Robust and Distributed Approaches to Power System Optimization

A dissertation submitted to

ETH Zurich

for the degree of

Doctor of Sciences

presented by

Joseph Warrington
M.Eng., University of Cambridge
born 07.05.1986
citizen of the United Kingdom

accepted on the recommendation of

Prof. Dr. Manfred Morari (examiner)
Prof. Kameshwar Poolla (co-examiner)

2013
# Contents

Abstract v  
Zusammenfassung vii  
Dissertation Structure ix  
Summary of Contributions ix  
List of Figures xii  
List of Tables xiii  
List of Acronyms xiv  

I Introduction 1  

1 Introduction 3  
1.1 Motivation 3  
1.1.1 The growth of intermittent renewable energy 3  
1.1.2 Increasing reserve requirements 4  
1.1.3 Electricity market practice 7  
1.1.4 Online optimization for power systems 9  
1.2 Sources of uncertainty in power systems 12  
1.2.1 Consumption 13  
1.2.2 Intermittent renewables 13  
1.2.3 Contingencies 15  
1.3 Literature survey 16  
1.3.1 Optimal Power Flow 16
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.3.2 Robust optimization with recourse</td>
<td>18</td>
</tr>
<tr>
<td>1.3.3 Distributed optimization and pricing mechanisms</td>
<td>19</td>
</tr>
<tr>
<td>1.3.4 Novel electricity pricing mechanisms</td>
<td>22</td>
</tr>
<tr>
<td>II Policy-Based Reserves for Power Systems</td>
<td>25</td>
</tr>
<tr>
<td>2 Affine Reserve Policies</td>
<td>27</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>28</td>
</tr>
<tr>
<td>2.2 Power system model</td>
<td>30</td>
</tr>
<tr>
<td>2.2.1 Participant model</td>
<td>30</td>
</tr>
<tr>
<td>2.2.2 Network model</td>
<td>33</td>
</tr>
<tr>
<td>2.3 Optimal affine reserve policies</td>
<td>34</td>
</tr>
<tr>
<td>2.3.1 Finite horizon optimization</td>
<td>35</td>
</tr>
<tr>
<td>2.3.2 Equivalent tractable reformulation</td>
<td>37</td>
</tr>
<tr>
<td>2.3.3 Computational requirements</td>
<td>40</td>
</tr>
<tr>
<td>2.4 Numerical case study</td>
<td>41</td>
</tr>
<tr>
<td>2.4.1 System model</td>
<td>41</td>
</tr>
<tr>
<td>2.4.2 Uncertainty model</td>
<td>42</td>
</tr>
<tr>
<td>2.4.3 Comparison of reserve costs</td>
<td>45</td>
</tr>
<tr>
<td>2.4.4 Sensitivity analysis</td>
<td>47</td>
</tr>
<tr>
<td>2.5 Discussion</td>
<td>48</td>
</tr>
<tr>
<td>3 Implementation of Affine Reserve Policies</td>
<td>51</td>
</tr>
<tr>
<td>3.1 Market incorporation</td>
<td>51</td>
</tr>
<tr>
<td>3.1.1 Efficient reserve policy prices</td>
<td>53</td>
</tr>
<tr>
<td>3.1.2 Policy-based reserve products</td>
<td>55</td>
</tr>
<tr>
<td>3.1.3 Clearing mechanisms</td>
<td>56</td>
</tr>
<tr>
<td>3.1.4 Real time operation</td>
<td>58</td>
</tr>
<tr>
<td>3.1.5 Decision-making process</td>
<td>60</td>
</tr>
<tr>
<td>3.2 Technical incorporation</td>
<td>60</td>
</tr>
<tr>
<td>3.2.1 Existing control structure</td>
<td>62</td>
</tr>
<tr>
<td>3.2.2 Adapting AGC to implement the first step of a reserve sequence</td>
<td>63</td>
</tr>
<tr>
<td>3.2.3 Adapting tertiary reserves to implement the remaining steps of a reserve sequence</td>
<td>65</td>
</tr>
<tr>
<td>4 Extensions</td>
<td>67</td>
</tr>
</tbody>
</table>
# III Distributed Optimal Power Flow and Market Mechanisms

5 Distributed Solutions to Welfare Maximization Problems

5.1 Welfare maximization problems

5.1.1 Problem statement

5.1.2 Elastic and inelastic supply and demand

5.1.3 The central planner’s solution

5.1.4 Profit maximization and price coordination

5.1.5 Pricing in the absence of (strict) convexity

5.2 Solution mechanisms

5.2.1 Double-sided auction

5.2.2 Iterative pricing mechanisms

6 Multi-Period Optimal Power Flow

6.1 A dual ascent mechanism for solving multi-period OPF

6.1.1 Problem statement

6.1.2 Market-based solution

6.1.3 Modified algorithms for improved performance

6.1.4 Results

6.1.5 Further improvements

6.1.6 Conclusions

6.2 Receding horizon (rolling window) electricity market

6.2.1 Comparison with Economic MPC

6.2.2 Negotiated predictive dispatch

6.2.3 Conclusion
# IV Research Outlook

## 7 Short- and Medium-Term Objectives

<table>
<thead>
<tr>
<th>Objective</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affine reserve policies</td>
<td>135</td>
</tr>
<tr>
<td>Online decision-making tools</td>
<td>137</td>
</tr>
<tr>
<td>ADMM for large-scale combinatorial optimization problems relevant to power systems</td>
<td>138</td>
</tr>
<tr>
<td>Pricing mechanisms</td>
<td>138</td>
</tr>
<tr>
<td>AC optimal power flow</td>
<td>139</td>
</tr>
</tbody>
</table>

## 8 Long-Term Challenges

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business as usual</td>
<td>141</td>
</tr>
<tr>
<td>Sharing costs and profits at high renewable penetration</td>
<td>142</td>
</tr>
</tbody>
</table>

# Appendix: DC load flow approximation

# Bibliography

# Publication List

# Curriculum Vitae
Abstract

This dissertation describes optimization-based approaches to improving the short-run efficiency of electrical power systems that operate under uncertainty, in particular uncertainty arising from large quantities of intermittent renewable energy. Its contributions fall into two categories.

