a feasible solution is found and the algorithm terminates. Since its initial proposal, a considerable amount of research effort has been devoted to making developments and assessing the theoretical properties of the FP heuristic. Achterberg and Berthold (2007) developed a variant of the FP for MILPs which goal is to improve the quality of feasible solutions obtained by the FP heuristic. The Feasibility Pump 2.0 by Fischetti and Salvagnin (2009) modifies the original FP by utilizing constraint propagation, from which they demonstrate improvements in both solution times, success rates and solution quality. Experiments of using nondifferentiable concave penalty functions as a measure of integrality within an FP framework has been shown to reduce solution times and iterations for binary MILPs (De Santis et al., 2013). Boland et al. (2012) presents an interpretation of the OFP as a discrete proximal point algorithm (Daniilidis and Lemarechal, 2005), and extend the cycle handling in the FP 2.0 with generation of cutting-planes. Their approach is shown to outperform the FP 2.0 in terms of robustness

and solution quality, however, with a substantial increase in required computation time.

Two different extensions of the Feasibility Pump have been developed for convex MINLPs. The first variant, presented in Bonami et al. (2008b), uses outer approximations (OA) of the nonlinear constraints to generate a sequence of points satisfying the integrality condition, thereby solving a sequence of both MILPs and NLPs to find a feasible solution. The second variant, presented in Bonami and Gonçalves (2010), bears a closer resemblance to the original FP for MILPs, and only solves a sequence of NLPs. In a computational study by Bonami and Gonçalves (2010), the former OA-based variant is reported to be slower and on average to find worse solutions than the latter rounding-based variant. The OA-based variant, however, is proven to not cycle if a constraint qualification holds (Bonami et al., 2008b). An FP heuristic for nonconvex MINLPs is presented in D'Ambrosio et al. (2012), which applies a combination of metaheuristic techniques in conjunction with a scheme resembling the OA-based FP for convex MINLPs.

The original FP for MILPs has been shown to generate feasible solutions in short computation times for a wide range of difficult MILPs (Achterberg and Berthold, 2007). The low computation time required for finding feasible solutions comes, however, often at the expense of poor solution quality with respect to the original objective (Bertacco et al., 2007). A similar result is also observed for the FP applied to MINLPs (Bonami and Gonçalves, 2010). The already mentioned FP approach for MILPs by Achterberg and Berthold (2007), called the Objective Feasibility Pump (OFP), tries to overcome this drawback of the original FP scheme (Fischetti et al., 2005) by defining the FP objective as a convex combination of a distance function and the original objective. The OFP is shown to yield improvements in the objective value with only a minor increase in the required computation time. To the best of the authors' knowledge, this approach has not yet been developed for MINLPs. However, adapting the OFP to MINLPs renders several difficulties, among other due to changes in magnitude of the gradients, which are needed to formulate the objective function used in the OFP developed by Achterberg and Berthold (2007). To this end, we formulate an OFP for convex MINLPs as a multi-objective optimization problem, and show how such a formulation can be utilized towards a structured development of an OFP for convex MINLPs. Finally, we assess extension of the proposed multi-objective OFP scheme to nonconvex MINLPs, in which we apply the algorithm on a difficult nonconvex MINLP model for shale-well scheduling (Knudsen et al., 2014a).

The remainder of this paper is structured as follows. In Section 5.1.1, we describe the FP for convex MINLPs from Bonami and Gonçalves (2010) and the OFP for MILPs from Achterberg and Berthold (2007). In Section 5.2, we present our proposed implementation for an OFP for convex MINLPs. In Section 5.3, we describe the computational environment and the results of applying the algorithm to a large set of convex MINLP test problems. Section 5.4 presents an MINLP shale-well scheduling problem, and associated numerical results of using the OFP. Concluding remarks in Section 5.5 ends the paper.

5.1.1 The Feasibility Pump

Consider the following general MINLP

$$\begin{aligned} \min_{x,y} \quad & f(x,y) \\ \text{s.t.} \quad & g(x,y) \leq 0, \\ & x^L \leq x \leq x^U, \\ & y \in \{0,1\}^q, \end{aligned} \tag{5.1}$$

where $f : \mathbb{R}^n \times \mathbb{R}^q \to \mathbb{R}$ and $g : \mathbb{R}^n \times \mathbb{R}^q \to \mathbb{R}^p$ are continuously differentiable functions and $x^L, x^U \in \mathbb{R}^n$ define lower and upper bounds on the continuous variable x. The problem (5.1) is said to be a convex MINLP if the functions $f, g_1, \ldots, g_p$ are convex. Although the algorithm presented in this paper can be extended to MINLPs with general integer variables (Bertacco et al., 2007), we will assume that the integrality constraint restricts y to be a binary variable as the majority of MINLPs are modeled using binary variables.

We will use the two following definitions for classifying the properties associated with points that are found during the execution of the Feasibility Pump algorithm.

Definition A point $(\bar{x}, \bar{y})$ is said to be constraint feasible for (5.1) if $g(\bar{x}, \bar{y}) \leq 0$, $x^L \leq \bar{x} \leq x^U$ and $\bar{y} \in [0,1]^q$.

Definition A point $(\tilde{x}, \tilde{y})$ is said to be integer feasible for (5.1) if $\tilde{y} \in \{0,1\}^q$.

The Feasibility Pump heuristics search for a feasible point to (5.1), that is, a point which is both constraint feasible and integer feasible, by iteratively solving a sequence of projection problems. The FP heuristic starts by computing a constraint feasible point, $(\bar{x}^0, \bar{y}^0)$, obtained by solving the NLP relaxation

$$\begin{aligned} \min_{x,y} \quad & f(x,y) \\ \text{s.t.} \quad & g(x,y) \leq 0, \\ & x^L \leq x \leq x^U, \\ & y \in [0,1]^q, \end{aligned} \tag{5.2}$$

that is, the continuous relaxation of (5.1). If the integer variables, $y_j,\ j \in J = \{1 \ldots q\}$, are integer feasible at this point, then a feasible solution to (5.1) has been found and the heuristic terminates. Otherwise, if some of the integer variables are not integer feasible, then the rounding $\tilde{y}^0 = [\bar{y}^0]$ of the constraint feasible point to the nearest integer point is computed, and the following ℓ_1-projection problem is solved:

$$\begin{aligned} \min_{x,y} \quad & \Delta(y, \tilde{y}) \\ \text{s.t} \quad & g(x,y) \leq 0, \\ & x^L \leq x \leq x^U, \\ & y \in [0,1]^q, \end{aligned} \tag{5.3}$$

with $\tilde{y} = \tilde{y}^0$, and where the distance function in the objective of (5.3) corresponds to the ℓ_1 norm, i.e.

$$\Delta(y, \tilde{y}) := ||y - \tilde{y}||_1 . \tag{5.4}$$

For the case of binary variables y, the norm in (5.4) can be rewritten as the linear function

$$\Delta(y, \tilde{y}) = \sum_{j \,:\, \tilde{y}_j = 0} y_j + \sum_{j \,:\, \tilde{y}_j = 1} (1 - y_j). \tag{5.5}$$

The solution to (5.3) yields the constraint feasible point, $(\bar{x}^1, \bar{y}^1)$. If the objective value at the solution to (5.3) is zero, then a feasible solution to (5.1) has been found and the heuristic terminates. In the case when the objective value is strictly positive, a new integer feasible point is found by rounding the current solution, i.e. $\tilde{y}^1 = [\bar{y}^1]$, and subsequently solving (5.3) using this new integer feasible point in (5.4). The process of solving the projection problem (5.3) and rounding the solution continues until the objective value converges to zero or a termination criterion, e.g. an iteration or time limit, is met.

5.1.2 Algorithmic issues

Generally, there are two difficulties which may occur in the execution of the FP heuristic and hamper the efficiency of the algorithm: stalling and cycling (Fischetti et al., 2005). If the solution to (5.3) yields an integer infeasible point $(\bar{x}^i, \bar{y}^i)$ at an iteration i such that the subsequent rounding of this point is equal to the integer feasible point from the previous iteration, i.e. $\tilde{y}^i = \tilde{y}^{i-1}$, then the algorithm is said to stall. The most widely used method to deal with stalling is to flip a random number of the entries of $\tilde{y}^i$ which correspond to the largest values of $|\bar{y}^i_j - \tilde{y}^i_j|,\ j = 1 \dots q$. For MILPs Fischetti et al. (2005) suggest flipping 10-30 entries, while for MINLPs Bonami and Gonçalves (2010) report that the best results were achieved when only a single entry was flipped.

After a certain number of iterations the algorithm may enter a cycle in which the same sequence of points $(\bar{x}^i, \bar{y}^i)$ and $(\tilde{x}^i, \tilde{y}^i)$ are continuously revisited without ever converging to a feasible solution of (5.1). To overcome this issue, Fischetti et al. (2005) proposed the following random perturbation mechanism: If a new solution to the problem (5.3) was also found in one of the previous three iterations, then a uniformly distributed random number $\rho_j \in [-0.3, 0.7],\ j = 1, \dots, q$ is generated and the value of $\tilde{y}^i_j$ is flipped if the condition $|\bar{y}^i_j - \tilde{y}^i_j| + \max\{\rho_j, 0\} > 0.5$ is satisfied.

5.1.3 The Objective Feasibility Pump for MILPs

The Feasibility Pump described in the previous section generates a sequence of constraint feasible and integer feasible points which may or may not converge to a feasible point. Initially, the constraint feasible point is found by solving the continuous relaxation of the original mixed-integer problem. For the remainder of the solution process, the original objective function is not further included, causing the solution quality with respect to the objective value to often be poor (Bertacco et al., 2007; Bonami and Gonçalves, 2010). The Objective Feasibility Pump for MILPs proposed by Achterberg and Berthold (2007) seeks to overcome

this drawback of the basic FP heuristic by including the original MILP objective also after solving the LP relaxation to obtain the initial constraint feasible solution. In order to describe the algorithm of Achterberg and Berthold (2007), consider the following MILP

$$\begin{aligned} \min_{x,y} \quad & c^T x + d^T y \\ \text{s.t.} \quad & Ax + By \leq b, \\ & x^L \leq x \leq x^U, \\ & y \in \{0,1\}^q . \end{aligned}$$

The objective function used in the OFP for MILPs is formulated as a convex combination of the original objective of the MILP and the distance function $\Delta(y, \tilde{y})$,

$$\begin{aligned} \min_{x,y} \quad & \frac{(1-\alpha_i)}{||\Delta||_2}\Delta(y,\tilde{y}) + \frac{\alpha_i}{||[c\ d]^T||_2}(c^T x + d^T y) \\ \text{s.t.} \quad & Ax + By \leq b, \\ & x^L \leq x \leq x^U, \\ & y \in [0,1]^q, \end{aligned} \tag{5.6}$$

where $\alpha_i \in [0,1]$ is the scalar weight and $[c\ d]^T$ denotes the concatenation of the two objective function vectors. In the case where all y_j are binary, the norm $||\Delta||_2$ is simply the square root of the number of binary variables. At each iteration i, the coefficient α_i is geometrically reduced by a factor $\phi \in (0,1)$, i.e. $\alpha_i = \phi\alpha_{i-1}$ with $\alpha_0 \in (0,1]$[1]. The computational efficiency of the OFP (5.6) for MILPs relies on the ability to easily compute normalization factors for balancing the feasibility and objective terms, hence achieving a well-scaled objective function. Computing a similar normalization for general nonlinear functions is not straightforward (Ding et al., 2006). Consequently, developing an objective feasibility pump for MINLPs requires additional effort and a more comprehensive algorithmic approach.

5.2 An Objective Feasibility Pump for convex MINLPs

The objective function in (5.6) is a normalized convex combination of two functions with conflicting goals: while the first term seeks to satisfy the integrality condition, the second term minimizes the original objective. Each term is normalized by the Euclidean norm of the corresponding gradient. An intuitive approach for extending this method to MINLPs would be to approximate the objective function by a linearization at the point computed in the previous FP iteration, i.e.

$$f(x,y) \approx f(\bar{x}^{i-1}, \bar{y}^{i-1}) + \nabla f^{i-1^T} \begin{pmatrix} x - \bar{x}^{i-1} \\ y - \bar{y}^{i-1} \end{pmatrix},$$

where $\nabla f^{i-1} = \nabla f(\bar{x}^{i-1}, \bar{y}^{i-1})$. This linear approximation can be used to derive the same type of normalization as in the objective function of (5.6). This leads to

[1] Note that $\alpha_0 = 0$ would correspond to the original FP.

the formulation

$$\begin{aligned} \min_{x,y} \quad & \frac{(1-\alpha_i)}{||\Delta||_2}\Delta(y,\tilde{y}) + \frac{\alpha_i}{||\nabla f^{i-1}||_2} f(x,y) \\ \text{s.t.} \quad & g(x,y) \leq 0, \\ & x^L \leq x \leq x^U, \\ & y \in [0,1]^q, \end{aligned} \tag{5.7}$$

where ∇f^{i-1} is ensured to satisfy $\left|\left|\nabla f^{i-1}\right|\right|_2 > 0$. Initial experimental testing, however, revealed that this approach is ineffective. This is most likely due to the random perturbations introduced during the execution of the FP algorithm to tackle convergence issues caused by stalling and cycling. When binary variables of an integer feasible point, $\tilde{y}$, are flipped, the projection term in (5.7) may result in a search direction which can differ significantly from the previous iterations. Hence, using the gradient at a point computed in a previous iteration to construct a local linear approximation of the objective function is likely to be unreliable due to the stochastic nature of the FP heuristic.

To overcome these shortcomings of an ad-hoc OFP formulation (5.7), we use a multi-objective optimization approach for developing an OFP for MINLPs. The two optimization problems (5.2) and (5.3) with equal constraint set can be formulated as one multi-objective problem,

$$\begin{aligned} \min_{x,y} \quad & \{\Delta(y,\tilde{y}), f(x,y)\} \\ \text{s.t.} \quad & g(x,y) \leq 0, \\ & x^L \leq x \leq x^U, \\ & y \in [0,1]^q, \end{aligned}$$

which can be cast into a single-objective optimization problem as a weighted sum of the two objectives

$$\begin{aligned} \min_{x,y} \quad & w_1\Delta(y,\tilde{y}) + w_2 f(x,y) \\ \text{s.t.} \quad & g(x,y) \leq 0, \\ & x^L \leq x \leq x^U, \\ & y \in [0,1]^q, \end{aligned} \tag{5.8}$$

where w_1 and w_2 are user defined weights. As previously mentioned, the two terms in the objective function of (5.8) may differ by orders of magnitude, and it is hence unlikely that we can achieve satisfactory results without scaling the two functions sufficiently. A common technique in multi-objective optimization is to scale each term with the difference in their Nadir and Utopia points (Miettinen, 1999, Ch.2.4). This normalization bounds each function by the range of which they may vary on the Pareto optimal set (Miettinen, 1999, Ch.2.2). Consider the following general multi-objective optimization problem

$$\begin{aligned} \min_{x} \quad & \{f_1(x), f_2(x), \ldots, f_k(x)\} \\ \text{s.t.} \quad & x \in X, \end{aligned} \tag{5.9}$$

where $f_1, \ldots, f_k : \mathbb{R}^n \to \mathbb{R}$ are (possibly) conflicting objective functions and $X \subseteq \mathbb{R}^n$ is the feasible set. The Utopia point for (5.9), $z^U \in \mathbb{R}^k$, provides a lower bound of the Pareto optimal set and is defined as the point at which each individual objective attains its optimal value on the feasible set

$$z_i^U = f_i(x^{[i]}) \quad \text{where} \quad x^{[i]} := \arg\min_x \{f_i(x) : x \in X\}, \ \forall i = 1, \ldots, k.$$

The Nadir point for (5.9), $z^N \in \mathbb{R}^k$, provides an upper bound of the Pareto optimal set and is defined as

$$z_i^N = \max_{1 \leq j \leq k} f_i(x^{[j]}), \ \forall i = 1, \ldots, k.$$

An illustration of what the Nadir and Utopia points may look like in a multi-objective optimization problem with two objectives is given in Fig. 5.1.

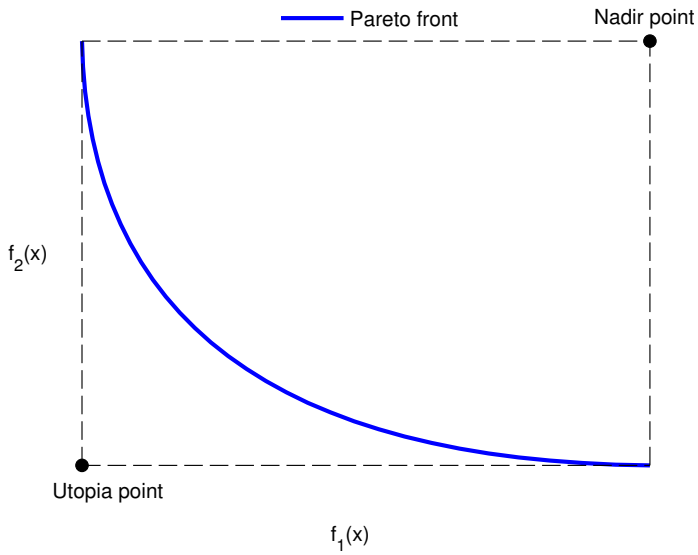


Fig. 5.1: Example of a Pareto front formed by two objective functions $f_1(x)$ and $f_2(x)$.

Nadir and Utopia points for $\Delta(y, \tilde{y})$ and $f(x, y)$ can be computed by the solutions of (5.2) and (5.3). Let $(\bar{x}^0, \bar{y}^0)$ and $(\bar{x}^i, \bar{y}^i)$, $i > 0$ be solutions of the NLP relaxation (5.2) and the ℓ_1-projection problem (5.3), respectively. The Utopia point for $f(x, y)$ is merely the optimal value of the NLP relaxation, while a Nadir point for $f(x, y)$ can be computed from the solution of (5.3), i.e. $f(\bar{x}^i, \bar{y}^i)$. Similarly, the Utopia point for the distance function $\Delta(y, \tilde{y})$ is given by its value at the optimal solution of (5.3), $\Delta(\bar{y}^i, \tilde{y}^{i-1})$, $i > 0$, while its Nadir point is its objective value at the solution of (5.2), $\Delta(\bar{y}^0, \tilde{y}^{i-1})$. Let η_1^i and η_2^i be normalization factors for $\Delta(y, \tilde{y})$ and $f(x, y)$, respectively. From the Nadir and Utopia points, we compute

these normalizations as

$$\eta_1^i = \frac{1}{\Delta(\bar{y}^0, \tilde{y}^{i-1}) - \Delta(\bar{y}^i, \tilde{y}^{i-1})}, \qquad i > 0, \tag{5.10a}$$

$$\eta_2^i = \frac{1}{f(\bar{x}^i, \bar{y}^i) - f(\bar{x}^0, \bar{y}^0)}, \qquad i > 0. \tag{5.10b}$$

This allows us to express (5.8) as the following normalized weighted sum problem:

$$\begin{aligned} \min_{x,y} \quad & u_1 \eta_1^i \Delta(y, \tilde{y}) + u_2 \eta_2^i f(x, y) \\ \text{s.t.} \quad & g(x, y) \leq 0, \\ & x^L \leq x \leq x^U, \\ & y \in [0, 1]^q, \end{aligned} \tag{5.11}$$

where u_1 and u_2 are user defined weights for the two objectives terms. Equivalent to the OFP for MILPs (Achterberg and Berthold, 2007), the goal of our algorithm is to initially minimize the original objective, $f(x, y)$, and then subsequently shift the focus to finding a feasible solution. This is achieved by introducing a weighting factor α_i which is reduced geometrically at each iteration, i.e. $\alpha_i = \phi\alpha_{i-1}$ for some $\phi \in (0, 1)$ and $\alpha_0 \in (0, 1]$. The problem to be solved at iteration i of the OFP can hence be expressed as

$$\begin{aligned} \min_{x,y} \quad & (1 - \alpha_i) u_1 \eta_1^i \Delta(y, \tilde{y}) + \alpha_i u_2 \eta_2^i f(x, y) \\ \text{s.t.} \quad & g(x, y) \leq 0, \\ & x^L \leq x \leq x^U, \\ & y \in [0, 1]^q. \end{aligned} \tag{5.12}$$

It is worth pointing out that α_i may, in addition to u_1 and u_2, be regarded as a user defined weight. However, we will apply α_i strictly as a geometrically reduced, built-in algorithmic parameter. We therefore include the additional weights u_1 and u_2 to allow the user to select whether integer feasibility, and hence low computation time, or quality of the objective function value should be emphasized in the search. This is further elaborated in Section 5.3.1.