First, robust optimization and control principles are used to define a new approach to the provision of electrical reserves, which maintain power system integrity under uncertainty. This approach relies on commitments referred to as affine reserve policies, which are planned multi-stage responses to uncertainty, chosen by solving an optimization problem based on the current system state and forecast information. For each participating device (such as a generator, energy storage unit, deferrable load, or curtailable wind farm) the policy consists of a nominal schedule plus a system of time-coupled adjustments that act when the values of uncertain variables are discovered in future time steps. These uncertain variables may model any random parameter affecting system operation, but may most usefully model poorly-predicted intermittent renewable energy infeeds, and the optimization framework exploits the fact that these uncertainties are typically correlated in time. The resulting reserve policies are guaranteed to be feasible for every realization of the uncertainty (within some assumed bounds) that could arise.

Affine reserve policies are shown via numerical experiments to have two key benefits. The first is that they are able to reduce the dynamic component of system costs, such as those associated with generator ramping, by planning more systematically how the sequence of operating points will be adjusted when uncertainties are discovered. The result of this is that the total cost of making the power system robust to disturbances is reduced. The second benefit is obtained when the policy optimization is coupled with a unit commitment (generator switching decision) problem. In this case it is found that the use of expensive peaking generators may often be reduced, owing to the fact that time-coupled policies make the control of conventional generators and energy storage less conservative.

In addition, it is reported that the reserve policy optimization problem can be solved for large problem instances by using distributed optimization approaches. This is facilitated by the linear structure of the constraints, which allow the interests of individual participating devices to be separated from those of the network operator via a price-like function.
Second, distributed optimization principles are used to propose price clearing mechanisms which are able to solve Optimal Power Flow (OPF) problems efficiently over a finite planning horizon. The pricing mechanisms are based on dual ascent algorithms, which have the interpretation of a price negotiation between the various market participants coordinated by a market operator, who adjusts prices until an optimal utilization of the generation and storage assets is obtained. Convergence of the method is guaranteed by ensuring that the characteristics of the participants adhere to the generic convexity assumptions upon which convergence proofs of subgradient algorithms are based. The market model is characterized by an arbitrary number of independent, profit- or utility-maximizing, price-taking participants. These are constrained by an alternating-current (AC) transmission network of arbitrary topology, in which real and reactive power flows, voltages, and power flows must satisfy various constraints. The algorithm makes use of a recently-discovered tight semidefinite relaxation of the static OPF problem, and heuristic approaches are used to improve the convergence speed significantly.

A market operation mode based on a receding horizon (also known as rolling window) principle is then described, as a logical extension of existing electricity markets that increasingly resort to intra-day trading in order to accommodate fluctuating renewable infeeds. This mechanism enables adjustments to the schedules of market participants to be made in the light of constantly-updating nominal forecasts of these infeeds, by conducting a re-negotiation of their actions. Such an approach can reduce system costs by allowing generator ramping actions and the operation of energy storage to be re-planned more effectively than under existing redispatch mechanisms.
Zusammenfassung

Diese Arbeit beschäftigt sich mit Optimierungsalgorithmen zur Verbesserung der kurzfristigen Effizienz von elektrischen Stromnetzen, die unter Unsicherheiten betrieben werden - insbesondere Unsicherheiten welche durch die Einspeisung erneuerbarer Energieträger entstehen. Die wissenschaftlichen Beiträge lassen sich in zwei Kategorien einteilen.


Simulationen zeigen, dass Rückführungsstrategien für Energiereserven zwei wesentliche Vorteile haben. Erstens werden die dynamischen Komponenten der Systemkosten, z.B. solche, die durch das Anfahren von Generatoren verursacht werden, reduziert, in dem systematisch geplant wird, wie die Abfolge der Arbeitspunkte angepasst wird wenn Abweichungen erkannt werden. Als Konsequenz sinken die Gesamtkosten, die aufgewendet werden müssen, um das Stromversorgungssystem robust gegenüber Störungen zu machen. Der zweite Vorteil offenbart sich, wenn die Rückführungsstrategieoptimierung mit einer Frage des Kraftwerkseinsatzes verbunden wird. Für diesen Fall finden sich Belege, dass die Verwendung teurer Spitzenlastgeneratoren häufig reduziert werden kann, weil die zeitlich übergreifenden Rückführungsstrategien eine weniger zurückhaltende Regelung konventioneller Generatoren und Energiespeicher möglich machen.


Dissertation Structure

Part I: Chapter 1 is an introduction containing a motivation for the work described in the dissertation, detailing challenges facing power system operators that have arisen from the incorporation of intermittent renewables. It also contains a survey of relevant literature from Optimal Power Flow, distributed optimization, and robust optimization.

Part II: Chapters 2 to 4 describe a multi-stage approach to accommodating large, correlated renewable in-feeds in power systems, in which optimal affine reserve policies are chosen for all flexible power market participants. The benefits of the method are shown via case studies, and extensions of the method are described.

Part III: Chapter 5 describes various dual-based distributed optimization techniques, and their interpretations as market mechanisms. Chapter 6 presents an application of these methods to predictive dispatch in power systems via price negotiation, and finally describes a receding-horizon market clearing concept.

Part IV: Chapter 7 proposes ways in which the ideas presented in Parts II and III could be developed further. Chapter 8 describes two broader emerging issues in power systems brought about by large renewable penetrations.

Summary of Contributions

The main contributions are contained in Parts II and III of this dissertation. Citations in brackets denote publications from which content was drawn.