An issue with computing the normalizations (5.10) within an OFP algorithm is that they require the solution of an NLP and the FP problem in (5.3). This is due to the fact that each iteration of the OFP algorithm will solve (5.12) and subsequently compute a new integer feasible point $\tilde{y}^i$. Hence, the optimal solution of (5.3) and the normalization parameters (5.10) will need to be recomputed. In order to avoid spending excessive time in computing the constraint feasible points required in the normalization factors (5.10), we only compute the normalizations for the first iteration, and apply these for all subsequent iterations. The background for this decision was that we observed that there was little variation in the parameters (5.10) when we solved (5.12) on the set of convex MINLP test problems used in the computational study in Section 5.3. Experimental testing revealed that this decision had no significant impact on the final objective value. The computation time, however, was greatly reduced. From this simplification, let $\nu_1 := \eta_1^1$ and

$\nu_2 := \eta_2^1$ be the associated normalization factors computed during the first iteration of the OFP. The OFP problem which will be used for the computational study in Section 5.3 is therefore,

$$\begin{aligned} \min_{x,y} \quad & (1-\alpha_i)u_1\nu_1\Delta(y,\tilde{y}) + \alpha_i u_2 \nu_2 f(x,y) \\ \text{s.t.} \quad & g(x,y) \leq 0, \\ & x^L \leq x \leq x^U, \\ & y \in [0,1]^q, \end{aligned} \tag{5.13}$$

where the solution to (5.13) will from here on be denoted by $(\hat{x}, \hat{y})$.

Since the continuous relaxation (5.2) is assumed to be convex, it follows that (5.13) is also a convex NLP, since the ℓ_1-norm is a convex function and the non-negative sum of two convex functions is convex (Boyd and Vandenberghe, 2004).

5.2.1 Rounding

The rounding scheme applied during execution of the FP algorithm may greatly impact its computational efficiency (Bonami and Gonçalves, 2010; Baena and Castro, 2011). Exploiting structures in the problems to improve the rounding procedure may hence translate into a great reduction of required FP iterations. Of particular interest is the presence of constraints

$$\sum_{j \in S} y_j = 1, \tag{5.14}$$

where $S \subseteq \{1, \ldots, q\}$, referred to as type 1 special order set constraints (SOS1) (Beale and Tomlin, 1970) or generalized upper bound constraints (GUB). These constraints enforce the condition that for all binary variables which belong to the set S only a single variable may take a nonzero value. Out of the 146 MINLPs in our set of test problems 135 contain at least one SOS1-constraint (Beale and Tomlin, 1970). The FP implementation in `BONMIN` (Bonami and Lee, 2007) applies so-called *GUB rounding* (Bonami and Gonçalves, 2010) on variables appearing in SOS1 constraints (5.14) to ensure that the rounded point $\tilde{y}$ does not violate any of these constraints. This is achieved by setting $\tilde{y}_j = 1$ for $j = [t]$ and $\tilde{y}_j = 0$ otherwise, where

$$t = \sum_{j \in S} j\hat{y}_j,$$

Remaining binary variables which do not belong to any set S in an SOS1 constraint (5.14) are rounded to their nearest integer value.

To evaluate the default rounding procedure in `BONMIN` when applied to the proposed OFP algorithm, we extend the GUB-based rounding with a technique called *directed rounding from nonlinear constraint violation* presented in Kumar (2008, Ch. 5.2.1). This rounding method can be described as follows: let the solution to (5.13) be denoted by $(\hat{x}, \hat{y})$. For all $\hat{y}_j \notin 0, 1$ which are not in any SOS1 constraints (5.14), we store a temporary solution by rounding this value to zero and to one, respectively, such that

$$\tilde{y}_{j,0} := (\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_{j-1}, 0, \hat{y}_{j+1}, \ldots, \hat{y}_q)$$

and

$$\tilde{y}_{j,1} := (\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_{j-1}, 1, \hat{y}_{j+1}, \ldots, \hat{y}_q).$$

Based on each of these roundings, compute and evaluate the constraint violations

$$\tilde{c}_{j,0} := ||\max(0, g(\hat{x}, \tilde{y}_{j,0})|| \quad \text{and} \quad \tilde{c}_{j,1} := ||\max(0, g(\hat{x}, \tilde{y}_{j,1})|| \,. \tag{5.15}$$

We then keep the rounding which yields the smallest constraint violation and apply GUB rounding for all of the variables which appear in SOS1 constraints. Our experiments revealed that this rounding technique resulted in only a marginal improvement of average computation time and number of iterations, while the average objective value was slightly worse when compared to the results from using the integer- and GUB-rounding only. Hence we apply the default rounding procedure implemented in `BONMIN` in the results reported in our computational study presented in Section 5.3. We did, however, find that the above approach for measuring constraint violation was beneficial in the flipping technique in the stall handling within the OFP algorithm.

5.2.2 Stall and cycle handling with SOS1 constraints

Handling of stalling in the Feasibility pump, i.e. preventing the algorithm from getting stuck in an integer infeasible point, is as mentioned in Section 5.1.2 normally performed by a perturbation strategy. While Fischetti et al. (2005) implements a multi-variable flipping strategy in the original FP for MILPs, Bonami and Gonçalves (2010) observed that best results of their FP implementation for MINLPs where obtained by flipping only a single variable $\tilde{y}_j$ corresponding to the most fractional binary from the previous FP iteration. Though building upon the FP implementation of Bonami and Gonçalves (2010), we observed during experiments with our OFP implementation that improved convergence speed was obtained when flipping a number, drawn randomly from the set $T \in \{1, \ldots, 10\}$, of the most fractional binaries. However, improved stall handling can be obtained by taking into account presence of SOS1 constraints (5.14); when binary variables are flipped, there is a possibility that the resulting point will be infeasible with respect to one or more of the SOS1 constraints. In order to ensure that SOS1 constraints are feasible after flipping variables, we check if any of the flipped variables belongs to a SOS1 set, and subsequently if any SOS1 constraints are infeasible. If any SOS1 constraint is violated, we set to one the binary y_k for $k \in S$, which yields the smallest constraint violation according to the directed rounding constraint violation measurement (5.15) described above, and conversely the remaining binaries y_j, $j \in S \setminus \{k\}$ to zero. A pseudocode for this stall-handling is given in Algorithm 2.

Depending on the value of the geometric reduction factor, ϕ, the *relative* scaling between the two terms in the objective function in (5.13) may not differ significantly between consecutive iterations of the OFP. Thus, for some iterations, the solution to (5.13) may be equal to a previously computed solution. However, the lack of progress in this case is not necessarily due to a cycle, but more likely caused by a lack of significant change in the relative scaling between the two objective terms. To increase the probability that a cycle is detected correctly, we utilize the method

Input: $T, \tilde{y}, (\hat{x}, \hat{y})$
Output: $\tilde{y}^*$

Flip T entries of $\tilde{y}$ corresponding to the largest values of $|\hat{y}_j - \tilde{y}_j|$
Store the indices of the T entries in the index set F
Let $R \leftarrow \emptyset$ contain the indices of the SOS1-constraints in which the flipped variables appear
for $j \in F$ **do**
  if y_j appears in a SOS1-constraint **then**
    Let $g_r(x, y)$ correspond to the SOS1-constraint in the constraint vector
    Let S contain the indices of the variables that appear in the constraint $g_r(x, y)$
    if $\sum_{k \in S} \tilde{y}_k \neq 1$, **then**
      $R \leftarrow \{r\} \cup R$
    end
  end
end
`minViolation` $\leftarrow \infty$
for $r \in R$ **do**
  Let S contain the indices of the variables that appear in the constraint $g_r(x, y)$
  for $k \in S$ **do**
    $\tilde{y}_k \leftarrow 1$
    for $l \in S \setminus \{k\}$ **do**
      $\tilde{y}_l \leftarrow 0$
    end
    `violation` $\leftarrow \left|\left|\max\left\{0, g(\hat{x}^i, \tilde{y})\right\}\right|\right|$
    if `violation` $<$ `minViolation` **then**
      $\tilde{y}^* \leftarrow \tilde{y}$
      `violation` $\leftarrow$ `minViolation`
    end
  end
end
return $\tilde{y}^*$

Algorithm 2: Stall handling in OFP

for cycle detection in the OFP for MILPs (Achterberg and Berthold, 2007): If a solution $(\hat{x}^i, \hat{y}^i)$ is equal to a solution found in a previous iteration $(\hat{x}^k, \hat{y}^k)$, for some $k = i - C$, $1 \leq C < i$, the algorithm is considered to be cycling *if and only if*

$$\alpha_k - \alpha_i \leq \delta_\alpha, \tag{5.16}$$

where δ_α is set to some small positive value. When a cycle is detected, a perturbation is introduced by generating uniformly random numbers $\rho_j \in [-0.3, 0.7]$ for $j = 1, \ldots, q$, and flipping the values of $\tilde{y}_j$ if the condition $|\hat{y}_j - \tilde{y}_j| + \max\{\rho_j, 0\} > 0.5$ holds.

Due to the geometric reduction of the weighting factor α, i.e. $\alpha_i = \phi\alpha_{i-1}$, it may be difficult for the user of the algorithm to give δ_α a meaningful value. From $\alpha_{i-C} - \alpha_i \leq \delta_\alpha$, i.e. by substituting $k = i - C$ in (5.16) and noting that C is the cycle length, it is clear that short cycles are detected before longer cycles. However, it is not clear at which *iteration* $i = N$ a cycle of length C will be detected. This can be specified by further substituting with $\alpha_i = \phi^i\alpha_0$ in (5.16), and rearranging to obtain the relationship

$$\delta_\alpha = \frac{\alpha_0}{C}(1 - C\phi)\phi^{N-1}, \tag{5.17}$$

which gives the value of δ_α so that cycles of length C are detected for iterations $i \geq N$. By setting $C = 1$, which is the shortest possible cycle length, we calculate δ_α from (5.17) by specifying at which iteration N the cycle detection should begin. As discussed above, we do not wish to detect cycles early in the algorithm. On the other hand, a too large N may for many problems prevent detection of cycles; setting $N = 30$ was observed through computational testing to be a good compromise for the value of N. This gives $\delta_\alpha = 0.0047$, which is the default value used for δ_α in our OFP algorithm.

The complete proposed objective Feasibility Pump heuristic for convex MINLP is given by the pseudocode in Algorithm 3, comprising computation of normalization factors, weight selection, rounding procedure and extended stall and cycle handling.

5.3 Computational results

In this section, we describe and present results of the computational experiments with the proposed OFP algorithm. The algorithm is implemented in `BONMIN` version 1.6, which uses `IPOPT` (Wächter and Biegler, 2005) v3.10.2 as NLP solver. We set an upper limit of one solution and no heuristics for `BONMIN`, while the rest of the settings are left to their default values. Our set of test problems consists of 146 convex MINLPs[2] which are described in (Bonami et al., 2008a). The experiments were performed on a personal computer running the 64-bit Ubuntu v. 12.04.3 operating system with Intel Core i7-2600 3.40 GHz CPU and 16 GB of RAM.

The algorithm parameters which are given as input in Algorithm 3 were chosen partly based on those which were used in (Achterberg and Berthold, 2007; Bonami and Gonçalves, 2010) and the rest were determined experimentally: The iteration limit is set to $I_l = 200$, and the OFP parameters are set to $\alpha_0 = 1$, $\phi = 0.9$, and $T \in \{1, 2, \ldots, 10\}$. The user weight for the ℓ_1-distance function (5.4) is set to $u_1 := 1$ and the objective function user weight to $u_2 := 100$.

We use performance profiles (Dolan and Moré, 2002) to present a benchmark of the OFP, FP, and the first solution found by the branch-and-bound method. The performance profile for a solver is the cumulative distribution function for a performance metric, such as the CPU time, the number of nodes processed, the number of function evaluations or the objective value. Given a set of problems $\mathcal{P}$ where $|\mathcal{P}| = n_p$ and a set of solvers $\mathcal{S}$ where $|\mathcal{S}| = n_s$, the performance profile is

[2]These problems can be downloaded from http://egon.cheme.cmu.edu/ibm/page.htm

Input: MINLP of the form (5.1), user weights, u_1, u_2, initial iteration weight α_0, geometric reduction factor ϕ, set of entries to flip T, and iteration limit I_l
Output: Feasible solution (x^*, y^*), when successful

$(\bar{x}^0, \bar{y}^0) \leftarrow$ Solution to the NLP (5.2)
if $\bar{y}^0 \in \{0,1\}^q$ **then**
  $(x^*, y^*) \leftarrow (\bar{x}^0, \bar{y}^0)$
  return (x^*, y^*)
end
$\tilde{y}^0 \leftarrow [\bar{y}^0]$
$(\bar{x}^1, \bar{y}^1) \leftarrow$ Solution to the FP problem (5.3)
$\nu_1, \nu_2 \leftarrow$ Normalization parameters from (5.10a) and (5.10b) with $i = 1$
$\delta_\alpha \leftarrow$ Cycle detection parameter calculated from (5.17) with $C = 1$ and $N = 30$
`/* Main loop */`
while $i < I_l$ **do**
  $i \leftarrow i + 1$
  $\alpha_i \leftarrow \phi \alpha_{i-1}$
  $(\hat{x}^i, \hat{y}^i) \leftarrow$ Solution to the OFP problem (5.13)
  if $\hat{y}^i \in \{0,1\}^q$ **then**
    $(x^*, y^*) \leftarrow (\hat{x}^i, \hat{y}^i)$
    return (x^*, y^*)
  end
  `/* Stall detection and handling */`
  if $[\hat{y}^i] \neq \tilde{y}^{i-1}$ **then**
    $\tilde{y}^i \leftarrow [\hat{y}^i]$
  else
    $\tilde{y}^i \leftarrow$ `handleStall` $\left(T, \tilde{y}^{i-1}, (\hat{x}, \hat{y})\right)$ from Algorithm 2
  end
  `/* Cycle detection and handling */`
  if Cycle detected **and** $\alpha_k - \alpha_i \leq \delta_\alpha$ **then**
    Generate uniform random numbers $\rho_j \in [-0.3, 0.7]$ for $j = 1, \ldots, q$
    if $|\hat{y}^i_j - \tilde{y}^i_j| + \max\{\rho_j, 0\} > 0.5$ **then**
      Flip $\tilde{y}^i_j$
    end
  end
end

Algorithm 3: Objective Feasibility Pump for MINLPs

generated by comparing the results of applying all solvers $s \in \mathcal{S}$ on all problems $p \in \mathcal{P}$.

For each problem p and solver s, the performance $t_{p,s}$ is defined as

$$t_{p,s} := \text{Performance metric on problem } p \text{ by solver } s.$$

The performance on problem p by solver s is compared with the best performance by any solver s on this problem by defining a *performance ratio*

$$r_{p,s} = \frac{t_{p,s}}{\min\{t_{p,s} : s \in \mathcal{S}\}}. \tag{5.18}$$

A parameter r_M is specified so that $r_M \geq r_{p,s}$ for all p, s, and $r_M = r_{p,s}$ if and only if solver s *does not* solve problem p. The cumulative distribution function for the performance ratio is defined as

$$\psi_s(\kappa) = \frac{1}{n_p}\text{size}\{p \in \mathcal{P} : r_{p,s} \leq \kappa\}. \tag{5.19}$$

Hence, $\psi_s(\kappa)$ is the probability that a solver s yields a performance ratio $r_{p,s}$ that is at most worse by a factor of κ of the best ratio.

We construct performance profiles for the MINLP heuristics using two performance metrics: the CPU time to find the first integer feasible solution, and a relative optimality *gap* of this solution defined by

$$\text{Gap} := 100 \times \frac{|\text{Best possible BB sol.} - \text{First feasible sol.}|}{\text{Best possible BB sol.}}, \tag{5.20}$$

where the Best possible BB solution for each of the test problems is computed by running the BB method in `BONMIN` with default settings for *three* hours. In order to compute average values we use a modified geometric mean, which is given by

$$\left(\prod_{j=1}^{N} \max\{1, v_j\}\right)^{\frac{1}{N}}, \tag{5.21}$$

where v_j is the result of a performance metric on a given instance j, and N is the total number of instances.

The OFP, FP, and BB all find a solution for 138 out of the 146 test problems. Results from these test problems are used to calculate the geometric means and sums which are given in Table 5.1. A table with the results from each instance is given in Appendix 5.A.

The results in Table 5.1 show that the OFP finds a solution for 143 instances and the FP for 140 out of 146 test problems. Out of the 138 problems for which all of the methods find a feasible solution, the OFP finds a solution with an improved objective value compared to the FP in 97 of these instances. In 9 of these problems the OFP finds a solution with a relative optimality gap less than 10^{-6}. Moreover, the OFP finds a solution with a gap strictly less than the one found by BB in 31 of the instances. The performance profile in Fig. 5.2 shows a comparison of the three methods where the relative optimality gap is used as performance metric.

Table 5.1: Aggregated results of OFP, FP, and the first solution found by the branch-and-bound method on the set of convex MINLP test problems.

	OFP	FP	BB
GM gap [%][a]	11.10	33.27	5.57
GM time [s][a]	1.74	1.04	7.28
Sum of solution times	288.90	42.54	4966.56
Sum of optimality gaps	11503.59	22676.74	10224.86
Solution found[b]	143	140	145
(better:draw:worse) gap[c]	97:18:23	-	112:11:15
(better:draw:worse) time[c]	10:3:125	-	4:0:134

[a] : geometric mean.
[b] : out of 146.
[c] : compared to the FP.

The figure shows that the OFP finds a solution which is superior to the solution found by the FP for a majority of the test problems. Comparing the performance profile of OFP with both FP and BB, the OFP is seen to overall generate solutions with a relative optimality gap somewhere in between the value computed by BB and FP.

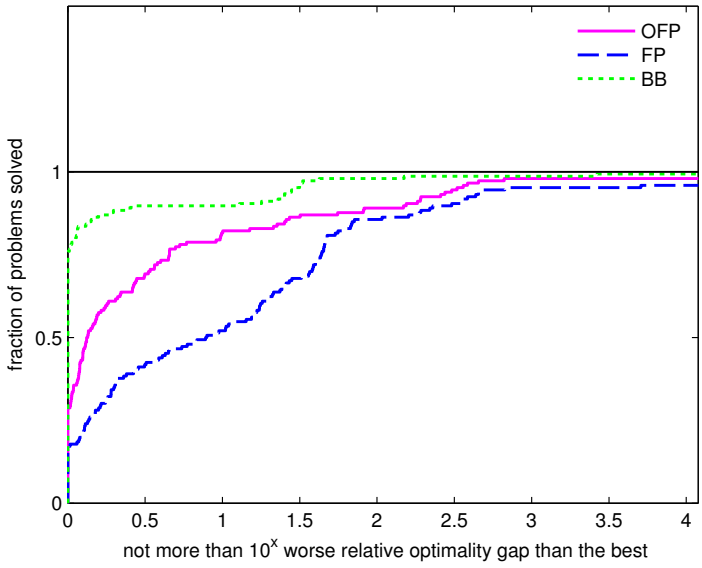


Fig. 5.2: Performance profile of the optimality gap (5.20) with the respect to the best known solution, comparing the results of OFP, FP and the first solution found by the branch-and-bound (BB) method on the convex MINLP test problems.