Part II:

- The concept of a multi-stage affine reserve policy, as an improved means of handling errors in the prediction of uncertain loads or renewable in-feeds. An affine policy is a set of rules that a market participant (such as a generator, energy storage unit, or flexible load) must follow when providing reserves to the grid. The corrective actions supplied by the market participant become time-coupled in a way that exploits the correlation of uncertainties along a planning horizon ([WGMM12b]).

- A tractable and scalable optimization approach, based on recent advances in robust predictive control, for choosing optimal affine reserve policies for
bounded uncertainties ([WGMM12b, WGMM13]).

- The demonstration, using closed-loop simulations with a stochastic wind model, that reserve costs using such policies can be lowered compared to the best possible time-decoupled reserve actions ([WGMM13]).

- An elaboration on the implementation details of such a reserve mechanism, in terms of technical requirements (for example adapting generator control actions) and its incorporation into modern electricity markets ([WMM13]).

- An extension to include coupled unit commitment decisions with the choice of reserve policy. It is shown that costs could be reduced by avoiding the activation of expensive peaking generation in some situations ([WHGM14]).

Part III:

- A reformulation of the AC Optimal Power Flow (OPF) problem to include generic market participants with arbitrary convex time-coupled costs and constraints. This is more appropriate for the consideration of market interactions than conventional formulations ([WGMM12a]).

- A market-clearing mechanism for multi-period AC OPF based on a dual subgradient method. The mechanism exploits a recently-discovered tightness property of a semidefinite relaxation of the static OPF problem, in order to obtain exact global solutions to an initially non-convex problem ([WGMM12a]).

- Two heuristics for improving the convergence performance of the standard dual subgradient method, based on physical insights related to the network physics ([WGMM12a]).

- A receding-horizon (also known as a rolling window) market where prices are negotiated over a time horizon in response to continually-updating wind forecasts. The approach allows energy storage and generator ramping actions to be managed more efficiently than existing market clearing approaches based on time-decoupled optimization ([WMM11]).
List of Figures

1.1 Continental-scale weather fronts in Europe ........................................... 5
1.2 Future short-term operating reserve requirements in the UK .................. 6
1.3 Reserve costs as a function of wind penetration ..................................... 7
1.4 ENTSO-E reserve products ...................................................................... 8
1.5 Locational marginal prices cleared by the Midwest ISO, 1st November 2013 .............................................................. 10
1.6 Zonal prices in the Nordic spot market, 3rd November 2013 .................... 11
1.7 Nominal and measured wind power outputs ........................................... 14
2.1 Affine reserve policy structure ................................................................. 36
2.2 Modified 39 bus test network .................................................................. 42
2.3 Example output of correlated uncertainty model $q_k$ ................................. 44
2.4 Power output traces under affine reserve policies ..................................... 46
2.5 Sensitivity of reserve costs to wind penetration level ................................. 47
3.1 Structure of reserve policy matrix ............................................................. 52
3.2 Auctions for two types of reserve policy commitment ............................... 57
3.3 Batchwise and receding horizon reserve policy decisions .......................... 61
3.4 General decision-making optimization scheme for system operators ......... 61
3.5 Standard primary and secondary control structure for generators ............. 62
4.1 Nominal prediction of net load, with shaded uncertainty bounds ............... 72
4.2 Test system use for coupled unit commitment and policy optimization ....... 73
4.3 Structure of policy matrices $D_i$ permitted in cases 1 and 2 ....................... 75
4.4 Number of time steps in which expensive peaking plants were required, with and without time-coupled policies ............................... 76
4.5 Expected finite horizon cost (fuel plus switching costs) with and without time-coupled policies .................................................... 77
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.6</td>
<td>Case 1: Robust unit commitment schedule without time-coupled policies</td>
<td>78</td>
</tr>
<tr>
<td>4.7</td>
<td>Case 2: Robust unit commitment schedule with time-coupled policies</td>
<td>79</td>
</tr>
<tr>
<td>4.8</td>
<td>Feedback policy matrices with and without time coupling</td>
<td>80</td>
</tr>
<tr>
<td>4.9</td>
<td>Number of iterations needed by ADMM algorithm</td>
<td>84</td>
</tr>
<tr>
<td>4.10</td>
<td>ADMM residuals for nominal and disturbance power matching constraints</td>
<td>84</td>
</tr>
<tr>
<td>4.11</td>
<td>Segregation of one-dimensional disturbance into two components</td>
<td>86</td>
</tr>
<tr>
<td>5.1</td>
<td>Profit maximizing actions under different convexity assumptions</td>
<td>94</td>
</tr>
<tr>
<td>5.2</td>
<td>Standard double-sided auction</td>
<td>97</td>
</tr>
<tr>
<td>6.1</td>
<td>Schematic of dual subgradient method</td>
<td>114</td>
</tr>
<tr>
<td>6.2</td>
<td>Modified 39 bus test network</td>
<td>119</td>
</tr>
<tr>
<td>6.3</td>
<td>Convergence behaviour of prices for real and reactive power</td>
<td>121</td>
</tr>
<tr>
<td>6.4</td>
<td>Equality constraint residuals for real and reactive power</td>
<td>122</td>
</tr>
<tr>
<td>6.5</td>
<td>Nodal real power prices over 10 steps</td>
<td>123</td>
</tr>
<tr>
<td>6.6</td>
<td>Real and reactive power schedules for generators and storage units</td>
<td>123</td>
</tr>
<tr>
<td>6.7</td>
<td>Complex bus voltages for step 1 of AC OPF solution</td>
<td>124</td>
</tr>
<tr>
<td>6.8</td>
<td>Receding horizon market</td>
<td>127</td>
</tr>
<tr>
<td>6.9</td>
<td>PJM-like market mechanism and receding-horizon modification</td>
<td>128</td>
</tr>
</tbody>
</table>
List of Tables