The OFP is slower than the FP in 125 out of the 138 instances; it yields an 67.3% increase in the geometric mean of the time to the first solution, and correspondingly

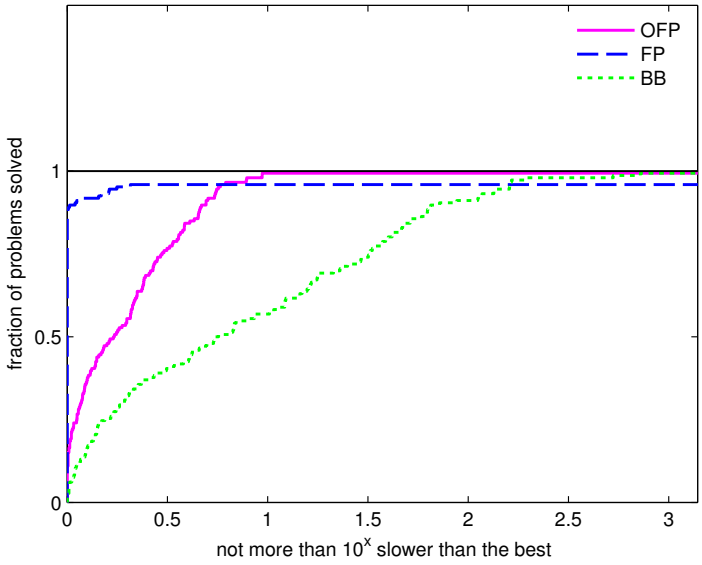


Fig. 5.3: Performance profile of the solution time obtained by comparing the results of OFP, FP and the first solution found by the branch-and-bound method on convex MINLP test problems.

an increase by a factor of 6.8 in the sum of the solution times compared to the FP. In contrast, the OFP has a solution time which is strictly less than BB in 127 of the instances, with the geometric mean of the solution time reduced by a factor of 4.2 compared to BB. The large difference in the average times of the OFP and FP is to some extent caused by a particular large discrepancies in solution time for some of the test problems as displayed in Table 5.A. Hence when comparing the performance profile with the solution time as performance metric in Fig. 5.3, it can be seen that the OFP is not that much slower than the original FP, while being significantly faster than BB for a large fraction of the test problems.

5.3.1 Results of varying the user defined weight

To evaluate the impact of the user defined weight u_2 in (5.13), we perform an experiment where u_2 is varied among the values $\{0.01, 0.1, 1, 10, 100, 1000\}$, while $u_1 = 1$. The OFP is tested with each of these different weights on the set of convex MINLP test problems. Of these 146 test problems, only 116 instances were solved successfully by both the FP and all of the OFP variants with different weights. Consequently, the results of the solution time and relative optimality gap given in detail in Table 5.2 are computed from these 116 instances only. To further illustrate the impact of choosing the weight of u_2, we show in Fig. 5.4 the variation in geometric mean of the solution time and the relative optimality gap for the FP and the OFPs.

The results of varying the value of the user defined weight confirms the notion that the choice of this weight is a trade-off between solution time and solution

quality. As would be expected, a larger value for the user weight u_2 yields an improvement in the average optimality gap. For instance, OFP1000 returns an average optimality gap which is smaller by a factor of 3.07 compared to the average optimality gap found by OFP0.1. However, the average solution time of OFP1000 is nearly twice as large as that of OFP0.1. The trends displayed in Fig. 5.4 show that the behavior of the OFP progressively tends towards the FP as the value of u_2 is decreased. It can also be observed that OFP0.01 is slightly better than the original FP both in terms of solution quality and solution time.

Table 5.2: Detailed results of varying the user weight u_2 in (5.13) on 146 convex MINLP test instances. The number appended to "OFP" indicates the value of u_2.

	FP	OFP0.01	OFP0.1	OFP1	OFP10	OFP100	OFP1000
GM gap [%][a]	35.54	35.29	30.09	17.95	15.15	11.57	9.79
GM time [s][a]	1.041	1.027	1.044	1.170	1.398	1.778	2.039
Sum of solution time	37.37	25.93	43.39	89.41	157.23	265.61	342.03
Sum of optimality gaps	21276	20252	18526	12258	13983	10087	7843
Solution found[b]	140	138	139	141	139	145	140

[a] : geometric mean.
[b] : out of 146.

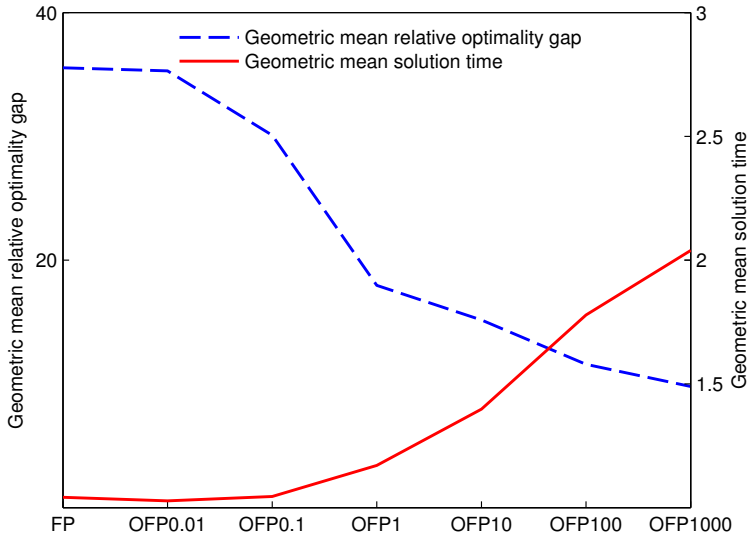


Fig. 5.4: Comparison of geometric mean of the solution time and relative optimality gap obtained by varying the user weight u_2 in (5.13) on 146 convex MINLP test instances. The number appended to "OFP" indicates the value of u_2.

5.4 Example: Shale-well scheduling

The computational results in Section 5.3 show that the proposed OFP heuristic compares favorably in terms of balancing short computation time and solution quality for convex MINLP. These results raise two questions; the first is how the heuristic applies to large-scale real-world problems, for instance as part of industrial decision-support tools (DSTs). In this type of application, where optimization problems are solved repeatedly with a certain time limit, low computation times and sufficient solution qualities are often more important than global optima. The second question is to which extent the results of the proposed algorithm rely on convexity of the NLP relaxation (5.2), and as such if the algorithm can be extended favorably to nonconvex MINLPs. A detailed study and assessment of these questions are beyond the scope of this paper. However, to briefly explore how the proposed OFP performs on these difficult types of problems, we apply the algorithm on an industrial type shale-well scheduling problem taken from Knudsen et al. (2014a).

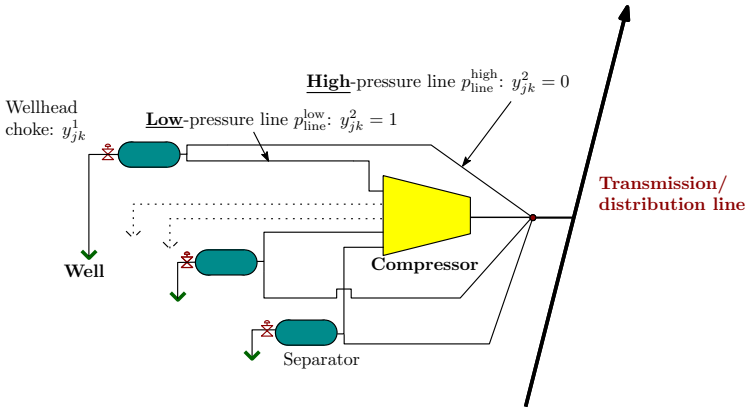


Fig. 5.5: Illustration of shale-well scheduling problem.

Consider the shale-well and compressor system illustrated in Fig. 5.5. The system consists of $|\mathcal{J}|$ geographically distributed wells, each connected with two flowlines to a gathering system: one low-pressure line connected upstream of a shared compression unit, and one high-pressure line bypassing the compressor. The compression is either performed by a midstream company, requiring a fraction C^{C} of the gas sales price G, or the compression is performed by the well operator, causing an equivalent compression cost. Routing gas flow to the low-pressure line increases well deliverability by allowing the well to be operated at a lower wellhead pressure, while routing the gas flow past the compressor gives the operator the full sales price. The compressor requires a minimum inflow rate $q^{\text{low}}_{\text{tot}}$ to avoid surge, and has a maximum load capacity $q^{\text{up}}_{\text{tot}}$.

Each well $j \in \mathcal{J}$ is described by a dynamic, nonlinear well and reservoir model given as a set of constraints $F_j(m_{jk+1}, q_{jk+1}, p_{\text{t},jk}, p_{\text{wf},jk})$, where m_{jk} is a pseudo-pressure, $p_{\text{t},jk}$ is tubinghead pressure, $p_{\text{wf},jk}$ is bottomhole pressure and q_{jk} is gas rate. The reservoir model is constructed by an I-dimensional spatial discretization

of a parabolic partial differential equation, while $k \in \mathcal{K}$ is the discrete time index. Each well has to be operated above the respective line pressure, $p_{\text{line}}^{\text{low}}$ and $p_{\text{line}}^{\text{high}}$, depending on the line scheduling modeled by the binary y_{jk}^2; if $y_{jk}^2 = 1$ the gas flow is routed to the low-pressure line, and if $y_{jk}^2 = 0$, the gas flow is bypassed the compressor, cf. Fig. 5.5. Moreover, to avoid so-called liquid loading, the gas rate q_{jk} has to be above a lower bound q_{gc} (Turner et al., 1969), or the well must be shut-in (Knudsen and Foss, 2013). We model this on/off state of the wells with a binary y_{jk}^1. See Knudsen et al. (2014a) for further details on the model formulation. Summarized, the shale-well scheduling problem is modeled by the condensed nonconvex MINLP

$$\max G \sum_{k\in\mathcal{K}} \sum_{j\in\mathcal{J}} (1 - C^{\text{C}} y_{jk}^2) q_{jk} \Delta k, \tag{5.22a}$$

s.t.

$$\sum_{j\in\mathcal{J}} q_{jk} y_{jk}^2 \le q_{\text{tot}}^{\text{up}}, \qquad \forall k \in \mathcal{K}, \tag{5.22b}$$

$$\sum_{j\in\mathcal{J}} q_{jk} y_{jk}^2 \ge q_{\text{tot}}^{\text{low}}, \qquad \forall k \in \mathcal{K}, \tag{5.22c}$$

$$F_j\left(m_{jk+1}, q_{jk+1}, p_{\text{t},jk}, p_{\text{wf},jk}\right) = 0, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K}^{\text{m}}, \tag{5.22d}$$

$$m_{j0} = m_j^{\text{init}}, \qquad \forall j \in \mathcal{J}, \tag{5.22e}$$

$$p_{\text{t},jk} \ge p_{\text{line}}^{\text{low}} y_{jk}^2 + (1 - y_{jk}^2) p_{\text{line}}^{\text{high}}, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K}, \tag{5.22f}$$

$$q_{jk} \ge y_{jk}^2 y_{jk}^1 q_{\text{gc}}^{\text{low}} + (1 - y_{jk}^2) y_{jk}^1 q_{\text{gc}}^{\text{high}}, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K}, \tag{5.22g}$$

$$y_{jk}^1, y_{jk}^2 \in \{0, 1\}, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K},$$

$$q_{jk}, p_{\text{t},jk}, p_{\text{wf},jk} \in \mathbb{R}, \;\; m_{jk} \in \mathbb{R}^I, \qquad \forall j \in \mathcal{J}, k \in \mathcal{K}.$$

5.4.1 Results

To asses the performance of the proposed OFP in Algorithm 3 on the shale-well scheduling problem, we construct an example of (5.22) consisting of $|\mathcal{J}| = 6$ wells, and a two-month planning horizon $\mathcal{K}$ with a fixed $\Delta k = 2$-day time step. The resulting nonconvex MINLP consists of 3276 constraints, 1620 continuous variables and 360 binary variables. We include a set of test problems with totally 15 instances of the problem (5.22), in which each problem has a different initial condition m_j^{init}. The computational environment and parameter values for the OFP were set to those given in Section 5.3, except for the user-defined weight u_2 for the original objective function which is set to $u_2 := 1$. That is, we assume the users desires an equal emphasis on solution quality and quickly obtaining a feasible solution. We impose a maximum CPU time of 7200 seconds.

There are a few instances from the set of test problems used for the computational study in Section 5.3 which have similar dimensions to the shale-well scheduling problem. This includes the problems: RSyn0810M04H, RSyn0815M04H, RSyn0830M03H, RSyn0840M03H and Syn40M04H. The dimensions of these problems can be found in Appendix 5.A.

Table 5.3 compares average results from the 15 instances of the shale-well scheduling problem when applying the proposed OFP in Algorithm 3, the FP heuristics of Bonami and Gonçalves (2010), and the BB method in `BONMIN`. It can be seen that the OFP turns out to be superior to the FP both with respect to solution time and solution quality when applied to (5.22). The average solution time of the OFP is smaller by a factor of 2.64 compared to the FP, and the average optimality gap of the OFP is approximately 12 times smaller than that of the FP. This is in contrast to the results of the OFP from Section 5.3, where the OFP was observed to improve the solution quality compared to the FP for a majority of the test problems, while requiring on average more time to compute a feasible solution. This discrepancy in performance is likely to be a result of an intrinsic property of the problem formulation (5.22), in which the feasible space makes it possible for the OFP to quickly compute feasible solutions. The original FP does also suffer from excessive stalling on these instances. To escape integer infeasible points causing stalling of the algorithm, the structure of the optimization problem must be changed, either by the feasible region or by the objective function (Boland et al., 2012). The changing weight α_i in the OFP objective may as such help reducing stalling, by serving as an additional escape mechanism to the variable flipping. The FP variant applied for comparison in this paper resembles the FP-3 variant for nonconvex MINLPs described in D'Ambrosio et al. (2012), which is reported to be very fast but often hits the time limit due to a high number of iterations. As this may indicate frequent stalling, a further study on nonconvex MINLPs could assess if the proposed OFP circumvents some of these issues and hence is a viable FP-approach also for nonconvex MINLPs.

Table 5.3: Average objective values and CPU times for the OFP, FP and the first solution found by the branch-and-bound method on 15 test instances of the shale-well scheduling problem (5.22)

	OFP	FP	BB
GM obj [10^6 \$][a]	1.12	0.70	1.14
GM gap [%]	3.21	39.32	2.11
GM time [s]	6.64	17.5	252.4

[a] : geometric mean.

The performance of the OFP with regards to objective value is in the first row of Table 5.3 seen to be nearly as good as that of the first solution found by the branch-and-bound method. While the average optimality gap of BB is smaller by a factor of 0.65 compared with the OFP, the average solution time of BB is larger by a factor of 38. The significant contrasts in performance of the various methods applied to (5.22) is clearly visible in the performance profiles for the relative optimality gap and solution time shown in Fig. 5.6 and 5.7, respectively. The large relative distance between the individual performance profiles for the different methods indicates clearly which solver is superior for each of the two selected performance metrics.

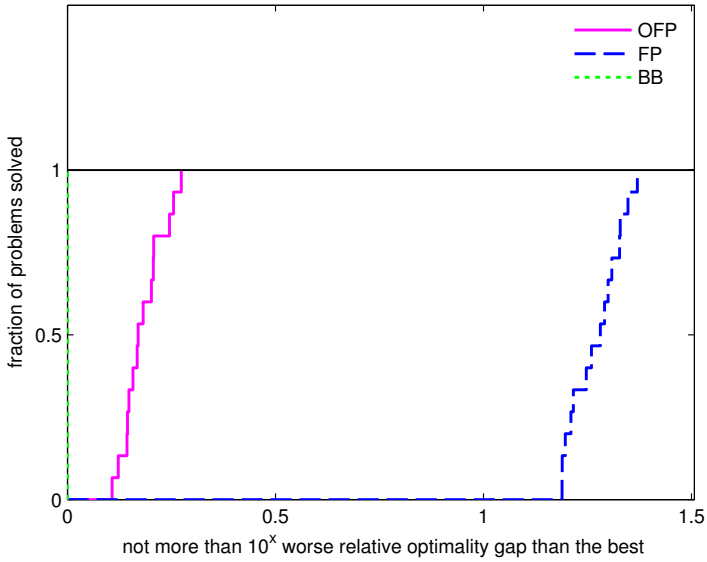


Fig. 5.6: Performance profile of the optimality gap (5.20) with the respect to the best known solution, comparing the results of OFP, FP and the first solution found by the branch-and-bound method on 15 test instances of the shale well scheduling problem (5.22).

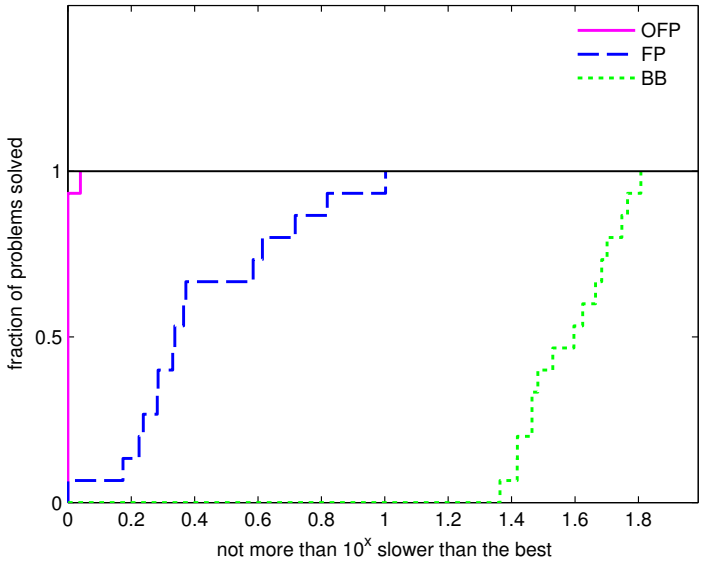


Fig. 5.7: Performance profile of the CPU time required to find a feasible solution, comparing the results of OFP, FP and the first solution found by the branch-and-bound method on 15 test instances of the shale well scheduling problem (5.22).

5.5 Concluding remarks

In this paper we have outlined the development of an Objective Feasibility Pump for convex MINLPs. The computational results show that on average the proposed OFP heuristic finds better solutions than the original Feasibility Pump, while the improvement is not as significant as what was observed for the OFP for MILPs. Although the computation time for the OFP increases compared to the original FP, the experiments with varying user-defined weights u_i demonstrates how this trade-off between solution quality and computation time can be adjusted to the specific application and requirements of the user. The results from the shale-well example demonstrate as such that the OFP can be very effective for finding high quality solutions in short computation times on certain problems.

Although we have limited the scope of the algorithmic development to convex MINLPs, we believe that the proposed OFP heuristic may be applied successfully to nonconvex MINLPs. This is supported by the promising results from the nonconvex shale-well scheduling problem in Section 5.4, which merit a further investigation into such an extension of the algorithm. Finally, we remark that improved handling of cycling in the OFP algorithm may be obtained by including for instance a tabu-list as suggested in D'Ambrosio et al. (2012). It may also be possible to derive a direct updating strategy for the weight parameters in the objective function of the OPF as a means of mitigating stalling, in addition to the two objectives of integer feasibility and solution quality. A comprehensive numerical and theoretical analysis of the effect of objective weighting on stalling is, however, still a research topic.

5.A Detailed computational results

Table 5.4: Results of the 146 convex MINLP test problems from applying the OFP, FP and the first solution found by the branch-and-bound method in `BONMIN` with default settings . The '#Bin', '#Cont' and '#Const' columns display the number of binary variables, continuous variables and constraints, respectively, for each problem.

				OFP		FP		BB	
Instance	#Bin	#Cont	#Const	Gap [%]	Time [s]	Gap [%]	Time [s]	Gap [%]	Time [s]
BatchS101006M	129	149	1019	**3.51**	1.53	6.77	**0.05**	0.00	6.15
BatchS121208M	203	203	1511	**1.62**	**2.90**	-	-	0.00	14.70
BatchS151208M	203	242	1781	**1.23**	4.03	5.06	**0.11**	0.03	17.29
BatchS201210M	251	307	2327	**4.03**	7.46	42.62	**1.40**	0.01	28.26
CLay0203H	18	72	132	35.04	**0.17**	**0.33**	0.28	31.90	0.54
CLay0203M	18	12	54	**35.04**	0.16	**35.04**	**0.02**	0.39	0.24
CLay0204H	32	132	234	**744.43**	**0.17**	1034.71	0.73	990.76	1.64
CLay0204M	32	20	90	**40.57**	0.07	1007.93	**0.06**	743.60	0.40
CLay0205H	50	210	365	-	-	-	-	-	-
CLay0205M	50	30	135	216.48	**0.09**	**202.03**	0.12	7.62	0.80
CLay0303H	21	78	150	**56.40**	**0.69**	56.50	1.72	105.62	0.85
CLay0303M	21	12	66	**3.45**	**0.54**	77.31	1.23	56.40	0.26
CLay0304H	36	140	258	**53.91**	**0.75**	-	-	44.94	2.58
CLay0304M	36	20	106	**46.76**	**0.31**	95.08	0.88	52.78	0.45
CLay0305H	55	220	395	**183.34**	**15.76**	-	-	7.62	5.96
CLay0305M	55	30	155	**37.62**	0.04	**37.62**	**0.01**	7.62	1.05
FLay02H	4	42	51	**0.00**	0.20	**0.00**	**0.01**	0.00	0.03
FLay02M	4	10	11	**0.00**	0.13	**0.00**	**0.01**	0.00	0.02
FLay03H	12	110	144	**0.00**	0.40	**0.00**	**0.01**	0.00	0.28
FLay03M	12	14	24	**0.00**	0.18	0.00	**0.01**	0.00	0.10
FLay04H	24	210	282	**1.07**	0.73	**1.07**	**0.02**	0.02	0.86
FLay04M	24	18	42	**0.00**	0.23	1.07	**0.01**	24.41	0.25

Table 5.4 – (continued)

				OFP		FP		BB	
Instance	#Bin	#Cont	#Const	Gap [%]	Time [s]	Gap [%]	Time [s]	Gap [%]	Time [s]
FLay05H	40	342	465	30.82	1.47	**0.00**	**0.04**	0.00	2.25
FLay05M	40	22	65	**0.00**	0.40	0.00	**0.01**	0.00	0.42
FLay06H	60	506	693	**44.59**	2.37	47.62	**0.04**	11.55	4.91
FLay06M	60	26	93	20.71	0.49	**3.77**	**0.01**	3.77	0.77
RSyn0805H	296	424	1886	**29.27**	4.57	32.07	**0.13**	2.01	51.56
RSyn0805M	69	101	286	**15.55**	0.82	18.03	**0.02**	0.68	2.88
RSyn0805M02H	148	552	1045	**1.66**	1.35	25.16	**0.13**	0.02	12.76
RSyn0805M02M	148	212	769	**9.50**	2.27	36.79	**0.04**	3.77	14.03
RSyn0805M03H	222	828	1698	**1.69**	1.83	18.54	**0.37**	0.00	28.16
RSyn0805M03M	222	318	1284	-	-	**25.67**	**0.08**	1.81	29.48
RSyn0805M04H	296	1104	2438	**0.00**	2.26	9.47	**0.46**	0.15	41.77
RSyn0805M04M	296	424	1886	**29.37**	4.68	32.16	**0.13**	2.14	51.52
RSyn0810H	336	484	2140	182.69	5.46	**167.91**	**0.19**	266.05	73.28
RSyn0810M	74	111	312	**16.29**	0.87	17.28	**0.03**	4.08	3.34
RSyn0810M02H	168	622	1188	**2.22**	1.50	31.91	**0.20**	0.17	16.22
RSyn0810M02M	168	242	866	**58.11**	2.54	73.07	**0.06**	38.00	15.64
RSyn0810M03H	252	933	1935	**3.94**	2.25	17.84	**0.36**	0.09	42.90
RSyn0810M03M	252	363	1452	**30.58**	3.58	47.08	**0.12**	29.38	40.88
RSyn0810M04H	336	1244	2784	**2.50**	3.72	22.80	**0.59**	0.11	62.49
RSyn0810M04M	336	484	2140	**45.03**	5.49	47.90	**0.19**	28.82	72.88
RSyn0815H	376	564	2430	**58.52**	6.38	89.68	**0.33**	46.63	177.52
RSyn0815M	79	126	347	**13.35**	0.97	22.57	**0.05**	0.47	4.67
RSyn0815M02H	188	710	1361	**1.93**	1.71	62.60	**0.28**	0.05	18.00
RSyn0815M02M	188	282	981	**54.20**	2.88	100.20	**0.08**	31.24	33.34
RSyn0815M03H	282	1065	2217	**4.94**	2.70	38.33	**0.37**	0.01	49.57
RSyn0815M03M	282	423	1647	**44.90**	4.27	81.20	**0.13**	28.11	72.37
RSyn0815M04H	376	1420	3190	**2.30**	4.17	42.95	**0.77**	0.06	73.30
RSyn0815M04M	376	564	2430	**58.52**	6.27	89.68	**0.32**	46.63	179.70
RSyn0820H	416	604	2676	**3.38**	6.83	17.71	**0.35**	99.23	150.75
RSyn0820M	84	131	371	**19.61**	1.04	27.87	**0.03**	0.37	4.12
RSyn0820M02H	208	770	1500	**2.80**	1.97	70.37	**0.21**	0.00	20.43
RSyn0820M02M	208	302	1074	**78.24**	3.35	94.14	**0.13**	62.20	25.70
RSyn0820M03H	312	1155	2448	**4.67**	2.79	43.29	**0.46**	0.07	59.95
RSyn0820M03M	312	453	1809	**61.12**	4.62	78.92	**0.14**	57.57	70.35
RSyn0820M04H	416	1540	3528	**0.30**	4.30	52.65	**0.89**	0.00	96.07
RSyn0820M04M	416	604	2676	**83.50**	6.70	85.95	**0.36**	65.98	150.84
RSyn0830H	496	744	3192	95.22	11.20	**91.06**	**0.30**	71.68	212.43
RSyn0830M	94	156	425	**69.73**	1.33	77.75	**0.05**	47.57	6.31
RSyn0830M02H	248	924	1794	**0.96**	1.63	58.93	**0.23**	0.31	20.69
RSyn0830M02M	248	372	1272	106.13	4.08	**96.87**	**0.13**	75.25	42.73
RSyn0830M03H	372	1386	2934	**0.40**	3.23	46.68	**0.50**	0.04	49.93
RSyn0830M03M	372	558	2151	**76.27**	6.76	93.21	**0.26**	72.55	109.56
RSyn0830M04H	496	1848	4236	**0.47**	4.91	38.32	**0.57**	0.02	90.58
RSyn0830M04M	496	744	3192	95.22	11.19	**91.06**	**0.30**	71.68	211.62
RSyn0840H	576	864	3728	**16.33**	11.89	114.02	**0.65**	571.39	263.96
RSyn0840M	104	176	484	**107.98**	1.47	109.29	**0.05**	59.70	8.00
RSyn0840M02H	288	1072	2106	**0.60**	1.71	58.49	**0.24**	0.36	20.94
RSyn0840M02M	288	432	1480	101.92	4.09	**98.90**	**0.13**	86.04	53.48
RSyn0840M03H	432	1608	3447	**3.96**	3.86	17.20	**0.44**	0.09	66.89
RSyn0840M03M	432	648	2508	73.56	7.20	**73.31**	**0.78**	61.36	126.86
RSyn0840M04H	576	2144	4980	**0.41**	4.94	37.93	**0.65**	0.08	102.92
RSyn0840M04M	576	864	3728	96.37	11.96	**93.32**	**0.65**	79.04	264.84
SLay04H	24	116	174	**359.77**	0.05	**359.77**	**0.02**	0.00	0.43
SLay04M	24	20	54	**452.32**	0.02	**452.32**	**0.01**	0.00	0.24
SLay05H	40	190	290	**303.25**	0.05	**303.25**	**0.02**	0.00	0.97
SLay05M	40	30	90	**383.88**	0.02	**383.88**	**0.01**	0.00	0.50
SLay06H	60	282	435	**291.96**	0.06	**291.96**	**0.04**	0.53	2.12
SLay06M	60	42	135	**277.48**	0.03	333.62	**0.01**	0.53	0.95
SLay07H	84	392	609	**145.67**	0.11	332.53	**0.05**	0.78	3.36
SLay07M	84	56	189	**338.27**	0.07	407.84	**0.02**	0.78	1.43
SLay08H	112	520	812	**663.43**	0.33	669.50	**0.07**	0.00	6.03
SLay08M	112	72	252	**328.56**	0.07	454.93	**0.02**	0.00	2.45
SLay09H	144	666	1044	**434.70**	0.36	553.15	**0.08**	0.59	11.39
SLay09M	144	90	324	**257.30**	0.16	434.38	**0.02**	0.00	4.55
SLay10H	180	830	1305	**470.96**	0.51	488.24	**0.10**	2.05	16.58
SLay10M	180	110	405	**274.86**	0.14	363.22	**0.02**	2.38	5.90
Syn05H	5	37	58	**0.03**	**0.01**	**0.03**	**0.01**	0.00	0.03
Syn05M	5	15	28	0.75	0.13	**0.00**	**0.01**	0.00	0.04
Syn05M02H	20	84	151	**0.00**	**0.01**	**0.00**	**0.01**	0.00	0.08
Syn05M02M	20	40	101	**0.20**	0.22	1.86	**0.02**	0.00	0.26
Syn05M03H	30	126	249	**0.00**	0.03	**0.00**	**0.02**	0.00	0.28
Syn05M03M	30	60	174	**0.15**	0.28	1.52	**0.02**	0.10	0.46
Syn05M04H	40	168	362	**0.00**	**0.03**	**0.00**	**0.03**	0.00	0.46

Table 5.4 – (continued)

Instance	#Bin	#Cont	#Const	OFP Gap [%]	OFP Time [s]	FP Gap [%]	FP Time [s]	BB Gap [%]	BB Time [s]
Syn05M04M	40	80	262	**0.11**	0.43	5.06	**0.02**	1.19	0.73
Syn10H	10	67	112	**0.00**	0.02	**0.00**	**0.01**	0.00	0.04
Syn10M	10	25	54	**1.66**	0.18	2.21	**0.02**	0.00	0.11
Syn10M02H	40	154	294	**0.00**	0.08	37.07	**0.03**	30.64	0.20
Syn10M02M	40	70	198	**0.34**	0.35	58.21	**0.02**	0.90	0.82
Syn10M03H	60	231	486	**0.00**	0.12	31.41	**0.04**	27.01	0.41
Syn10M03M	60	105	342	**1.75**	0.68	31.41	**0.03**	0.79	1.79
Syn10M04H	80	308	708	**0.00**	0.25	35.57	**0.09**	30.55	0.68
Syn10M04M	80	140	516	**1.44**	0.81	35.57	**0.03**	1.29	3.31
Syn15H	15	106	181	**0.00**	0.05	**0.00**	**0.01**	0.00	0.04
Syn15M	15	40	89	5.28	0.24	**4.93**	**0.01**	1.99	0.17
Syn15M02H	60	242	467	**0.00**	0.12	54.95	**0.04**	0.00	0.16
Syn15M02M	60	110	313	**1.77**	0.53	94.14	**0.02**	1.38	1.79
Syn15M03H	90	363	768	**0.00**	0.18	46.54	**0.06**	0.00	1.06
Syn15M03M	90	165	537	**1.30**	1.04	92.59	**0.03**	1.01	4.04
Syn15M04H	120	484	1114	**0.00**	0.38	44.19	**0.14**	41.30	0.62
Syn15M04M	120	220	806	**0.97**	1.54	91.73	**0.05**	4.07	7.30
Syn20H	20	131	233	**0.00**	0.09	**0.00**	**0.02**	24.20	0.10
Syn20M	20	45	113	**54.63**	**0.26**	-	-	0.00	0.33
Syn20M02H	80	302	606	**0.00**	0.25	67.97	**0.06**	0.00	1.10
Syn20M02M	80	130	406	**2.72**	0.77	63.66	**0.03**	4.42	3.19
Syn20M03H	120	453	999	**0.00**	0.42	66.42	**0.09**	0.00	2.73
Syn20M03M	120	195	699	**1.32**	1.79	61.73	**0.04**	2.04	7.91
Syn20M04H	160	604	1452	**0.00**	0.54	44.33	**0.18**	0.00	5.33
Syn20M04M	160	260	1052	147.21	2.22	**0.00**	**0.06**	148.83	16.27
Syn30H	30	198	345	**2.99**	0.30	19.03	**0.02**	0.00	0.38
Syn30M	30	70	167	21.55	0.43	**13.25**	**0.02**	0.00	0.74
Syn30M02H	120	456	900	**0.36**	0.62	3.08	**0.07**	0.00	1.51
Syn30M02M	120	200	604	40.59	1.15	**33.93**	**0.04**	39.05	6.03
Syn30M03H	180	684	1485	**0.47**	1.23	2.34	**0.14**	0.26	4.19
Syn30M03M	180	300	1041	53.63	1.90	**49.38**	**0.06**	57.11	16.79
Syn30M04H	240	912	2160	**0.00**	2.05	5.84	**0.14**	1.41	12.86
Syn30M04M	240	400	1568	65.55	3.98	**61.94**	**0.09**	72.34	34.35
Syn40H	40	262	466	20.36	0.48	**9.62**	**0.03**	1.35	1.20
Syn40M	40	90	226	114.27	0.54	**65.80**	**0.01**	30.70	1.28
Syn40M02H	160	604	1212	2.32	0.71	**0.44**	**0.06**	0.00	1.79
Syn40M02M	160	260	812	114.43	1.66	**61.90**	**0.04**	67.45	10.14
Syn40M03H	240	906	1998	**0.32**	1.42	19.36	**0.12**	0.40	10.42
Syn40M03M	240	390	1398	**85.56**	3.14	86.94	**0.08**	87.57	34.92
Syn40M04H	320	1208	2904	**0.26**	3.22	8.21	**0.15**	0.01	32.50
Syn40M04M	320	520	2104	**77.80**	3.77	77.83	**0.11**	80.46	56.05
Water0202	7	106704	107209	**1874.15**	10.01	2448.37	**3.93**	1625.22	32.01
Water0202R	7	188	283	**0.00**	1.45	5085.12	**0.16**	2647.83	0.60
Water0303	14	107208	108217	-	-	**1455.45**	**4.64**	1455.45	68.50
Water0303R	14	370	556	**26.29**	15.63	1857.62	**11.25**	819.46	9.54
fo7	42	72	211	**36.35**	**0.10**	88.31	1.27	38.03	24.21
fo7-2	42	72	211	**93.11**	**0.40**	127.70	0.42	72.10	34.03
fo8	56	90	273	**121.70**	**0.13**	130.10	0.26	73.80	595.36
fo9	72	110	343	**168.77**	**0.14**	-	-	191.58	820.62
o7	42	72	211	75.53	1.23	**75.31**	**0.87**	55.94	359.39
o7-2	42	72	211	**85.37**	**0.13**	96.64	0.83	78.30	131.35

Table 5.5: Results from 15 test problems of the shale well model described in Section 5.4

Instance	OFP Gap [%]	OFP Time [s]	FP Gap [%]	FP Time [s]	BB Gap [%]	BB Time [s]
SW1	**2.35**	**8.12**	39.13	42.36	1.67	236.62
SW2	**3.06**	**8.19**	38.71	17.81	2.39	396.10
SW3	**2.58**	**5.70**	39.43	12.21	1.85	263.41
SW4	**3.48**	**4.57**	35.36	18.79	2.29	266.52
SW5	**2.40**	**4.48**	36.33	29.50	1.63	250.72
SW6	**4.03**	**5.71**	42.00	9.88	2.15	240.41
SW7	**3.15**	**8.83**	41.90	13.18	2.19	231.15

SW8	**3.88**	9.03	38.38	**8.24**	2.44	215.74
SW9	**4.05**	**6.63**	41.81	12.77	2.30	224.65
SW10	**3.00**	**5.67**	39.63	13.15	1.86	223.93
SW11	**3.26**	**4.58**	38.04	10.80	2.46	294.18
SW12	**4.08**	**8.06**	40.06	30.96	2.27	245.08
SW13	**3.28**	**4.92**	36.43	9.42	2.21	247.40
SW14	**3.33**	**11.86**	41.19	119.23	2.07	273.80
SW15	**2.89**	**7.59**	42.26	12.74	2.08	221.11

Chapter 6

Concluding Remarks and Recommendations for Further Work

Shale gas has emerged as one of the main resources of natural gas. Its impact on energy generation, electric power production and as feedstock in chemical process systems already is and will continue to be significant. The focus of this thesis has been to derive and analyze models and optimization algorithms for efficient shut-in scheduling of shale-gas systems. Several system structures and problem definitions have been considered and analyzed, leading to the development of suitable proxy models, MIP formulations, heuristics and Lagrangian relaxation based decomposition schemes.

Proxy models and parameter estimation
Two highly related, while still qualitatively different shale-gas reservoir proxy models are developed in the thesis. Although the cylindrical SRV-based model applied in Chapter 2 and 3 and the linear-flow based, explicit fracture model used in Chapter 4 and 5 to a large extent may be tuned and used interchangeably, there are principles behind the models that make each of them favorable for different applications. The cylindrical SRV-based model with a crushed-rock region may better capture depletion of the entire fracture system further into the life-time of a well, as the three-region model better reflects the total compound conductivity in fractured shale-gas reservoirs. Reservoirs assumed to contain large networks of natural fractures may also benefit from this interpretation and approximation of the SRV, compared to explicitly modeling single planar fractures. Consequently, the cylindrical SRV-based model may be the most suitable for scheduling of shut-ins to prevent liquid loading during late-life times of wells. The slab proxy model based on modeling a quarter section of a planer fracture and matrix block, captures the frequently observed linear flow in shale-gas reservoirs. This model is as such more accurate during the initial gas flow from shale-gas wells, and may also be more accurate for predicting pseudo steady-state gas rates. The slab-based proxy model will thus generally cover a wider life-span of a shale-gas well, and may require fewer updates compared to the cylindrical SRV-based model in order to maintain accept-

able accuracy. Both models, when used in a closed-loop production optimization scheme, may benefit from a more comprehensive updating strategy compared to the one proposed in Chapter 3. It is also worth noting that the nonlinear tubing model applied together with the slab model assumes remote control of tubinghead pressure, while the simplified, nonsmooth well and wellbore model applied in Chapter 2 and 3 assumes implicitly that a closed-loop pressure control is installed at the wellhead valves.

An intrinsic difficulty in model tuning by weighted nonlinear least squares is appropriately selecting weights. Although the cross-validations in Chapter 2 demonstrate that the selected model structure and weight selection in the WNLS tuning renders a good model fit, the difficult lies in making this approach generalizeable to models with different data sets and properties. The correction term for cumulative error improves the model fit for longer prediction horizons, however, once more presenting challenges in terms of relative weight-selection and appropriate scaling when used in combination with the sum of squares for gas-rate prediction errors. The prediction-error filtering approach in Chapter 3 gives a more general approach for weight selection in tuning of the proxy models. With the governing model being nonlinear, the approach will of course only allow a heuristic interpretation of weighting of frequencies in the models. Nevertheless, it provides an enhanced procedure for emphasizing different modes in the proxy model compared to the ad-hoc weighting applied in Chapter 2. Note that this approach is also analyzed in Appendix A for the Cartesian slab model applied in Chapter 4 and 5.