2.1 Parameters of elastic and inelastic participants . . . . . . . . . . . 43
2.2 Cost comparison for different structural restrictions on $D_i$ . . . . . . 46
2.3 Variation of results with installed wind power capacity . . . . . . . 48
4.1 Participant parameters . . . . . . . . . . . . . . . . . . . . . . . 74
6.1 AC OPF simulation parameters . . . . . . . . . . . . . . . . . . . . 120
6.2 Convergence time and optimality results for Algorithms 1, 2, and 3 . . 120
6.3 Cost comparison for receding horizon market implementations . . . 130
## List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>Alternating Current</td>
</tr>
<tr>
<td>ACCPM</td>
<td>Analytic-Centre Cutting Plane Method</td>
</tr>
<tr>
<td>ACE</td>
<td>Area Control Error</td>
</tr>
<tr>
<td>ADMM</td>
<td>Alternating Direction Method of Multipliers</td>
</tr>
<tr>
<td>AGC</td>
<td>Automatic Generator Control</td>
</tr>
<tr>
<td>DC</td>
<td>Direct Current</td>
</tr>
<tr>
<td>DSO</td>
<td>Distribution System Operator</td>
</tr>
<tr>
<td>ENTSO-E</td>
<td>European Network of Transmission System Operators for Electricity</td>
</tr>
<tr>
<td>ISO</td>
<td>Independent System Operator</td>
</tr>
<tr>
<td>LDR</td>
<td>Linear Decision Rule</td>
</tr>
<tr>
<td>LMP</td>
<td>Locational Marginal Price</td>
</tr>
<tr>
<td>LP</td>
<td>Linear Program</td>
</tr>
<tr>
<td>LR</td>
<td>Lagrangian Relaxation</td>
</tr>
<tr>
<td>MIQP</td>
<td>Mixed-Integer Quadratic Program</td>
</tr>
<tr>
<td>MLD</td>
<td>Mixed Logical-Dynamical (system)</td>
</tr>
<tr>
<td>MPC</td>
<td>Model Predictive Control</td>
</tr>
<tr>
<td>OPF</td>
<td>Optimal Power Flow</td>
</tr>
<tr>
<td>PJM</td>
<td>Pennsylvania-Jersey-Maryland (market interconnection)</td>
</tr>
<tr>
<td>RMS</td>
<td>Root Mean Square</td>
</tr>
<tr>
<td>RO</td>
<td>Robust Optimization</td>
</tr>
<tr>
<td>SCOPF</td>
<td>Security-Constrained Optimal Power Flow</td>
</tr>
<tr>
<td>TSO</td>
<td>Transmission System Operator</td>
</tr>
<tr>
<td>UC</td>
<td>Unit Commitment</td>
</tr>
<tr>
<td>VoLL</td>
<td>Value of Lost Load</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

This chapter provides a motivation and introduction to the topic of this dissertation. Section 1.1 describes the transition that many electrical grids are making towards a reliance on intermittent renewable sources of electricity, and explains the challenges that this is bringing to system operation. Section 1.2 describes the various sources of uncertainty that must be accommodated in daily decision-making for power systems. Section 1.3 is a survey of related literature from related areas of optimal power flow, market mechanism design, and robust optimization.

1.1 Motivation

1.1.1 The growth of intermittent renewable energy

Electricity provision is undergoing a transition from predominantly fossil-fuelled generation to a high fraction of generation from renewable energy sources, particularly intermittent ones such as wind and solar plants. The grid built up in the 20th century depended almost solely on energy sources such as coal, gas, and nuclear power, whose real-time operation was comparatively straightforward. In more recent years renewable sources have become an increasingly attractive alternative, in part due to their low marginal costs of generation and in part due to the imperative on the part of governments to bring about reductions in greenhouse gas emissions. However, the intermittent outputs of these fast-growing sources poses a great challenge to the reliable and economic provision of electricity [New10].
1. Introduction

In some countries the penetration of intermittent electricity sources is already advanced. As of 2013, countries such as Denmark (27%), Portugal (17%), and Spain (16%) already obtain considerable percentages of their electrical energy needs from the wind \[WM13\]. The penetration of solar energy sources lags considerably behind that of wind, but brings with it the same inherent intermittency problems. Many countries, particularly in Europe, have set ambitious targets for renewable electricity generation in order to meet greenhouse gas emissions targets for the coming decades. These percentages will therefore continue to rise rapidly, bringing about several technical and economic challenges that form the motivation for this dissertation.

1.1.2 Increasing reserve requirements

Power generation is scheduled in advance, in order to match a forecast of system load (consumption plus network losses) as efficiently as possible. However since this forecast will be inaccurate to some degree, a reserve mechanism is needed to regulate power generation to match the real requirements as they are discovered in real time. This dissertation is concerned with providing such a mechanism more efficiently by exploiting more granular or up-to-date prediction information.

Wind and solar power sources are prone to extremely fast ramps in output, both upward and downward. It is possible for multiple gigawatts of output to be gained or lost over periods as short as 15 minutes \[HFD13\], Fig. 2.1) due to the way weather fronts pass over the clustered generation sites, and these fluctuations cause difficulties to system operators responsible for balancing the grid \[XT12\]. Although installing renewable sources over a very large area (thereby reducing the degree of correlation between their outputs) can mitigate this risk, a large backup of fast-ramping reserves must always be operated to compensate for these rapid fluctuations. This is for two reasons:

1. Transmission connections between areas, particularly across national boundaries, are typically too weak to fully protect local grids from fluctuations from their own renewables.

2. Weather fronts frequently evolve on a continental scale, as shown in Fig. \[1.1\] and in the worst case there will be strong correlation even over a large area. This means, for example, that a drop in wind power output can potentially be synchronized across a country’s entire wind farm portfolio.
1.1. Motivation

Figure 1.1: Continental-scale weather fronts dividing air masses of different temperatures, humidities, and densities. Source: http://www.wetter.net, accessed 29th October 2013.