Formulations of optimization problems

The approach of using reference-rate tracking introduced in Chapter 2 is extended to *multi-pad* systems in Chapter 3, applying a systematic approach based on GDP modeling to derive an improved MILP model compared to the one used in the single-pad study. Note that the MILP reformulation applied for this embedded GDP turned out to be most efficient using a combined convex hull (CH) and big-M reformulation, as the complete CH reformulation gave little improvement in tightness of the formulation while significantly increasing the problem size. In comparison, the GDP model in Chapter 4 was reformulated using the convex hull only, with disaggregated variables utilized directly in the description of line flow-rates. See Vecchietti et al. (2003) and Grossmann and Trespalacios (2013) for discussion on the choice and combination of CH and big-M reformulations of GDPs. The approach presented in Chapter 3 allows for scheduling field-wide gas production with respect to a given demand. Compared to the approach in Chapter 2, it also computes optimal pad reference rates with respect to the total demand and the capacity for each pad. By tracking each of these sub reference-rates sufficiently well, the total rate tracks the field reference rate, while also stabilizing the rates in gathering pipelines from the well pads to the compression unit. An assumption of the model is that each pad is connected with a separate pipeline to the compression unit, which serves as boundary for the system considered. A further study of the proposed approach could hence include more complex surface pipeline networks with intermediate pipeline connections. One approach to circumvent the increased complexity this type of networks introduces, may be to shift the boundary point of the system considered to the intermediate pipeline connection, and subsequently

decompose the system in several levels.

The Lagrangian relaxation based decomposition schemes
Two decomposition schemes based on Lagrangian relaxation have been constructed and implemented to compute field-wide shut-in schedules for large-scale shale-gas systems. The first scheme assumes a multi-pad structure, and decompose the primal problem into one subproblem per pad, while the second scheme decomposes the primal problem directly into one subproblem per well. Both schemes demonstrate good numerical efficiency as implemented, although parallelization of subproblems would clearly reduce total computation significantly. Still, combining the two approaches is deemed possible, and may improve the computational efficiency. For the first LR scheme, applied in Chapter 3, single-well pads may be directly included, however, with the definition of a pad-reference becoming ambiguous. The computational efficiency of this LR-based decomposition scheme is mainly limited by the time spent on solving each MILP subproblem. Several possibilities for improved efficiency present themselves, with a bi-level decomposition scheme probably being the most suitable approach. Possible approaches for decomposing the pad subproblems include using temporal decomposition, following the lines of Beccuti et al. (2004); Terrazas-Moreno et al. (2011); Kopanos and Pistikopoulos (2014), or spatially decomposing the problem as in Chapter 4, although obtaining a decomposition scheme with strong Lagrangian bounds for ℓ_1- or $\ell_\infty-$tracking objectives is more involved.

The choice of stabilization device in a bundle method for solving the Lagrangian dual, comparing the trust-region approach used in Chapter 3 and the proximal bundle method applied in Chapter 4, the most appropriate choice depends on the application. The trust-region approach works more as a switch, defining an acceptable region for trusting the CP-model, and and as such for accepting or discarding iterates to update the multipliers, while the proximal bundle method adds a strictly convex term with associated benefits, e.g. second-order information, see e.g. Lemaréchal (2001). From the computational experience applying these two closely related, yet different approaches, the proximal bundle method appears to work somewhat better on problems with a large number of dual variables. The effort required in implementing the trust-region approach is, however, less compared to the proximal bundle method, in that a choosing and implementing a good updating strategy for the penalty parameter γ^n is not straight forward, while at the same time being important for the numerical efficiency.

Lagrangian- and MIP heuristics developed in the thesis
This thesis describes the development of several mixed integer programming heuristics. MIP heuristics are used in many forms within the solution process of MIP problems, including constructing and improving solutions within an branch-and-bound framework, as standalone algorithms and for repairing nearly feasible solutions, for instance in Lagrangian relaxation schemes. Chapter 3 and 4 both contain novel Lagrangian heuristics that are shown to efficiently generate primal feasible solutions from the Lagrangian. The heuristic used in Chapter 3 resorts to an integration of the classical Lagrangian heuristic based on fixing binary variables, together with the Local Branching heuristic used for improving the well sched-

ule generated from the binary-fixing. A similar combination of a construction and improvement heuristic is applied in Knudsen et al. (2013), however, with a more numerically demanding construction heuristic due to lack of applicability of binary fixing. A possible extension of the latter Lagrangian heuristic is to integrate the local branching constraint directly in a construction heuristic, starting from the solution of the Lagrangian, see Fischetti and Lodi (2008). Chapter 4 contains a novel Lagrangian heuristic inspired by the FP and RINS heuristics. This heuristic overcomes the infeasibility of classic binary-fixing heuristics encountered in many applications, obtained by constructing a constraint feasible point that is closest (in the ℓ_1 sense) to the solution of the Lagrangian. The closest point will generally not be integer feasible, thereby initiating a fixing phase with subsequent branching. By dropping the penalty term in the objective of (4.28), the developed heuristic renders a generic structure, which can be utilized to construct primal feasible solutions for a wide range of MIP problems solved with LR schemes, though up to a certain problem size as discussed in Section 4.6.3. A possible extension of the proposed heuristic may be to explore a convex combination of the ℓ_1 feasibility term and the original objective, following lines of the OFP approach described in Chapter 5. This approach was, however, briefly tested and found less efficient than the selected approach with a penalty term only; sufficient scaling and a careful updating strategy for the (convex) weights must be applied.

The Objective Feasibility Pump heuristic for convex MINLPs presented in Chapter 5 seeks to explore the computationally most efficient FP variant for convex MINLPs (Bonami and Gonçalves, 2010) with a strategy for improving the quality of solutions obtained. Obtaining a balanced combination of efficiently generating feasible solutions and retaining solution quality is significantly more involved for nonlinear than linear MIPs. The structured development of an OFP for convex MINLPs demonstrates that an FP approach balancing solution quality and effectiveness is also possible for MINLPs. Developing a completely generic OFP for convex MINLP is, however, observed to be challenging, while the case study on the shale-gas problem illustrates that the approach may be very efficient when tailored to a particular application. Using a heuristic as the OFP may as such be beneficial in DSTs where short computation time is crucial for the applicability of the tool. In light of the shale-gas case being nonconvex, a topic to be further explored is the use of similar OFP variants on broader classes of nonconvex MINLPs.

A significant contribution in this thesis is the development of algorithms for solving large-scale scheduling problems containing *dynamic* models. Scheduling problems tend to consider only static, possibly piecewise linear or polynomial models, as these are often proper and sufficiently accurate models for the applications. Shale-gas wells with the applications considered are, however, too dynamic to use static models. Efficiently combining reduced-order proxy models, appropriate tuning and constructive decomposition schemes has therefore been crucial to be able to solve these associated large-scale mixed integer problems. The dynamic scheduling problems were further integrated in a moving horizon type scheme in Chapter 4. Moving horizon or EMPC applications are more common on a lower level in the classical process hierarchy, see e.g. Skogestad and Postlethwaite (2005), than on a scheduling or RTO level. In the future, however, with more advanced problems to be solved and increased computational power, more applications of dynamic

scheduling or RTO may arise.

Shale gas in the context of natural-gas based electric-power production Chapter 4 presents a novel scheme for shifting storage of gas for meeting varying demands, from LDCs to the gas producers themselves, obtained through structured shut-in scheduling. The approach is formulated from the gas producers' perspective, and shown to potentially increase profit for well operators, while ensuring a firm gas supply for NGPP operators. Circumventing external gas storage operated by LDCs leads to an *energy saving* by avoiding fuel consumption associated with injection and withdrawal of gas into underground storage, as well as preventing loss of injected gas. A challenge for the scheme may be to ensure sufficient transmission pipeline capacity. Pipeline capacity is currently squeezed in some areas of the US due to large volumes of produced shale gas, while expansions are in progress (ISO New England, 2008). The average gas production from applying the proposed scheduling scheme in a field with many wells, would be close to the production obtained with each well producing at their maximum and transmitting the gas to LDCs. The supply may, however, increase during periods of peak demand. These peak demands are today mainly covered by LDCs, hence requiring sufficient pipeline capacity between their facilities and end-users. The proposed scheme would require a similar capacity further upstream of the natural-gas value chain, that is, closer to the producers.

Several extensions of the scheme proposed in Chapter 4 may be considered, for instance integrating more tightly the demand side with shale-gas production, e.g. as variations of the UC problem. See Liu et al. (2011) for possible directions. Further analysis on the efficiency of shut-ins of shale-gas wells may also be conducted, using a more comprehensive SENSOR simulation model to study effects of gravity, capillary, and osmotic pressures, as well as a multi-phase well model, see e.g. Cheng (2012) and Fakcharoenphol et al. (2013). The LR based moving-horizon scheme may also be considered with dynamic programming for solving the subproblems, avoiding the use of SOS2 approximations for the tubing model, and approaches for reusing cutting-planes from one moving-horizon iteration to the next, thereby reducing the number of LR iterations required. A possible approach in this direction may be to explore the use of inexact bundle methods (Kiwiel, 2006).

The explosion in recovery of shale gas has unquestionably changed the global gas market, and had a great impact on the natural-gas value chain and industry, safety of supply and energy generation in the US. The future of shale-gas production, as well as its sustainability, depend on many factors, including advances in drilling, stimulation and recovery technology. Sustainable water recycling and management are also crucial for long-term viability of the industry, together with improvements in well design in order to minimize emissions related to drilling and completion. The future gas price will also impact the development of dry-gas fields. Shale-gas operators may have to seek tighter integration with end-users to retain profit, in which the approach proposed in Chapter 4 may be an important step. The future of natural-gas usage in electricity generation depends on the development in coal-plant retirement, the slow down in construction of nuclear plants (Macmillan et al., 2013) and development of renewable generation. Until we develop an energy resource that provides the same flexibility and versatility as natural gas and

hydro power, areas without access to hydro-plants will rely on natural gas to ensure acceptable generation security in conjunction with intermittent and stochastic renewable generation sources. Countries with a well developed natural-gas value chain and access to vast shale-gas resources, may hence have the possibility to *integrate* gas production and renewable generation in the effort towards low-carbon, reliable electricity generation.

Bibliography

Abou-Kassem, J. H., Ali, S. F., and Islam, M. R. (2006). *Petroleum Reservoir Simulation, A Basic Approach*. Gulf Publishing Company, Houston.

Achterberg, T. and Berthold, T. (2007). Improving the feasibility pump. *Discrete Optimization*, 4(1):77–86.

Al Ahmadi, H., Almarzooq, A., and Wattenbarger, R. (2010). Application of linear flow analysis to shale gas wells – field cases. In *SPE Unconventional Gas Conference*, Pittsburgh, Pennsylvania, USA. SPE 130370.

Al-Hussainy, R., Ramey, H. J., and Crawford, P. (1966). The flow of real gases through porous media. *Journal of Petroleum Technology*, 18(5):624–636.

Ambrose, R. J., Clarkson, C. R., Youngblood, J., Adams, R., Nguyen, P., Nobakht, M., and Biseda, B. (2011). Life-cycle decline curve estimation for tight/shale gas reservoirs. In *SPE Hydraulic Fracturing Technology Conference*, The Woodlands, Texas, USA. SPE 140519.

Amrit, R., Rawlings, J. B., and Angeli, D. (2011). Economic optimization using model predictive control with a terminal cost. *Annual Reviews in Control*, 35(2):178–186.

Anderson, D., Nobakht, M., Moghadam, S., and Mattar, L. (2010). Analysis of production data from fractured shale gas wells. In *SPE Unconventional Gas Conference*, Pittsburgh, Pennsylvania, USA. SPE 131787.

Arps, J. (1945). Analysis of decline curves. *Trans. AIME*, 160:228–247.

Arvesen, O., Medbø, V., Fleten, S.-E., Tomasgard, A., and Westgaard, S. (2013). Linepack storage valuation under price uncertainty. *Energy*, 52:155–164.

Avery, W., Brown, G. G., Rosenkranz, J. A., and Wood, R. K. (1992). Optimization of purchase, storage and transmission contracts for natural gas utilities. *Operations Research*, 40(3):446–462.

Avraam, M., Shah, N., and Pantelides, C. (1998). Modelling and optimisation of general hybrid systems in the continuous time domain. *Computers & Chemical Engineering*, 22(98):221–228.

Awoleke, O. O. and Lane, R. H. (2011). Analysis of data from the Barnett shale using conventional statistical and virtual intelligence techniques. *SPE Reservoir Evaluation & Engineering*, 14(5):544–556. SPE 127919.

Aziz, K. and Settari, A. (1979). *Petroleum Reservoir Simulation*. Applied Science Publishers Ltd, London.

Baena, D. and Castro, J. (2011). Using the analytic center in the feasibility pump. *Operations Research Letters*, 39(5):310–317.

Baihly, J., Altman, R., Malpani, R., and Luo, F. (2010). Shale gas production decline trend comparison over time and basins. In *SPE Annual Technical Conference and Exhibition*, Florence, Italy. SPE 135555.

Balakrishna, S. and Biegler, L. T. (1992). Targeting strategies for the synthesis and energy integration of nonisothermal reactor networks. *Industrial & Engineering Chemistry Research*, 31(9):2152–2164.

Balas, E. (1985). Disjunctive programming and a hierarchy of relaxations for discret optimization problems. *SIAM Journal on Algebraic and Discrete Methods*, 6(3):466–486.

Barton, P. I. and Lee, C. K. (2004). Design of process operations using hybrid dynamic optimization. *Computers & Chemical Engineering*, 28(6-7):955–969.

Baumrucker, B. and Biegler, L. (2009). MPEC strategies for optimization of a class of hybrid dynamic systems. *Journal of Process Control*, 19(8):1248–1256.

Beale, E. and Tomlin, J. (1970). Special facilities in a general mathematical programming system for non-convex problems using ordered sets of variables. In *Proc. of the 5th Int. Conf. on Operations Research*, pages 447–454.

Beccuti, A., Geyer, T., and Morari, M. (2004). Temporal Lagrangian decomposition of model predictive control for hybrid systems. In *IEEE Conference on Decision and Control*, pages 2509–2514.

Bello, R. O. and Wattenbarger, R. A. (2010). Modelling and analysis of shale gas production with a skin effect. *Journal of Canadian Petroleum Technology*, 49(12):37–48.

Belotti, P., Kirches, C., Leyffer, S., Linderoth, J., Luedtke, J., and Mahajan, A. (2013). Mixed-integer nonlinear optimization. *Acta Numerica*, 22:1–131.

Bemporad, A. and Morari, M. (1999). Control of systems integrating logic , dynamics , and constraints. *Automatica*, 35(3):407–427.

Bertacco, L., Fischetti, M., and Lodi, A. (2007). A feasibility pump heuristic for general mixed-integer problems. *Discrete Optimization*, 4(1):63–76.

Berthold, T. (2006). Primal Heuristics for Mixed Integer Programs. Master's thesis, TU Berlin, Berlin, Germany.

Berthold, T. (2013). Primal MINLP heuristics in a nutshell. Technical Report 13-42, Zuse-Informationstechnik Berlin (ZIB), Berlin, Germany.

Berthold, T. and Gleixner, A. M. (2013). Undercover: a primal MINLP heuristic exploring a largest sub-MIP. *Mathematical Programming*, 144(1-2):315–346.

Biegler, L. T. (2010). *Nonlinear Programming: Concepts, Algorithms, and Applications to Chemical Processes*. SIAM, Philadelphia, PA.

Bieker, H. P., Slupphaug, O., and Johansen, T. A. (2007). Real-time production optimization of oil and gas production systems: a technology survey. *SPE Production & Operations*, 22(4):382–391.

Bixby, R. and Rothberg, E. (2007). Progress in computational mixed integer programming – A look back from the other side of the tipping point. *Annals of Operations Research*, 149(1):37–41.

Boland, N. L., Eberhard, A. C., Engineer, F., and Tsoukalas, A. (2012). A new approach to the feasibility pump in mixed integer programming. *SIAM Journal on Optimization*, 22(3):831–861.

Bonami, P., Biegler, L. T., Conn, A. R., Cornuéjols, G., Grossmann, I. E., Laird, C. D., Lee, J., Lodi, A., Margot, F., Sawaya, N., and Wächter, A. (2008a). An algorithmic framework for convex mixed integer nonlinear programs. *Discrete Optimization*, 5(2):186–204.

Bonami, P., Cornuéjols, G., Lodi, A., and Margot, F. (2008b). A feasibility pump for mixed integer nonlinear programs. *Mathematical Programming*, 119(2):331–352.

Bonami, P. and Gonçalves, J. (2010). Heuristics for convex mixed integer nonlinear programs. *Computational Optimization and Applications*, 51(2):729–747.

Bonami, P. and Lee, J. (2007). *Bonmin user's manual*.

Borghetti, A., Frangioni, A., Lacalandra, F., and Nucci, C. (2003). Lagrangian heuristics based on disaggregated bundle methods for hydrothermal unit commitment. *IEEE Transactions on Power Systems*, 18(1):313–323.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press, New York.

Brooke, A., D.Kendrick, Meeraus, A., and Raman, R. (2011). *GAMS - A user's guide*.

Cagienard, R., Grieder, P., Kerrigan, E., and Morari, M. (2007). Move blocking strategies in receding horizon control. *Journal of Process Control*, 17(6):563–570.

Cardoso, M. A. and Durlofsky, L. J. (2010). Use of reduced-order modeling procedures for production optimization. *SPE Journal*, 15(2):426–435.

Carlson, E. S. and Mercer, J. C. (1991). Devonian shale gas production: mechanisms and simple models. *Journal of Petroleum Technology*, 43(4):476–482.

Chen, C. and Mangasarian, O. L. (1996). A class of smoothing functions for nonlinear and mixed complementarity problems. *Computational Optimization and Applications*, 5(2):97–138.

Chen, H. and Baldick, R. (2007). Optimizing short-term natural gas supply portfolio for electric utility companies. *IEEE Transactions on Power Systems*, 22(1):232–239.

Chen, Z. (2007). *Reservoir Simulation: Mathematical Techniques in Oil Recovery*. SIAM, Philadelphia, PA.

Cheng, Y. (2012). Impact of water dynamics in fractures on the performance of hydraulically fractured wells in gas-shale reservoirs. *Journal of Canadian Petroleum Technology*, 51(2):143 – 151.

Chupin, G., Hu, B., Haugset, T., Sagen, J., and Claudel, M. (2007). Integrated wellbore/reservoir model predicts flow transients in liquid-loaded gas wells. In *SPE Annual Technical Conference and Exhibition*, Anaheim, California. SPE 110461.

Cipolla, C. L., Lolon, E., Erdle, J., and Rubin, B. (2010). Reservoir modeling in shale-gas reservoirs. *SPE Reservoir Evaluation & Engineering*, 13(4):638–653.

Coats, B. K., Fleming, G. C., Watts, J. W., Rame, M., and Shiralkar, G. S. (2004). A generalized wellbore and surface facility model, fully coupled to a reservoir simulator. *SPE Reservoir Evaluation & Engineering*, 7(2):3–5.

Coleman, S. B., Clay, H. B., McCurdy, D. G., and Norris, H. L. (1991). A new look at predicting gas-well load-up. *Journal of Petroleum Technology*, 43(3):329–333.

D'Ambrosio, C., Frangioni, A., Liberti, L., and Lodi, A. (2012). A storm of feasibility pumps for nonconvex MINLP. *Mathematical programming*, 136(2):375–402.

Daniilidis, A. and Lemarechal, C. (2005). On a primal-proximal heuristic in discrete optimization. *Mathematical Programming*, 128:105–128.

Danna, E., Rothberg, E., and Pape, C. L. (2004). Exploring relaxation induced neighborhoods to improve MIP solutions. *Mathematical Programming*, 102(1):71–90.

De Santis, M., Lucidi, S., and Rinaldi, F. (2013). A new class of functions for measuring solution integrality in the feasibility pump approach. *SIAM Journal on Optimization*, 23(3):1575–1606.

Ding, Y., Gregov, S., Grodzevich, O., Halevy, I., Kavazovic, Z., Romanko, O., Seeman, T., Shioda, R., and Youbissi, F. (2006). Discussions on normalization and other topics in multi-objective optimization. In *Fields-MITACS, Fields Industrial Problem Solving Workshop*, Toronto, Canada.

Dolan, E. D. and Moré, J. J. (2002). Benchmarking optimization software with performance profiles. *Mathematical programming*, 91(2):201–213.