Reserve margins have traditionally been dominated by a need to maintain security against load forecast inaccuracies and the worst contingency scenario envisaged by the system operator [Web10]. In practice this means fast-acting reserves are needed to maintain supply in case one of the largest generators is unexpectedly taken offline for some reason. Compared to this, the uncertainty brought about by inaccurate predictions of net load (load minus any contributions from uncontrolled power in-feeds) has historically been somewhat small. However, high-penetration renewables are starting to change this situation in several countries, since the resulting uncertainty margins are starting to exceed the original contingency-based requirements [Pre12]. For example, the UK’s National Grid procures a near-constant 5 GW of fast-acting tertiary reserves today, relative to a peak load of 60 GW, but expects this to become increasingly variable, with the maximum requirement to increase sharply to 18 GW by 2025, almost a third of expected peak demand.
1. Introduction

Figure 1.2: Future short-term operating reserve requirements in the UK, as projected by National Grid [HFD13]. Requirements will increase sharply with increasing wind power penetration. The dotted lines represent the bounds on reserve requirements, which vary according to wind conditions.

Forecast for that year ([HFD13], Fig. 2.2). In other words, the source of a third of that country’s electrical power will at times be indeterminate even only a few hours ahead of delivery, and at off-peak times the fraction could be higher still.

Increased reserve margins are an unavoidable result of introducing an increasingly uncertain component into the electricity generation mix. Several studies have been carried out to quantify this need and its associated costs [HFD13, AAD+07, HMK+08], see Figs. 1.2 and 1.3. Reserve capacity needs increase with the share of intermittent renewables, but in addition the marginal cost of providing this capacity increases. This is because grids incorporating large shares of renewables experience very fast ramps in net power requirements. These fast ramps can only be tracked by the most agile plants, such as inefficient open-cycle gas turbines, which in turn means that cheaper, slower-acting plants are only able to supply a lower fraction of total energy.

Reserves are provided on several levels. The fastest, known as Frequency Containment Reserves (or primary reserves), act automatically within seconds based on local frequency measurements in order to keep the grid frequency within safe bounds. The second-fastest, known as Frequency Restoration Reserves (or secondary reserves), act on a regional Area Control Error signal and ensures that the
1.1. Motivation

**Figure 1.3:** Reserve costs as a function of wind penetration, from Ackermann et al. [AAD+07].

grid frequency is regulated back to its nominal value. Lastly, Replacement Reserves (or tertiary reserves) restore the margins within which the faster reserves operate, in order to ensure that future reserve activation does not come up against generator constraints. Fig. 1.4 shows a range of such products provided by market participants in several European countries.

There is a trend from automatic to manual activation of these reserves as the activation time increases from the top to the bottom of the table. The reliance on automatic controllers on fast timescales reflects the fact that security is the first priority in electricity provision (the costs of recent blackouts in Europe and the US have been catastrophic, and typical Value of Lost Load (VoLL) penalties are estimated to lie in the thousands of dollars per MWh [KA96], hundreds of times the average rate paid by consumers). Economic efficiency is only achievable on slightly longer timescales (5-15 minutes), usually by using a market mechanism to plan generation and consumption. This dissertation therefore focuses on improving system efficiency from this timescale up to horizons of order 24 hours.

### 1.1.3 Electricity market practice

Electricity provision began as a deregulated industry in the late 19th century, with some cities served by dozens of separate generation companies. It was quickly recognized, though, that electricity could be provided more efficiently by larger generation units and a single transmission and distribution infrastructure, i.e.
1. Introduction

natural monopoly. As a result, from 1907 in the US, state utility commissions appeared which would soon link up to manage huge synchronized grids ([Sto02], §1-1). The state-run utility became the worldwide norm until the 1990s, when countries such as the UK began deregulating electricity generation once more. The purpose of this was to promote efficiency by taking advantage of the fact that the interconnection of many generators over the same strong transmission network made a competitive environment increasingly viable. The transmission and distribution of the electricity itself, however, remain regulated natural monopolies. These issues are discussed in the introductory chapter of [Sto02].

Deregulation has resulted in several different trading practices for electricity. An important distinction between these is whether energy is traded bilaterally, between pairs of market participants, or in a pool, where all buyers and all sellers operate through a clearing house and the buyer of a particular seller’s energy cannot be identified. The pool model is common in the US, where an Independent System Operator (ISO) solves a constrained optimization problem to work out which participants’ bids are to be accepted. The shadow prices from this optimization are used to construct locational marginal prices (LMPs) for each node of the network,

Figure 1.4: Reserve products available in a selection of ENTSO-E countries [ENT12].
1.1. Motivation

an example of which is shown in Fig. 1.5. The pool model was also used in the United Kingdom until the complexity of that implementation led it to collapse in 2001 [Tov04]. Experience with the model in the US, which has several markets, has varied dramatically, with markets such as New England being reasonably competitive and successful while others, such as California and Montana, experienced disaster [BMS08]. One particular, though subjective, cause for concern is that the LMPs often appear to be unfairly distributed, in that two relatively nearby locations can experience large differences in electricity prices. An example of this effect is shown in Fig. 1.5.

The market approach that has emerged in much of Europe features large quantities of bilateral trades between participants, spot markets in which bids are not as strongly constrained to particular network nodes as in the case of the US, and a different treatment of reserves and balancing payments. An overview of the different mechanisms in Europe (as of 2012) governing spot markets, intraday markets, and reserves, is given in the ENTSO-E report [ENT12]. In contrast to the US, where LMPs are used as signals to market participants, Europe tends to favour price solidarity for political reasons. This leads to uniform pricing across a given country, or at least within large regions of a country as in the zonally-priced Nordic market, see Fig. 1.6.

1.1.4 Online optimization for power systems

An increase in poorly-predicted renewables means that the cost of providing adequate reserves is expected to grow to unacceptable levels in the coming years [Web10, New10]. To counteract this, there has been an increase in the sophistication of centralized decision-making in pool markets. One of the most forward-looking of system operators worldwide is the Pennsylvania-Jersey-Maryland (PJM) Interconnection in the United States. Online optimization based on the current system state has been used there to selectively activate market participants’ bids for the last decade, with sophisticated unit commitment algorithms and other heuristics used to improve market efficiency based on real-time information [Ott03, Ott10].

A common feature of such ISO optimizations is that they take as inputs bids from generators and consumers that are attached to specific locations on the grid. These are assumed to be a good representation of the real costs and utilities of market
1. Introduction

Figure 1.5: Locational Marginal Prices (LMPs) generated from a market-clearing optimization by the Midwest ISO at 04:45 local time on 1st November 2013. A high-price island (up to $90/MWh) in Iowa is surrounded by otherwise low prices (below $26/MWh). Immediately to the east of the high-price zone is an area where prices are near zero (dark blue). Both negative and extremely high LMPs are also commonplace. Source: http://www.misoenergy.org.
1.1. Motivation

Figure 1.6: Day-averaged zonal spot prices in Euro/MWh cleared for 3rd November 2013 in the Nordic EISpot Market, operated by Nord Pool. Similar prices between zones indicate that no congestion constraints are binding between them. Source: http://www.nordpools.com.
1. Introduction

participants, and therefore it is assumed that the outcomes of the optimizations performed by the ISO approximately minimize real system costs. However, when various constraints bind, such as when ramping limits of slower generators are hit or transmission lines become congested, three problems can arise:

1. Bids may fail to reflect true dynamic costs and constraints, and therefore when the ISO activates bids periodically in response to measurements of the network state, the outcome may be inefficient. For example, the Midwest ISO has for several years sought to improve the way it treats ramping decisions in real time [Eco13, Recommendations 10 and 11, p. (viii)]. A similar concern is shown by the California ISO [XT12].

2. Extreme price spikes (both positive and negative) can occur when the dynamically changing load and renewable in-feeds interact with the dynamic constraints of the market participants, for example due to ramping limitations of generators [WNPK+12].

3. Opportunities for market power abuse often arise, since only one or a few companies may physically be able to serve particular consumers. In these situations, bid prices may be inflated above cost, resulting in distorted prices that allow the abuser to extract large incomes in very short periods.

These problems all increase in severity as the penetration of intermittent renewables grows, because operational decisions must be made under uncertainty. This dissertation presents some contributions towards addressing problems 1 and 2 by proposing mechanisms that take account of this time coupling and exploit more granular prediction information.

1.2 Sources of uncertainty in power systems

This dissertation is concerned with the efficient accommodation of constantly-updating forecasts and granular characterizations of uncertainty in power system operations. The three potential sources of this uncertainty are now summarized.
1.2. Sources of uncertainty in power systems

1.2.1 Consumption

When aggregated over a large number of users, electricity consumption follows regular, fairly predictable patterns that are strongly correlated with the heating and cooling needs of the consumers. Consequently, a rather good prediction of the electrical load (errors in the range 1-4% at 24 hours ahead) can typically be obtained by incorporating forecasts of the air temperature and other weather effects for the following hours or days [PESMI+91]. For this reason the methods in Chapter 2 do not focus on load uncertainty, although they could be incorporated equally well into the model.

1.2.2 Intermittent renewables

In contrast to electrical loads, the contribution of intermittent renewables is far more difficult to forecast, even on relatively short time-scales of a few hours. A major cause of this difficulty is that even if the wind speed or incident radiation can be predicted well, the power injected into the grid is a highly nonlinear function of these drivers. In the case of wind, the instantaneous power produced is flat in some regimes and in others it is related to the cube of the wind speed. In the case of solar power it is dependent on the incident radiation level and the relative angle of the panels with respect to the sunlight. The potentially large difference between nominal and measured performance is shown for a wind farm in Fig. 1.7.

Accommodating wind or solar power efficiently therefore relies on a good prediction of how a very high-dimensional uncertain process (the weather system covering the grid in question) maps via the individual wind turbines or solar panels into power flows entering the network. For short time horizons of up to 2-6 hours, an extrapolation from the measured output of the renewable sources, known as a persistence model, is in fact the best available estimator of future power output, whereas for longer horizons a computationally intensive numerical weather prediction model is capable of providing better power output predictions. The reason for this difference is that numerical weather prediction models are generally poor at explaining current renewable power outputs, and therefore cannot provide especially accurate extrapolations of power output for the coming hours [Pat10].

Though forecast accuracy is improving, this improvement is being outstripped by the rate of growth of installed renewables, so that even only a few hours ahead of
1. Introduction

Figure 1.7: Above: Nominal output of a Vestas V44 600 kW turbine as a function of wind speed. Below: Measured output distribution of the Klim wind farm over six months. Source: [Pin13].

delivery the forecast contribution from wind power may have an error of multiple gigawatts [Klo10]. Studies such as [GSH07; PCK07; BDNL08; LSC+09; HLM+12] point to aggregated RMS errors of order 5-10% of total installed wind capacity over a time-scale of 6 hours, although forecast accuracy is highly dependent on weather conditions and the size of the aggregation area.
1.2.3 Contingencies

Load and renewable in-feeds present continuously varying uncertainty to the grid. However another category of uncertainty comes from events, or contingencies. These include the tripping of a generator, weather events causing a transmission line outage, or any other component failure. The list of possible events considered by system operators is typically of the order $10^4$ [CMRP+11], and many of these will result in the outage of multiple components.

Although lists of specific contingencies used by system operators are kept secret for security reasons, the different types of risk considered are published. For example, the contingency categories considered by the California ISO are as follows, in increasing order of severity [MD13]:

B: Loss of 1 generator; loss of one transformer; loss of one transmission line; loss of a single pole of a DC line; loss of one generator and one transmission line; loss of both poles of a particularly critical line.