Dong, Z., Holditch, S. A., Mcvay, D. A., and Ayers, W. B. (2012). Global unconventional gas resource assessment. *SPE Economics & Management*, 4(4):222–234.

Dousi, N., Veeken, C., and Currie, P. K. (2006). Numerical and analytical modeling of the gas-well liquid-loading process. *SPE Production & Operations*, 21(4):6–9.

Dubost, L., Gonzalez, R., and Lemarechal, C. (2005). A primal-proximal heuristic applied to the French unit-commitment problem. *Mathematical Programming*, 104(1):129–151.

Duran, M. A. and Grossmann, I. E. (1986). An outer-approximation algorithm for a class of mixed-integer nonlinear programs. *Mathematical Programming*, 36:307–339.

Economides, M. J. and Wang, X. (2010). Differences and similarities in the stimulation and production of shale gas reservoirs and other tight formations. In *Canadian Unconventional Resources & International Petroleum Conference*, Calgary, Alberta, Canada. SPE 137718.

EIA (2011). Review of Emerging Resources: U.S. Shale Gas and Shale Oil Plays. Technical report, The U.S. Energy Information Administration, Washington DC, USA.

EIA (2013). Annual Energy Outlook 2013. Technical report, The U.S. Energy Information Administration, Washington DC, USA.

EIA (2013a). Natural gas consumption by end use. `http://www.eia.gov/dnav/ng/ng_cons_sum_dcu_nus_m.htm`. [Online; accessed 08-May-2014].

EIA (2013b). Natural gas spot and futures prices (nymex). `http://www.eia.gov/dnav/ng/ng_pri_fut_s1_d.htm`. [Online; accessed 08-May-2014].

EIA (2014). Annual Energy Outlook 2014. Technical report, The U.S. Energy Information Administration, Washington DC, USA.

El-Banbi, A. H. (1998). *Analysis of Tight Gas Wells*. PhD thesis, Texas A&M University.

Engelder, T. (2011). Should fracking stop? *Nature*, 477(7364):271–274.

Fakcharoenphol, P., Torcuk, M., and Bertoncello, A. (2013). Managing shut-in time to enhance gas flow rate in hydraulic fractured shale reservoirs: a simulation study. In *SPE Annual Technical Conference and Exhibition*, New Orleans, Louisiana, USA. SPE 166098.

Feltenmark, S. and Kiwiel, K. C. (2000). Dual applications of proximal bundle methods, including Lagrangian relaxation of nonconvex problems. *SIAM Journal on Optimization*, 10(3):697–721.

FERC (2012). Energy Primer - A Handbook of Energy Market Basics. Technical report, Federal Energy Regulatory Commission.

Fetkovich, M. J. (1980). Decline curve analysis using type curves. *Journal of Petroleum Technology*, 32(6):1065–1077.

Fischetti, M., Glover, F., and Lodi, A. (2005). The feasibility pump. *Mathematical Programming*, 104(1):91–104.

Fischetti, M. and Lodi, A. (2003). Local branching. *Mathematical Programming*, 98(1-3):23–47.

Fischetti, M. and Lodi, A. (2008). Repairing MIP infeasibility through local branching. *Computers & Operations Research*, 35(5):1436–1445.

Fischetti, M. and Salvagnin, D. (2009). Feasibility pump 2.0. *Mathematical Programming Computation*, 1(2-3):201–222.

Fisher, M. L. (1981). The Lagrangian relaxation method for solving integer programming problems. *Management Science*, 27(1):1–18.

Fletcher, R. and Leyffer, S. (1994). Solving mixed integer nonlinear programs by outer approximation. *Mathematical Programming*, 66(1-3):327–349.

Foss, B. (2012). Process control in conventional oil and gas fields – challenges and opportunities. *Control Engineering Practice*, 20(10):1058–1064.

Foss, B., Gunnerud, V., and D. Díez, M. (2009). Lagrangian decomposition of oil-production optimization applied to the Troll west oil rim. *SPE Journal*, 14(4):646–652.

Foss, B. and Jensen, J. P. (2011). Performance analysis for closed-loop reservoir management. *SPE Journal*, 16(1):183–190.

Frangioni, A. (2005). About Lagrangian methods in integer optimization. *Annals of Operations Research*, 139(1):163–193.

Freeman, C., Moridis, G., Ilk, D., and Blasingame, T. (2013). A numerical study of performance for tight gas and shale gas reservoir systems. *Journal of Petroleum Science and Engineering*, 108:22–39.

Fu, Y., Shahidehpour, M., and Li, Z. (2005). Long-term security-constrained unit commitment: hybrid Dantzig-Wolfe decomposition and subgradient approach. *IEEE Transactions on Power Systems*, 20(4):2093–2106.

Gasbarri, S. and Wiggins, M. L. (2001). A dynamic plunger lift model for gas wells. *SPE Production & Facilities*, 16(2):89–96.

Geoffrion, A. M. (1974). Lagrangian relaxation and its uses in integer programming. *Mathematical Programming Study*, 2:82–114.

Glover, F. (1975). Improved linear integer programming formulations of nonlinear integer problems. *Management Science*, 22(4):455–460.

Glover, F. and Kochenberger, G., editors (2003). *Handbook of Metaheuristics*. Kluwer Academic Publishers, Norwell, MA.

Goffin, A. J. L., Haurie, A., and Vial, J. P. (1992). Decomposition and nondifferentiable optimization with the projective algorithm. *Management Science*, 38(2):284–302.

Golan, M. and Whitson, C. (1991). *Well Performance*. Prentice Hall, Upper Saddle River, New Jersey, second edition.

Gollmer, R., Nowak, M. P., and Werner, R. (2000). Unit commitment in power generation - a basic model and some extensions. *Annals of Operations Research*, 96(1-4):167–189.

Gomory, R. (1958). Outline of an algorithm for integer solutions to linear programs. *Bulletin of the American Mathematical Society*, 64:275–278.

Grossmann, I. E. (2002). Review of nonlinear mixed-integer and disjunctive programming techniques. *Optimization and Engineering*, 3(3):227–252.

Grossmann, I. E. and Trespalacios, F. (2013). Systematic modeling of discrete-continuous optimization models through generalized disjunctive programming. *AIChE Journal*, 59(9):3276–3295.

Guignard, M. (2003). Lagrangean Relaxation. *Top*, 11(2):151–228.

Guignard, M. and Kim, S. (1987). Lagrangean decomposition: a model yielding stronger Lagrangian bounds. *Mathematical Programming*, 39(2):215–228.

Guldmann, J.-M. and Wang, F. (1999). Optimizing the natural gas supply mix of local distribution utilities. *European Journal of Operational Research*, 112(3):598–612.

Gunnerud, V. and Foss, B. (2010). Oil production optimization – a piecewise linear model, solved with two decomposition strategies. *Computers & Chemical Engineering*, 34(11):1803–1812.

Guo, B., Lyons, W. C., and Ghalambor, A. (2007). *Petroleum Production Engineering, A Computer-Assisted Approach*. Elsevier Science & Technology Books, Burlington, MA.

Gupta, O. K. and Ravindran, A. (1985). Branch and bound experiments in convex nonlinear integer programming. *Management Science*, 31(12):1533–1546.

Hansen, P. C. (1998). *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. SIAM, Philadelphia, PA.

Harpel, J., Barker, L., Fontenot, J., Carroll, C., Thomson, S., and Olson, K. (2012). Case history of the Fayetteville shale completions. In *SPE Hydraulic Fracturing Technology Conference*, The Woodlands, Texas, USA. SPE 152621.

Held, M., Wolfe, P., and Crowder, H. P. (1974). Validation of subgradient optimization. *Mathematical Programming*, 6(1):62–88.

Helman, C. (2012). Chesapeake energy's new plan: desperate measures for desperate times. `http://www.forbes.com/sites/christopherhelman/2012/02/13/chesapeake-energys-new-plan-desperate-measures-for-desperate-times/`. [Online; accessed 13-February-2014].

Hemmecke, R., Matthias, K., Lee, J., and Weismantel, R. (2010). Nonlinear integer programming. In Jünger, M., Liebling, T. M., Naddef, D., Nemhauser, G. L., Pulleyblank, W. R., Reinelt, G., Rinaldi, G., and Wolsey, L. A., editors, *50 Years of Integer Programming 1958-2008*, pages 561–618. Springer Berlin Heidelberg.

Hiriart-Urruty, J.-B. and Lemarechal, C. (1993). *Convex Analysis and Minimization Algorithms II - Advanced Theory and Bundle Methods*. Springer-Verlag, Berlin.

Hooker, J. (2000). *Logic-Based Methods for Optimization : Combining Optimization and Constraint Satisfaction*. John Wiley & Sons, Inc., New York.

Hou, D., Luo, J., and Al-Tabbaa, A. (2012). Shale gas can be a double-edged sword for climate change. *Nature Climate Change*, 2(6):385–387.

Hughes, J. D. (2013). Energy: A reality check on the shale revolution. *Nature*, 494(7437):307–8.

IBM ILOG CPLEX (2011). *User's Manual*. version 12.3.

IEA (2012). Golden Rules for a Golden Age of Gas - World Energy Outlook - Special report on unconventional gas. Technical report, International Energy Agency, Paris, France.

IFC International (2013). Gas-Fired Power Generation in Eastern New York and its Impact on New England's Gas Supplies. Technical report, ISO New England Inc.

Ilk, D., Rushing, J. A., Perego, A. D., and Blasingame, T. (2008). Exponential vs. hyperbolic decline in tight gas sands – understanding the origin and implications for reserve estimates using Arps' decline curves. In *SPE Annual Technical Conference and Exhibition*, Denver, Colorado, USA. SPE 116731.

ISO New England (2008). CIGRE 2008 Case Study: Electric & Natural Gas Market Interdependencies Within New England. Technical report, ISO New England Inc.

Jayakumar, R. and Rai, R. (2014). Impact of uncertainty in estimation of shale-gas-reservoir and completion properties on EUR forecast and optimal development planning : a Marcellus case study. *SPE Reservoir Evaluation & Engineering*, 17(1):60–73.

Jayakumar, R., Sahai, V., and Boulis, A. (2011). A better understanding of finite element simulation for shale gas reservoirs through a series of different case histories. In *SPE Middle East Unconventional Gas Conference and Exhibition*, Muscat, Oman. SPE 142464.

Jenkins, C. D. and Boyer, C. M. (2008). Coalbed- and shale-gas reservoirs. *Journal of Petroleum Technology*, 60(2):92–99.

Jenner, S. and Lamadrid, A. J. (2013). Shale gas vs. coal: policy implications from environmental impact comparisons of shale gas, conventional gas, and coal on air, water, and land in the United States. *Energy Policy*, 53:442–453.

Johansen, T. A. (1996). Identification of non-linear systems using empirical data and prior knowledge–an optimization approach. *Automatica*, 32(3):337–356.

Johansen, T. A. and Foss, B. (1997). Operating regime based process modeling and identification. *Computers & Chemical Engineering*, 21(2):159–176.

Kaiser, M. J. (2012). Profitability assessment of Haynesville shale gas wells. *Energy*, 38(1):315–330.

Kallehauge, B., Larsen, J., and Madsen, O. B. (2006). Lagrangian duality applied to the vehicle routing problem with time windows. *Computers & Operations Research*, 33(5):1464–1487.

Karimi-Farad, M., Durlofsky, L. J., and Aziz, K. (2004). An efficient discrete-fracture model applicable for general-purpose reservoir simulators. *SPE Journal*, 9(2):227–236.

Katz, D. L. and Lee, R. L. (1990). *Natural Gas Engineering*. McGraw-Hill Publishing Company, New York.

Kazemi, H. (1969). Pressure transient analysis of naturally fractured reservoirs with uniform fracture distribution. *SPE J*, 9(4):451–462.

Kehlhofer, R. (2009). *Combined-Cycle Gas & Steam Turbine Power*. PennWell, Tulsa, Oklahoma.

Kennedy, R. L., Knecht, W. N., Georgi, D. T., and Hughes, B. (2012). Comparisons and contrasts of shale gas and tight gas developments, North American experience and trends. In *SPE Saudi Arabia Section Technical Symposium and Exhibition*, Saudi Arabia. SPE 160855.

Kerschen, G., Golinval, J.-c., Vakakis, A. F., and Bergman, L. a. (2005). The method of proper orthogonal decomposition for dynamical characterization and order reduction of mechanical systems: an overview. *Nonlinear Dynamics*, 41(1-3):147–169.

King, G. E. (2010). Thirty years of gas shale fracturing: what have we learned? In *SPE Annual Technical Conference and Exhibition*, pages 19–22, Florence, Italy. SPE 133456.

Kiwiel, K. C. (1990). Proximity control in bundle methods for convex nondifferentiable minimization. *Mathematical Programming*, 46(3):105–122.

Kiwiel, K. C. (2006). A proximal bundle method with approximate subgradient linearizations. *SIAM Journal on Optimization*, 16(4):1007–1023.

Knudsen, B. R. (2011). A numerical study of fracture geometry and gridding in shale gas reservoir models. Technical report, Norwegian University of Science and Technology.

Knudsen, B. R. and Foss, B. (2012). Mixed integer optimization of the late-life performance of shale-gas wells. Presentation at *the 21st International Symposium on Mathematical Programming (ISMP) held in Berlin, Germany.*

Knudsen, B. R. and Foss, B. (2013). Shut-in based production optimization of shale-gas systems. *Computers & Chemical Engineering*, 58:54–67.

Knudsen, B. R., Foss, B., C.H. Whitson, and A.R. Conn (2012). Target-rate tracking for shale-gas multi-well pads by scheduled shut-ins. In *Proc. of IFAC Symp. on Advanced Control of Chemical Processes*, pages 107–113, Singapore.

Knudsen, B. R., Foss, B., Grossmann, I. E., and Gupta, V. (2013). Lagrangian relaxation based production optimization of tight-formation wells. In *Proc. of IFAC Symp. on Dynamics and Control of Process Systems*, pages 147–152, Mumbai, India.

Knudsen, B. R., Grossmann, I. E., Foss, B., and Conn, A. R. (2014a). Lagrangian relaxation based decomposition for well scheduling in shale-gas systems. *Computers & Chemical Engineering*, 63:234–249.

Knudsen, B. R., Sharma, S., and Foss, B. (2014b). On MINLP heuristics for solving shale-well scheduling problems. In *Proc. of the 19th IFAC World Congress*, pages 2721–2726, Cape Town, South Africa.

Knudsen, B. R., Whitson, C. H., and Foss, B. (2014c). Shale-gas scheduling for natural-gas supply in electric power production. *Energy* (accepted).

Kolodziej, S., Castro, P. M., and Grossmann, I. E. (2013). Global optimization of bilinear programs with a multiparametric disaggregation technique. *Journal of Global Optimization*, 57(4):1039–1063.

Kopanos, G. M. and Pistikopoulos, E. N. (2014). Reactive scheduling by a multiparametric programming rolling horizon framework: a case of a network of combined heat and power units. *Industrial & Engineering Chemistry Research*, 53(11):4366–4386.

Kosmidis, V. D., Perkins, J. D., and Pistikopoulos, E. N. (2005). A mixed integer optimization formulation for the well scheduling problem on petroleum fields. *Computers & Chemical Engineering*, 29(7):1523–1541.

Kravits, M., Frear, R., and Bordwell, D. (2011). Analysis of plunger lift applications in the Marcellus shale. In *SPE Annual Technical Conference and Exhibition*, Denver, Colorado, USA. SPE 147225.

Kumar, A. (2008). *Topics in Mixed Integer Nonlinear Programming.* Phd thesis, Lehigh University, Bethlehem, PA.

Land, A. and Doig, A. (1960). An automatic method of solving discrete programming problems. *Econometrica*, 28:497–520.

Larsen, L. and Hegre, T. M. (1994). Pressure transient analysis of multifractured horizonal wells. In *SPE Annual Technical Conference and Exhibition*, New Orleans, Louisiana, USA. SPE 28389.

Larsson, T., Patriksson, M., and Strömberg, A.-B. (1999). Ergodic, primal convergence in dual subgradient schemes for convex programming. *Mathematical Programming*, 86(2):283–312.

Lea, J. F. (1982). Dynamic analysis of plunger lift operations. *Journal of Petroleum Technology*, 34(11):2617–2629.

Lea, J. F. and Nickens, H. V. (2004). Solving gas-well liquid-loading problems. *Journal of Petroleum Technology*, 56(4):30–36.

Lee, J., Leung, J., and Margot, F. (2004). Min-up/min-down polytopes. *Discrete Optimization*, 1(1):77–85.

Lee, S. and Grossmann, I. E. (2000). New algorithms for nonlinear generalized disjunctive programming. *Computers & Chemical Engineering*, 24(9-10):2125–2141.

Lee, S. and Grossmann, I. E. (2001). A global optimization algorithm for nonconvex generalized disjunctive programming and applications to process systems. *Computers & Chemical Engineering*, 25(11-12):1675–1697.

Lemaréchal, C. (2001). Lagrangian relaxation. In Jünger, M. and Naddef, D., editors, *Computational Combinatorial Optimization*, volume 2241 of *Lecture Notes in Computer Science*, pages 112–156. Springer Berlin Heidelberg.

Li, M., Li, S. L., Sun, L. T., State, T., and Reservoir, G. (2002). New view on continuous-removal liquids from gas wells. *SPE Production & Facilities*, 17(1):42–46.

Li, T., Eremia, M., Member, S., and Shahidehpour, M. (2008). Interdependency of natural gas network and power system security. *IEEE Transactions on Power Systems*, 23(4):1817–1824.

Li, Z. and Ierapetritou, M. G. (2012). Capacity expansion planning through augmented Lagrangian optimization and scenario decomposition. *AiChE Journal*, 58(3):871–883.

Liu, C., Shahidehpour, M., Fu, Y., and Li, Z. (2009). Security-constrained unit commitment with natural gas transmission constraints. *IEEE Transactions on Power Systems*, 24(3):1523–1536.

Liu, C., Shahidehpour, M., and Wang, J. (2011). Coordinated scheduling of electricity and natural gas infrastructures with a transient model for natural gas flow. *Chaos*, 21(2).

Liu, C. H., Martinez-Rahoe, B., Fleckenstein, W., Park, Y., and Shi, F. (2013). Economic feasibility analysis of the Marcellus shale, Pennsylvania. In *Unconventional Resources Technology Conference*, Denver, Colorado, USA. SPE 168886.

Ljung, L. (1999). *System Identification: Theory For The User*. Prentice Hall, New Jersey, second edition.

Lodi, A. (2010). Mixed integer programming computation. In Jünger, M., Liebling, T. M., Naddef, D., Nemhauser, G. L., Pulleyblank, W. R., Reinelt, G., Rinaldi, G., and Wolsey, L. A., editors, *50 Years of Integer Programming 1958-2008*, pages 619–645. Springer Berlin Heidelberg.

Macmillan, S., Antonyuk, A., and Schwind, H. (2013). Gas to Coal Competition in the U.S. Power Sector. Technical report, International Energy Agency/OECD.

Makhanov, K., Kuru, E., and Dehghanpour, H. (2013). Measuring liquid uptake of organic shales: a workflow to estimate water loss during shut-in periods. In *SPE Unconventional Resources Conference - Canada*, Calgary, Alberta, Canada. SPE 167157.

Marcano, L. (1994). Mechanistic design of conventional plunger lift installations. *SPE Advanced Technology Series*, 2(1):15–24.

Marsten, R. E., Hogan, W. W., and Blankenship, J. W. (1975). The boxstep method for large-scale optimization. *Operations Research*, 23(2):389–405.

Martin, R. K. (1998). *Large Scale Linear and Integer Optimization: A Unified Approach*. Kluwer Academic Publishers, Norwell, MA.

Mayerhofer, M. J., Bolander, J. L., Williams, L. I., Pavy, A., and Wolhart, S. (2005). Integration of microseismic fracture mapping , fracture and production analysis with well interference data to optimize fracture treatments in the Overton field , east Texas. In *SPE Annual Technical Conference and Exhibition*, Dallas, Texas, USA. SPE 95508.