C: Outage of a breaker or bus section; two sequential outages of a single outage with time for adjustment; loss of both poles of a DC line; double circuit tower line outage; stuck breaker plus Category B contingency; loss of two adjacent transmission lines on separate towers.

D: Loss of 2 nuclear units; loss of all generating units at a station; loss of all transmission on a common right-of-way; loss of substation; certain combinations of element and transmission line outages.

The operator’s response to events in each category of contingency is planned according to the severity and likelihood of the event, and this will affect the generation plan. Some contingencies are only permitted to bring about a transition to another safe network state, without allowing time for corrective action. Others are allowed to bring about an “unsafe” state as long as this is corrected within a short period on the order 10-30 minutes. The former is known as preventive security, and the latter corrective security [SA12].
1.3 Literature survey

This dissertation draws on recent work in optimal power flow, market design, and robust optimization. This section provides context for the contributions described in Chapters 2 and 5.

Subsection 1.3.1 summarizes advances in optimal power flow theory. Subsection 1.3.2 surveys relevant work in robust optimization and control, with a focus on recourse rules and adaptive solutions. Subsection 1.3.3 describes efforts at the interface of market mechanism design and distributed optimization, and 1.3.4 describes new pricing and market mechanisms for electricity networks and smart grids. Attention is restricted to areas contributing to the background of work described in later chapters. For topics too broad for a thorough review here, an up-to-date survey paper containing a more comprehensive list of references for that topic is given.

1.3.1 Optimal Power Flow

Optimal Power Flow (OPF) is the problem of satisfying electrical loads most efficiently, over a transmission network that poses constraints on the locations at which electrical power may flow. It has been studied academically since the mid-20th century. The problem of choosing which generators to switch on and off over a planning period is known as the Unit Commitment (UC) problem. An OPF problem may or may not include UC, and when it does not, it is often referred to as optimal dispatch. The binary decision variables needed to formulate a UC problem make it difficult to scale up to real-world (country or continent) instances, and the problem is especially hard once uncertainty is added. OPF with the constraint that system safety be assured under a list of contingency scenarios or in the presence of general uncertainty is known as Security-Constrained OPF, or Security-Constrained Unit Commitment.

Carpentier [Car62] and Dommel and Tinney [DT68] are considered to be pioneering authors on OPF. They provided the first rigorous definitions of the costs and nonlinear constraints that were needed to specify an OPF problem useful to system operators. Several methods based on Newton-type algorithms, such as Fletcher-Powell [Sas69] and BFGS [HI82], or gradient methods [AS74], emerged over the following decade. A typical feature of these was unpredictable convergence proper-
1.3. Literature survey

ties, because the exact implications of the nonlinear network constraints were not analyzed. Sequential quadratic or linear programming, interior point, and other more elaborate methods such as artificial neural networks, particle swarms, and fuzzy logic were also applied to the problem; these are reviewed in the surveys [MEHA99, PJ08]. A review of the evolving real requirements of OPF software for system operators is given in [SA12].

A promising recent advance of relevance to Section 6.1 of Chapter 5 of this dissertation is the discovery [BWFW08, LL12] that when the AC OPF problem is written as a linear or quadratic program with quadratic equality and inequality constraints, an often-tight semidefinite relaxation of the problem is available. This is obtained by writing the problem in terms of the outer product of the original complex nodal voltage variables and neglecting the fact that the resulting matrix must have rank 1 in order to be able to extract the global solution [LL12]. Instead, one finds that a rank-1 solution can very often be obtained from the rank-relaxed solution without further optimization. It has often been noted [LL12, Mol13] that the dual of this relaxation is computationally cheaper to solve than the primal problem. In the last two or three years, work has been done to extend the problem formulation [GT11] and to derive cases where the relaxation is tight [BGLC12] or fails [LMBD11]. Finding succinct conditions for this tightness still appears to be an open problem (investigated in detail in the recent dissertation of Molzahn [Mol13], which contains a more comprehensive account of recent efforts in this area), though it appears that in most practical cases a given solution obtained by relaxation can nevertheless be certified as globally optimal by confirming it satisfies a rank constraint.

The OPF problem including UC is a mixed-integer optimization problem with exponential worst-case complexity in the problem size. Although an exhaustive enumeration is sometimes possible for systems of relatively small size and was employed as early as the 1960s [HKH66, KSF66], it has been necessary to propose methods to solve this problem efficiently for large systems. These have included dual-based methods [WSK95, BLSJ83, PO13], cutting planes [MQ99], and simulated annealing [ZG90]. A review of these methods was written by Padhy [Pad04].

It is critical that OPF solutions be safe with respect to unforeseen events (called contingencies) that could disconnect or disable certain components. The system should remain stable and within its operating limits under the new steady state obtained, or should be receptive to an identified manual corrective action, under
1. Introduction

every scenario accounted for. This problem is referred to as Security-Constrained OPF (SCOPF), a thorough review of which was recently published by Capitanescu et al. [CMRP+11].

In real-world SCOPF, a common difficulty is that the number of possible contingencies is large, and the number is overwhelming once combinations of multiple simultaneous contingencies are counted. Solutions that are safe after the failure of any one line or generator are said to provide N-1 security [Car79]. To solve SCOPF under a large list of possible contingencies, a way of identifying and removing from consideration those contingencies that are not binding at the optimal solution is required. This typically requires an iterative approach to the problem, in which results may be refined by repeating the SCOPF optimization with constraints that have been modified based on the observed outcome. Due to their large number, it is often necessary to model post-contingency scenarios in a reduced manner; this approach is discussed in detail in [CMRP+11 §4.1].

1.3.2 Robust optimization with recourse

Power system reserves are a way of accommodating the various forms of uncertainty that impact on the grid by forcing devices such as generators to maintain capacity margins to respond to uncertain power requirements. A plan incorporating such reserves can be interpreted as a robust solution to an OPF problem. The following account reviews the background to the affinely-adjustable robust optimization methods applied in Chapter 2 of this dissertation.