Mayerhofer, M. J., Lolon, E. P., Warpinski, N. R., Cipolla, C. L., Walser, D., and Rightmire, C. M. (2010). What is stimulated reservoir volume? *SPE Production & Operations*, 25(1):89–98.

Mayne, D. Q., Rawlings, J. B., Rao, C. V., and Scokaert, P. O. M. (2000). Constrained model predictive control: stability and optimality. *Automatica*, 36(6):789–814.

McGlade, C., Speirs, J., and Sorrell, S. (2013). Unconventional gas – a review of regional and global resource estimates. *Energy*, 55:571–584.

Medeiros, F., Kurtoglu, B., Ozkan, E., and Kazemi, H. (2010). Analysis of production data from hydraulically fractured horizontal wells in shale reservoirs. *SPE Reservoir Evaluation & Engineering*, 13(3):559–568.

Miettinen, K. (1999). *Nonlinear Multiobjective Optimization*, volume 12. Kluwer Academic Publishers, Norwell, MA.

MIT Energy Initiative (2011). The Future of Natural Gas – An Interdisciplinary MIT Study. Technical report, MIT, Boston MA, USA.

Mohaghegh, S. D. (2013). Reservoir modeling of shale formations. *Journal of Natural Gas Science and Engineering*, 12:22–33.

Mouret, S., Grossmann, I. E., and Pestiaux, P. (2011). A new Lagrangian decomposition approach applied to the integration of refinery planning and crude-oil scheduling. *Computers & Chemical Engineering*, 35(12):2750–2766.

Munoz, J., Jimenez-Redondo, Perez-Ruiz, J., and Barquin, J. (2003). Natural gas network modeling for power systems reliability studies. In *Proc. of the IEEE Bologna PowerTech Conference*, Bologna, Italy.

Munoz-Estrada, J., Jimenez-Redondo, N., Perez-Ruiz, J., and Barquin, J. (2004). Including combined-cycle power plants in generation system reliability studies. In *Proc. of the Int. Conf. on Probabilistic Methods Applied to Power Systems*, pages 855–860, Ames, Iowa.

Nemhauser, G. L. and Wolsey, L. A. (1988). *Integer and Combinatorial Optimization*. John Wiley & Sons, Inc., New York.

NERC (2011). A Primer of the Natural Gas and Electric Power Interdependency in the United States. Technical report, The North American Electric Reliability Corporation.

Nobakht, M., Mattar, L., Moghadam, S., and Anderson, D. M. (2012). Simplified forecasting of tight/shale-gas production in linear flow. *Journal of Canadian Petroleum Technology*, 51(6):476–486.

Ockendon, J., Howison, S., Lacey, A., and Movchan, A. (2003). *Applied Partial Differential Equations*. Oxford University Press, New York, revised edition.

Okouma, V., Vittahal, S., Nair, N., and Miller, M. A. (2012). Play-wide well performance analysis in Montney siltstone. In *Canadian Unconventional Resources Conference*, Calgary, Alberta, Canada. SPE 162843.

Oldenburg, J. and Marquardt, W. (2008). Disjunctive modeling for optimal control of hybrid systems. *Computers & Chemical Engineering*, 32(10):2346–2364.

Ouorou, A. (2008). A proximal cutting plane method using Chebychev center for nonsmooth convex optimization. *Mathematical Programming*, 119(2):239–271.

Ozkan, E., Brown, M., and Raghavan, R. (2011). Comparison of fractured-horizontal-well performance in tight sand and shale reservoirs. *SPE Reservoir Evaluation & Engineering*, 14(2):248–259.

Partovi, F., Nikzad, M., Mozafari, B., and Ranjbar, A. M. (2011). A stochastic security approach to energy and spinning reserve scheduling considering demand response program. *Energy*, 36(5):3130–3137.

Pochet, Y. and Wolsey, L. A. (2006). *Production Planning by Mixed Integer Programming*. Springer, New York.

Quesada, I. and Grossmann, I. (1992). An LP/NLP based branch and bound algorithm for convex MINLP optimization problems. *Computers and Chemical Engineering*, 16:937–947.

Quist, A. J., Roos, K., Terlaky, T., van Geemert, R., and Hoogenboom, E. (2001). Reloading nuclear reactor fuel using mixed-integer nonlinear optimization. *Optimization and Engineering*, 2(3):251–276.

Rahmawati, S. D., Whitson, C. H., and Foss, B. (2009). Optimal production strategy for stranded tight-gas reservoirs. In *Seminar on stranded gas including low permeability reservoirs*, Yogyakarta, Indonesia.

Rajan, D. and Takriti, S. (2005). Minimum up/down polytopes of the unit commitment problem with start-up costs. Technical report, IBM Research Report.

Raman, R. and Grossmann, I. E. (1991). Relation between MILP modelling and logical inference for chemical process synthesis. *Computers & Chemical Engineering*, 15(2):73–84.

Raman, R. and Grossmann, I. E. (1994). Modelling and computational techniques for logic based integer programming. *Computers & Chemical Engineering*, 18(7):563–578.

Rao, C. V., Rawlings, J. B., and Lee, J. H. (2001). Constrained linear state estimation - a moving horizon approach. *Automatica*, 37(10):1619–1628.

Rawlings, J. and Mayne, D. (2009). *Model Predictive Control: Theory and Design*. Nob Hill Publishing, Madison, Wisconsin.

Redden, J. (2012). Haynesville operators look to exports for relief. *World oil: Shaletech*, 233(10):598–612.

Reuters (2012). Progress energy to shut in natural gas production. `http://www.reuters.com/article/2012/02/08/progressenergy-idUSL2E8D86SS20120208`. [Online; accessed 13-February-2014].

Reynolds, A., Bratwold, R., and Ding, W. (1987). Semilog analysis of gas well drawdown and buildup data. *SPE Formation Evaluation*, 2(4):657–670.

Rinaldi, G. and Padberg, M. (1991). A branch-and-cut algorithm for the resolution of large-scale symmetric traveling salesman problems. *SIAM Review*, 33(1):60–100.

Ruszczynski, A. (1995). On convergence of an augmented Lagrangian decomposition method for sparse convex optimization. *Mathematics of Operations Research*, 20(3):634–656.

Ryoo, H. S. and Sahinidis, N. V. (1996). A branch-and-reduce approach to global optimization. *Journal of Global Optimization*, 8(2):107–138.

Sagastizábal, C. (2012). Divide to conquer: decomposition methods for energy optimization. *Mathematical Programming*, 134(1):187–222.

Sager, S. (2009). Reformulations and algorithms for the optimization of switching decisions in nonlinear optimal control. *Journal of Process Control*, 19(8):1238–1247.

Sahinidis, N. V. (1996). BARON: A general purpose global optimization software package. *Journal of Global Optimization*, 8(2):201–205.

Sahinidis, N. V. (2004). Optimization under uncertainty: state-of-the-art and opportunities. *Computers & Chemical Engineering*, 28(6-7):971–983.

SENSOR (2011). *SENSOR Manual*. Coats Engineering Inc.

Sharma, S., Knudsen, B. R., and Grimstad, B. (2014). Towards an objective feasibility pump for convex MINLPs. *Manuscript submitted for publication (Computational Optimization and Applications)*.

Shelley, B., Grieser, B., Johnson, B. J., Fielder, E. O., Heinze, J. R., and Werline, J. R. (2008). Data analysis of Barnett shale completions. *SPE Journal*, 13(3):366–374.

Siripatrachai, N. and Ertekin, T. (2012). Alternate representations in numerical modeling of multistage hydraulically fractured horizontal wells in shale gas reservoirs. In *SPE Western Regional Meeting*, Bakersfield, California, USA. SPE 153813.

Sjöberg, J., Zhang, Q., Ljung, L., Benveniste, A., and Delyon, B. (1995). Nonlinear black-box modeling in system identification: a unified overview. *Automatica*, 31(12):1691–1724.

Skogestad, S. and Postlethwaite, I., editors (2005). *Multivariable Feedback Control: Analysis and Design,*. John Wiley & Sons, Inc., New York, 2nd edition.

Smith, E. and Pantelides, C. (1999). A symbolic reformulation/spatial branch-and-bound algorithm for the global optimisation of nonconvex MINLPs. *Computers & Chemical Engineering*, 23(4-5):457–478.

Spinelli, W., Piroddi, L., and Lovera, M. (2005). On the role of prefiltering in nonlinear system identification. *IEEE Transactions on Automatic Control*, 50(10):1597–1602.

Stefik, R. and Paulson, K. (2011). When unconventional becomes conventional. *Journal of Canadian Petroleum Technology*, 50(11):68–70.

Stevenson, B. K. and O'Shea, C. D. (2006). Case study: modeling of a large-scale tight-gas-gathering system. *SPE Projects, Facilities & Construction*, 1(3):1–5.

Strikwerda, J. C. (2004). *Finite Difference Schemes and Partial Differential Equations*. SIAM, Philadelphia, PA, second edition.

Sutton, R. P., Cox, S. A., and Barree, R. D. (2010). Shale gas plays: a performance perspective. In *SPE Tight Gas Completions Conference*, San Antonio, Texas, USA. SPE 138447.

Suwartadi, E. (2012). *Gradient-based Methods for Production Optimization of Oil Reservoirs*. PhD thesis, Norwegian University of Science and Technology.

Takriti, S. and Birge, J. (2000). Using integer programming to refine Lagrangian-based unit commitment solutions. *IEEE Transactions on Power Systems*, 15(1):151–156.

Takriti, S., Krasenbrink, B., and Wu, L. S.-Y. (2000). Incorporating fuel constraints and electricity spot prices into the stochastic unit commitment problem. *Operations Research*, 48(2):268–280.

Tang, Y. (2009). Plunger lift dynamic characteristics in single well and network system for tight gas well deliquification. In *SPE Annual Technical Conference and Exhibition*, New Orleans, Louisiana. SPE 124571.

Tawarmalani, M. and Sahinidis, N. V. (2004). Global optimization of mixed-integer nonlinear programs : A theoretical and computational study. *Mathematical Programming*, 591(3):563–591.

Terrazas-Moreno, S., Trotter, P. A., and Grossmann, I. E. (2011). Temporal and spatial Lagrangean decompositions in multi-site, multi-period production planning problems with sequence-dependent changeovers. *Computers & Chemical Engineering*, 35(12):2913–2928.

Thompson, M. L. and Kramer, M. A. (1994). Modeling chemical processes using prior knowledge and neural networks. *AIChE Journal*, 40(8):1328–1340.

Tomasgard, A., Fodstad, M., and Midthun, K. (2007). Optimization models for the natural gas value chain. In Hasle, G., Lie, K.-A., and Quak, E., editors, *Geometric Modelling, Numerical Simulation, and Optimization*, pages 521–558. Springer Berlin Heidelberg.

Tulleken, H. J. A. F. (1993). Grey-box modelling and identification using physical knowledge and Bayesian techniques. *Automatica*, 29(2):285–308.

Türkay, M. and Grossmann, I. E. (1996). Logic-based MINLP algorithms for the optimal synthesis of process networks. *Computers & Chemical Engineering*, 20(8):959–978.

Turner, R., Hubbard, M., and Dukler, A. (1969). Analysis and prediction of minimum flow rate for the continuous removal of liquids from gas wells. *Journal of Petroleum Technology*, 21(11):1475–1482.

Valkó, P. P. (2009). Assigning value to stimulation in the Barnett shale : a simultaneous analysis of 7000 plus production histories and well completion records. In *SPE Hydraulic Fracturing Technology Conference*, The Woodlands, Texas, USA. SPE 119369.

Vavasis, S. A. (1991). *Nonlinear Optimization: Complexity Issues*. Oxford University Press, Inc., New York.

Vecchietti, A. and Grossmann, I. E. (2000). Modeling issues and implementation of language for disjunctive programming. *Computers & Chemical Engineering*, 24:2143–2155.

Vecchietti, A., Lee, S., and Grossmann, I. E. (2003). Modeling of discrete/continuous optimization problems: characterization and formulation of disjunctions and their relaxations. *Computers & Chemical Engineering*, 27(3):433–448.

Wächter, A. and Biegler, L. T. (2005). On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106(1):25–57.

Warpinski, N. R., Kramm, R. C., Heinze, J. R., and Waltman, C. K. (2005). Comparison of single- and dual-array microseismic mapping techniques in the Barnett shale. In *SPE Annual Technical Conference and Exhibition*, Dallas, Texas, USA. SPE 95568.

Warren, J. and Root, P. (1963). The behavior of naturally fractured reservoirs. *SPE Journal*, 3(3):245–255.

Whitson, C. H. and Brule, M. R. (2000). *Phase Behavior*, volume 20 of *SPE Monograph Series*. Society of Petroleum Engineers.

Whitson, C. H., Rahmawati, S. D., and Juell, A. (2012). Cyclic shut-in eliminates liquid-loading in gas wells. In *SPE/EAGE European Unconventional Resources Conference and Exhibition*, Vienna, Austria. SPE 153073.

Wilson, K. C. and Durlofsky, L. J. (2013). Optimization of shale gas field development using direct search techniques and reduced-physics models. *Journal of Petroleum Science and Engineering*, 108:304–315.

Wolsey, L. A. (1998). *Integer Programming*. John Wiley & Sons, Inc., New York.

Appendix A

Discretization and Parameter Estimation in Slab-based Proxy Model

This appendix describes the parameter estimation technique used to tune the slab-based proxy model applied in Chapter 4 and 5 of the thesis. The governing model equations are restated for readability of the section. Some of the material below is presented in Knudsen et al. (2014b).

Consider the IBVP for the slab-based proxy presented in Section 4.2,

$$\phi\mu(p)c(p)\frac{\partial m}{\partial t} = \frac{\partial}{\partial x}\left(k(x)\frac{\partial m}{\partial x}\right), \tag{A.1a}$$

$$\left.\frac{\partial m}{\partial x}\right|_0 = q\frac{2Tp_{\mathrm{sc}}}{T_{\mathrm{sc}}\Delta z\Delta y_{\mathrm{f}}k_{\mathrm{f}}}, \tag{A.1b}$$

$$\left.\frac{\partial m}{\partial x}\right|_{\frac{L_{\mathrm{s}}}{2}} = 0, \tag{A.1c}$$

$$m(x,0) = m^{\mathrm{init}}. \tag{A.1d}$$

As mentioned in the introduction of the thesis, Section 1.2, the pressure dependency of total compressibility $c(p)$ is difficult to model explicitly in a pseudopressure-based model as (A.1). Consequently, we use the discretized grid compressibilities, c_i, as tuning parameters, while we evaluate the viscosity $\mu(p)$ at initial reservoir pressure. From a reservoir perspective, treating $c(p)$ as a constant mimics the modeling of slightly compressible fluids, however, with the governing PDE (A.1a) formulated in terms of pseudopressure $m(p)$. Furthermore, by using grid block compressibilities as tuning parameters, we are essentially estimating at which pressure $\hat{p}_i$ to evaluate the compressibilities to obtain the best fit of the proxy model, rather than deciding a priori at which pressure to evaluate $c(p)$. Let $\theta^T = [c_1\, c_2 \ldots c_I]$ be the vector unknown parameters in the discretized reservoir proxy model. By discretizing the spatial derivatives in (A.1) using central difference approximations, we obtain an I-dimensional semi-discretization, which can be formulated as the parametric state-

space model

$$\tilde{E}(\theta)\,\dot{m} = \tilde{A}m + \tilde{B}q, \tag{A.2a}$$

$$m(0) = m^{\text{init}}, \tag{A.2b}$$

where $m \in \mathbb{R}^I$ is a *vector* of grid pseudopressures. The matrix $\tilde{E}(\theta)$ is diagonal with elements

$$e_i = c_i \frac{\phi\mu T_{sc}}{2Tp_{sc}} \left(x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}\right) \Delta z \Delta y_{\text{f}}, \tag{A.3}$$

while the tridiagonal and singular matrix $\tilde{A}$ consists of transmissibilities for in and outflow of each grid block $i \in \mathcal{I}$. $\tilde{B}$ is a vector with -1 in the first element and zero otherwise. The spatial discretization of (A.1) is obtained by placing a grid point on each of the two boundaries, $x = 0$ and $x = L_{\text{s}}/2$, and distributing the interior grid points with a logarithmic increasing distance to improve model accuracy in the area close to the fracture, containing the fastest dynamics. Aziz and Settari (1979, Ch. 3.4) refers to this type of spatial discretization as the point-distribution technique for irregular grid, a scheme which guarantees consistency of the finite difference scheme. Discretization of the Neumann boundary conditions (A.1b) and (A.1c) is obtained by using the reflection-point technique.

Applying backwards Euler's approximations for time discretization of (A.2), and assembling the reservoir proxy model with a standard well inflow model and the nonlinear tubing model (4.5), we obtain the following complete shale-*well* proxy model parametrized by θ:

$$A(\theta)\,m_{k+1} = m_k + B(\theta)q_{k+1}, \tag{A.4a}$$

$$m_0 = m^{\text{init}}, \tag{A.4b}$$

$$\tilde{q}_k = \beta\,(m_{k1} - m_{\text{wf},k})\,, \tag{A.4c}$$

$$p^2_{\text{t},k} = \frac{1}{C^2_{\text{t}}} q^2_k + e^{-S} p^2_{\text{wf},k}, \tag{A.4d}$$

$$m_{\text{wf},k} := \tilde{a}_1 p^2_{\text{wf},k} + \tilde{a}_2, \tag{A.4e}$$

where $A(\theta) = I - \tilde{E}(\theta)^{-1}\tilde{A}$ and $B(\theta) = \tilde{E}(\theta)^{-1}\tilde{B}$. The map (A.4e) from pressure squared to pseudopressure is shown in Fig. A.1. As shale-gas wells normally operate on low wellhead pressures, the bottomhole pressure p_{wf} will be correspondingly low, in which the product $\mu(p)Z(p)$ is approximately constant, cf. Fig. 1.10. We can hence construct a map $p^2_{\text{wf}} \mapsto m_{\text{wf}}$ which is linear in p^2 to couple the reservoir inflow model (A.4c) with the tubing model (A.4d), using linear regression to obtain the parameters $(\tilde{a}_1, \tilde{a}_2)$.

The slab-based, explicit-fracture proxy model, based on discretizing a quarter-section of a planar transverse fracture, is generally too coarse to expect a good match in the entire frequency range of a shale-gas well. Similar to the tuning scheme applied in Chapter 3 for the cylindrical SRV-based proxy model, we hence apply prediction-error filtering to weight residuals in least-squares tuning of the proxy model. By including the nonlinear tubing model (A.4d), we can, however, obtain a better estimate of the bottomhole flowing pressure and thereby use this state in parameter estimation for the proxy model. To filter prediction errors, we

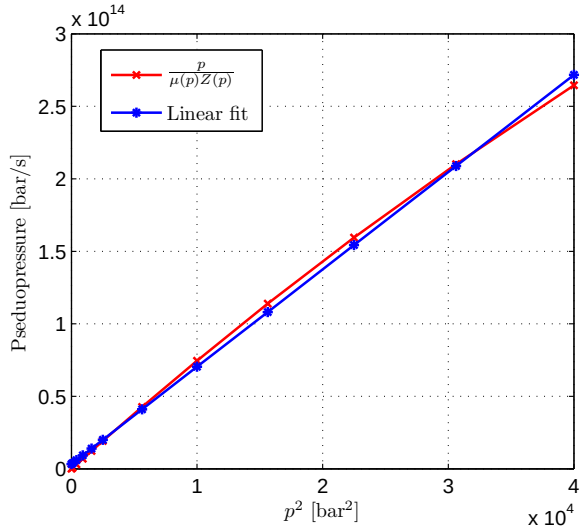


Fig. A.1: Conversion of bottomhole pressure squared to bottomhole pseudopressure, using linear regression to obtain parameters $(\tilde{a}_1, \tilde{a}_2)$ in the affine transformation $m_{\mathrm{wf}} = \tilde{a}_1 p_{\mathrm{wf}}^2 + \tilde{a}_2$.

may use any stable filter $L(\gamma)$, where γ^{-1} symbolizes the unit delay operator. For simplicity, we only consider linear infinite-impulse response (IIR) filters of the form

$$L(\gamma) = \frac{C(\gamma)}{D(\gamma)} = \frac{\sum_{m_1=0}^{M_1} c_{m_1}\gamma^{-m_1}}{1 + \sum_{m_2=1}^{M_2} d_{m_2}\gamma^{-m_2}}, \tag{A.5}$$

where c_{m_1} and d_{m_2} are coefficients in the filter. Let

$$\epsilon_k^{\mathrm{q}} := q_k - q_k^{\mathrm{MFR}}, \tag{A.6}$$

be gas-rate prediction errors (or residuals), where q_k^{MFR} is gas rate from a high-fidelity multi-fracture reference model (MFR), and let $\epsilon_k^{\mathrm{fq}} = L(\gamma)\epsilon_k^{\mathrm{q}}$ denote the *filtered* prediction errors. Filtering of prediction errors can be implemented as a set of linear constraints in an NLP for solving the WNLS problem,

$$\sum_{m_1=0}^{M_1} c_{m_1}\epsilon_{k-m_1} = \epsilon_k^{\mathrm{f}} + \sum_{m_2=1}^{M_2} d_{m_2}\epsilon_{k-m_2}^{\mathrm{f}}, \qquad \forall k \geq \max(M_1, M_2). \tag{A.7}$$

Implementing a recursive digital filter (A.5) as the constraints (A.7) hence yields ϵ_k^{fq} as a causal, weighted sum of past prediction errors.

We assume that measurements of the casing pressure p_{w} and/or tubinghead pressure p_{t} are available, in addition to measurements of the gas rate. In the former case, we may use the relation $p_{\mathrm{wf}} = e^{S/2}p_{\mathrm{w}}$ (Katz and Lee, 1990, Ch. 6) to convert measurements of casing pressures to bottomhole pressures, while (A.4d) or some

other tubing model can be used when tubinghead pressures are available. When using a high-fidelity simulator as SENSOR (SENSOR, 2011), bottomhole pressures will be readily available; we denote these $p^{\mathrm{MFR}}_{\mathrm{wf},k}$. For estimating parameters in the proxy model, we include two prediction-error filters, $L^{\mathrm{q}}(\gamma)$ for filtering of gas-rate prediction errors, ϵ^{q}_k, and $L^{\mathrm{p}}(\gamma)$ for filtering of bottomhole-pressure prediction errors ϵ^{p}_k. Furthermore, we use a high-fidelity MFR model with full-scale geometry as illustrated in Fig. 4.3 as reference model.

Let $\bar{y}$ be a vector of *predefined* binary valve-settings defining a shut-in schedule, i.e. $\bar{y}_k \in \{0, 1\}$, and let $p_{\mathrm{t,MFR}}$ be a vector of tubinghead pressures collected from the MFR model, which we use as input for the proxy model. The vector $\theta \in \mathbb{R}^I$ of unknown parameters in the shale-well proxy model is then estimated by solving the NLP

$$\min_{\theta} \sum_{k\in\hat{\mathcal{K}}} \left(\epsilon^{\mathrm{fq}}_k\right)^2 + \left(\epsilon^{\mathrm{fp}}_k\right)^2 \tag{A.8a}$$

s.t.

$$\epsilon^{\mathrm{fq}}_k = L^{\mathrm{q}}(\gamma)\epsilon^{\mathrm{q}}_k, \qquad \forall k \in \mathcal{K}^{\mathrm{q}} \tag{A.8b}$$

$$\epsilon^{\mathrm{fp}}_k = L^{\mathrm{p}}(\gamma)\epsilon^{\mathrm{p}}_k, \qquad \forall k \in \mathcal{K}^{\mathrm{p}} \tag{A.8c}$$

$$\epsilon^{\mathrm{q}}_k = q_k - q^{\mathrm{MFR}}_k, \qquad \forall k \in \mathcal{K} \tag{A.8d}$$

$$\epsilon^{\mathrm{p}}_k = p_{\mathrm{wf},k} - p^{\mathrm{MFR}}_{\mathrm{wf},k}, \qquad \forall k \in \mathcal{K} \tag{A.8e}$$

$$A(\theta)\, m_{k+1} = m_k + B(\theta) q_{k+1}, \qquad \forall k \in \mathcal{K} \setminus K \tag{A.8f}$$

$$m_0 = m^{\mathrm{init}}, \tag{A.8g}$$

$$\tilde{q}_k = \beta\,(m_{k1} - m_{\mathrm{wf},k})\,, \qquad \forall k \in \mathcal{K} \tag{A.8h}$$

$$m_{\mathrm{wf},k} = \tilde{a}_1 p^2_{\mathrm{wf},k} + \tilde{a}_2, \qquad \forall k \in \mathcal{K} \tag{A.8i}$$

$$\frac{1}{C^2_{\mathrm{t}}} q^2_k = \bar{y}_k \left(e^{-S} p^2_{\mathrm{wf},k} - \left(p^{\mathrm{MFR}}_{\mathrm{t},k}\right)^2\right), \qquad \forall k \in \mathcal{K} \tag{A.8j}$$

$$\theta^{\mathrm{lo}} \le \theta \le \theta^{\mathrm{up}}, \tag{A.8k}$$

where $\hat{\mathcal{K}}$ is a predefined estimation horizon and $\mathcal{K}^v = \{k \in \mathcal{K} : k \ge \max(M_1, M_2)\}$ for $v = \mathrm{q}, \mathrm{p}$. The lower and upper bounds on the compressibilities, θ^{lo} and θ^{up}, are computed from (1.3).

A.1 Grid selection and prefilters in WNLS

The selection of number of grid points I in the proxy model, and the construction of prefilters $L(\gamma)$ for the tuning, are both design issues that may be performed in conjunction with the *objective* of the proxy model. When using proxy models for production optimization, one typically seeks to construct a low-ordered model that retains sufficient accuracy in the frequency range of interest. An appropriate number of grid points for shale-well proxy models may hence be different if the model is designed to predict the initial steep decline, compared to if the model is designed to predict transients from intermittent shut-ins and changes line pressure, cf. Fig. A.2. To obtain a sufficiently accurate fit-for-purpose proxy model, we may

hence adjust both the number of grid points and the prefilter for the prediction errors.

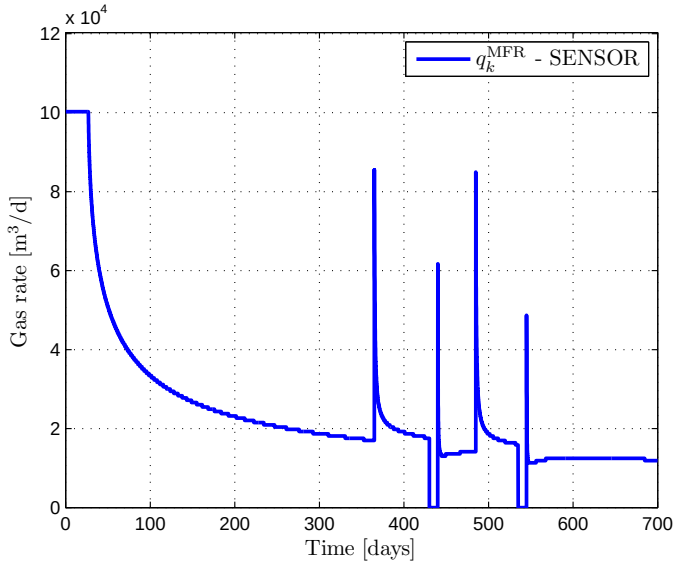
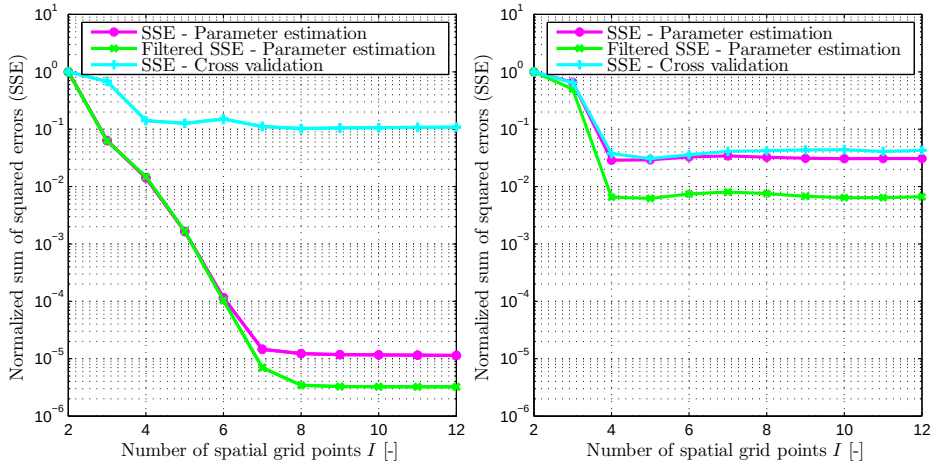


Fig. A.2: Different transients for a shale-gas well, generated from an implementation in SENSOR of the multi-fracture reference model shown in Fig. 1.7.

Butterworth filters are often applied as prefilters in system identification due to a flat magnitude in the bandpass and a sharp drop in magnitude outside the bandwidth of the filter. It should though be commented that nonlinear prediction-error filtering may introduce scattering of the spectrum at frequencies external to the bandwidth (Spinelli et al., 2005). Nevertheless, we apply lowpass 2nd-order digital Butterworth filters for both prefilters, $L^{\mathrm{q}}(\gamma)$ and $L^{\mathrm{p}}(\gamma)$, using a bandwidth $[0, 0.2]$ to reduce emphasis on high-frequency dynamics in the system.

In Fig. A.3 we compare estimation and cross-validation errors for different number of spatial grid points I, applying different horizons shown in Fig. A.2 for estimation and cross validation. In Fig. A.3a, we use the long initial transient for the parameter estimation, omitting the first plateau rate as we do not include a maximum rate $q^{\max}$ in the proxy model (A.4), while we use the subsequent two peak rates for cross validation, i.e. from day $k = 350$ to $k = 475$. The results shown in Fig. A.3b apply the horizon from day $k = 350$ to $k = 475$ for *parameter estimation*, and the remaining horizon from day $k = 476$ to $k = 700$ with the two last peak rates for cross validation. The sum of squared errors (SSE) are all normalized against the value obtained with $I = 2$ grid points. Note that filtered SSE in Fig A.3 corresponds to the value of (A.8a), while SSE for parameter estimation and cross validation is the *unfiltered* counterpart, i.e. actual SSEs. Each parameter estimation applies c_i evaluated at initial reservoir pressure as NLP starting point, and we used the NLP solver IPOPT v3.8 (Wächter and Biegler, 2005) to solve the WNLS problems.



(a) Tuning with initial transient in Fig. A.2 and cross validation with the two subsequent peak rates.

(b) Tuning with peak rates in Fig. A.2 from $k = 350$ to $k = 475$, and the remaining horizon for cross validation

Fig. A.3: Normalized sum of squared errors in parameter estimation and cross-validation with different number of grid points I and two different tuning horizons.

Comparing the two subfigures in Fig. A.3, it can be seen that more grid points are required for tuning of the proxy model against the initial transient compared to the mid-transients before the estimation error flattens out. The SSE from the cross validation in Fig. A.3a, however, shows that little improvement in accuracy is gained by using more than $I = 4$ grid points. Increasing the number of grid points from 4 to 8, hence mainly improves the prediction accuracy of the initial transient as seen in Fig. A.4. This may be beneficial for prediction of long-term pseudo steady-state rates, but does not render any perceivable improvement in accuracy for matching of shorter transients. Observe that the lowpass filters do not significantly impact the estimation errors when using the initial decline rate for tuning.

When tuning the proxy model to match short to medium-long transients, the reduction in estimation and cross validation errors as a function of increasing number of grid points I, can be seen in Fig. A.3b to be significantly from the tuning with the long initial transient. Both the estimation and cross validation errors drop quickly from $I = 2$ to $I = 4$ grid blocks, before the SSEs essentially converge to constant values. However, it can be observed in Fig. A.3b that the SSEs actually increase slightly from $I = 5$. The reason for this can be explained by comparing the gas-rates shown in Fig. A.5. As shown by the cross validations in Fig. A.5b, the gas rate for $I = 4$ and $I = 12$ are very similar, however, with the rate for $I = 4$ matching the peak-rate magnitude slightly better compared to the rate for $I = 12$, while the latter matches the curvature of the transients slightly better. As the magnitudes of prediction errors at the peak rate have a great impact on total estimation erros, the SSEs end up being slightly higher for increasing number of

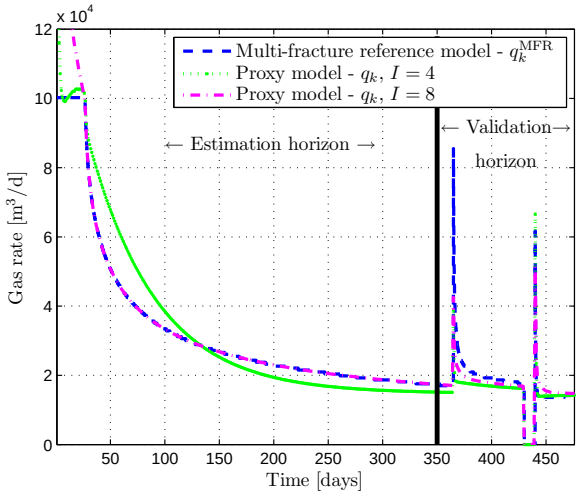
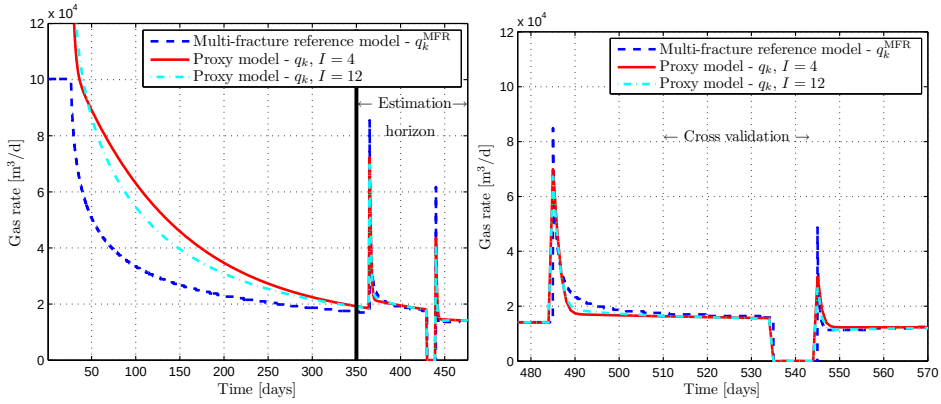


Fig. A.4: Gas rate for $I = 4$ and $I = 8$ grid points, using the initial transient up to day $k = 350$ for estimation, and the two subsequent transients for cross validation.

grid points. In this context, it is also important to note that (A.8) yields a non-convex NLP, with the consequence that optimal parameter vectors θ^* may depend on the provided NLP starting points.



(a) Parameter estimation. The initial decline rate, not included in the tuning, is also shown.

(b) Cross validation.

Fig. A.5: Model tuning, horizon from day $k = 350$ to $k = 475$ for parameter estimation, and the horizon from day $k = 476$ to $k = 700$ for cross validation.

Comparing Fig. A.3a and A.3b, it is observed that the lowpass filtering of prediction errors causes a much bigger difference in filtered and unfiltered estimation errors when tuning the model against short to medium-long transients than against the long initial transient. The lowpass filters effectively reduce the influence of prediction errors around the peak rates; matching tightly the magnitude of these peak

rates will translate into a loss in accuracy of pseudo steady-state rates. Although one may consider using a bandpass instead of a lowpass filter, using a lowpass filter is generally preferable in order to retain a good match of the pseudo steady-state rates. Fig. A.5a further shows that designing the proxy model to be accurate for short to medium-long transients comes at the expense of a reduced accuracy in prediction of the long initial transient.

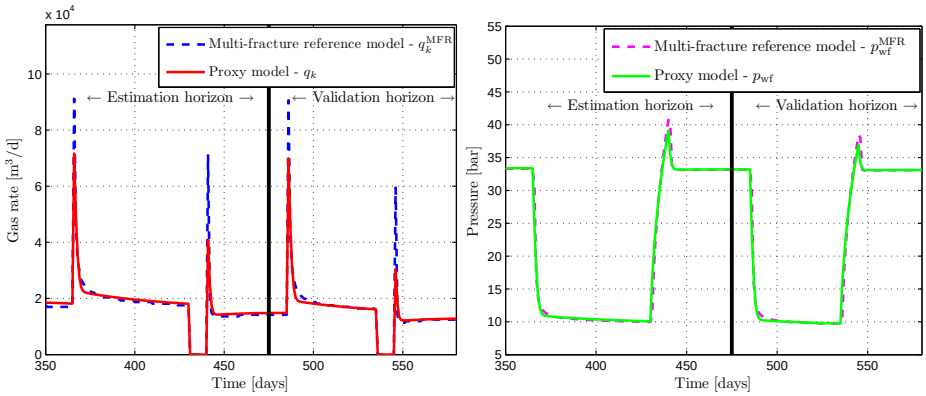
A.1.1 Discussion on the tuning scheme

Several approaches may be considered in the designing of the proxy model, both in terms of prefiltering, grid construction and set-up of the tuning procedure. Some shale-gas wells may occasionally during production exhibit multi-phase flow effects due liquids in the wellbore. Using the tubing model (A.4d) to estimate p_{wf} may in these situations lead to poor estimates, particularly when the well is producing, and thereby be detrimental for tuning of the proxy models. A possible remedy would be to replace the objective (A.8a) with $\bar{y}_k \left(\epsilon_k^{\text{fq}}\right)^2 + (1-\bar{y}_k)\left(\epsilon_k^{\text{fp}}\right)^2$, only using the estimates of p_{wf} to match bottomhole pressure build-up during shut-ins, while solely using the gas rates for matching during production of the well. Note that outliers in the production data may be suppressed by appropriate filtering.

A challenge when tuning the proxy model against short to medium-long transients is to retain a sufficient accuracy for prediction of long-term production rates. To improve this feature of the proxy model, a viable approach is to include a correction term for deviations in cumulative production as in Chapter 2. The correction term may be added added to the objective (A.8a), or as a constraint, defining a threshold δ_{cum} for acceptable error in cumulative production. Observe that the former technique of adding a correction term for deviation in cumulative production to the objective function, corresponds to adding an integrator-filter of the prediction errors ϵ_k^{q}. See Ljung (1999, Ch. 14.4) for a discussion on the design of prediction-error filters.

When adding more grid points to increase the accuracy of the proxy, it is important to be aware of certain pitfalls with respect to the tuning. Unless tuning parameters are combined for several of the grid blocks, then increasing the number of grid points increases linearly the size of θ, thereby requiring further excitation of the system to give reliable parameter estimates. Moreover, when the number of estimated parameters becomes high, one may end up fitting the proxy model to *local* effects and transients in the system that are not essential for the purpose of the model.

Finally, we show in Fig. A.6 tuning and cross validation of q_k and $p_{\text{wf},k}$ for the model applied in Chapter 4 and 5. Using the procedure described in the previous section with low-pass filtering of prediction errors, is seen to give a good match and cross-validation of both pseudo steady-state rates as well as the transients in the gas rate, cf. Fig A.6a. The cut-off frequencies of the lowpass filter is seen to leave a discrepancy in both the peak gas-rates and at the end of the pressure build-up during shut-ins as seen in Fig. A.6b. By also including matching of p_{wf}, we obtain a more correct pressure build-up during shut-ins, and hence overall enhanced



(a) Gas Rates. **(b)** Bottomhole pressures.

Fig. A.6: Parameter estimation and cross validation of proxy model with $I = 4$ grid blocks, used in studies in Chapter 4 and 5.

accuracy of transients in the proxy compared to the scheme used in Chapter 3 with matching against the gas rate only.