The idea of immunizing, or reducing the sensitivity of, the solution of an optimization problem to uncertain cost or constraint data arose in the 1990s in the fields of Operations Research and design [MVZ95, Tag86]. Since then the number of potential applications of Robust Optimization (RO) has grown as the price of computing power has dropped (see [BS07] for a summary of this historical development). Much work to formalize mathematical methods for solving robust linear, quadratic, and conic programs was carried out by Ben-Tal and Nemirovski [BTN02] and El Ghaoui [EGOL98] in the late 1990s, and (particularly in the areas of discrete optimization and network flows) a few years later by Bertsimas and Sim [BS03, BPS04, BS04]. A newer authoritative text is the book by Ben-Tal et al. [BTEGN09].

Of particular relevance to Chapter 2 of this dissertation is the idea of an adaptive,
or adjustable, solution to a robust problem, meaning one that can be changed in a second stage once the values of uncertain parameters have been discovered. This allows constraints to be satisfied less conservatively, in applications where the idea makes sense [BTGGN04]. The concept is also referred to as recourse, or in the language of control systems a kind of “planned feedback”. The notion of rigorous plans that govern how uncertain future information should be acted on goes back to the linear decision rules (LDRs) of the 1950s, starting with Holt’s applications to factory management and macroeconomics [HMS55, Hol62]. These gave the first formulations in which future actions depended linearly on as-yet-unknown quantities (e.g. demand for a product). However it was not until recently, with the results of Guslitser and Ben-Tal [Gus02, BTGGN04] that optimal LDRs satisfying hard constraints for every realization of the uncertainty could be found efficiently. A specific form of relevance to this dissertation is the multi-stage decision rule, which makes use of the structure of sequentially-revealed disturbance information by allowing recourse at every step of a planning horizon. This has been applied to predictive control of constrained dynamical systems [Löf03, GKM06] and portfolio optimization [Cal08, RK12]. Control applications that have since exploited these results include building climate control [OPJ12] and power electronics [VSWM13].

Linear or affine decision rules (and variants such as segregated responses, which apply different LDRs to different segments or parameterizations of the uncertainty set [CZ09, WOS10]) appear to be by far the most tractable parameterization for adjustable RO solutions. For this reason there has been some interest in measuring the suboptimality of affine policies with respect to general policies, resulting in algorithms for bounding suboptimality given a specific instance of a stochastic or robust optimization problem [KWG09, BG12]. These were then applied to robust MPC [HGK11, VPGM13]. Despite this historical preference for linear adjustments, recent work has included the investigation of quadratic decision rules [Roc12], and of ways of adjusting the integer components of a robust discrete optimization problem [BG13].

1.3.3 Distributed optimization and pricing mechanisms

This part of the survey is intended as background for Chapter 5 of this dissertation, which describes methods for efficiently coordinating power system operation using distributed optimization methods. Note that this dissertation makes the as-
assumption that market participants are price takers, meaning that the additional
game-theoretic aspects behind some aspects of market mechanism design are not
considered. Rather, the focus is on price discovery mechanisms.

A social welfare maximization problem provides the motivation for every market.
The assumptions that go into modelling this problem usually include a separable
cost function representing the costs or utilities seen by each participant as a result of
its actions, plus a constraint that supply should match demand for each commodity
involved. In addition, economic or technical constraints such as production limits
are also typically imposed on the actions of participants. Market mechanisms
are procedures for finding good solutions to this welfare maximization problem.
This is commonly achieved using prices, or equivalently with the introduction of a
numéraire commodity with unit value representing money, though explicitly money-
free mechanisms also exist \[\text{NRTV07}\].

Early theories explaining how real markets converge to suitable clearing prices
arose in the Victorian era, with Léon Walras first formalizing the idea of price
discovery through a process he called *tâtonnement* \[\text{Wal54}\]. He assumed that
a coordinator, either real or virtual (representing the aggregated judgments of
all market participants) adjusts prices downward when supply exceeds demand,
and increases them when demand exceeds supply. The idea of making repeated
offers, or pledges, before a binding trade was also proposed by Edgeworth \[\text{Edg81}\],
although there is some disagreement on how he intended the intermediate rounds of
negotiation to work, owing to the fact that “\[h\]is literary style was elusive” \[\text{Wal73}\].
The issue of convergence to competitive equilibria under price adjustments was
investigated further by Samuelson, who in the 1940s investigated the links between
local stability (via comparative statics) and global stability of pricing algorithms
\[\text{Sam41}\], and Arrow (and others), who formalized in continuous time what is known
as the Uzawa algorithm. This was in effect one of the first recognizable dual ascent
models, proposing “a perfectly competitive market with the price adjustment rate
proportionate to excess demand” \[\text{AH58}\].

The idea of price negotiation was generalized in the later 20\textsuperscript{th} century with the
advent of modern optimization theory. In particular, the formalization of duality
and the Karush-Kuhn-Tucker conditions made the notion of efficient prices, at
which incentives to maximize profit coincide with those that maximize total social
welfare, more concrete. The dual function of an optimization problem is always
concave, and under certain convexity and regularity assumptions, the maximization
of this function yields a solution of the original welfare optimization problem (posed as a minimization problem). Therefore the gradient step rules whose convergence for generic convex optimization problems was proven by Shor [SKR85] (and several others before) could be used to derive dual (sub-)gradient methods with guarantees of asymptotic convergence to a dual optimum (i.e., efficient prices).

Dual- (or price-) based methods for solving optimization problems (summaries of which can be found in [Lem01] and Chapter 6 of [Ber99]) are made more or less difficult by the characteristics of this dual function. In particular, the choice of algorithm depends on whether the dual function can be evaluated for a given candidate price, or whether only the gradient, or worse still, a single element of the subgradient, can be obtained. The latter point depends on whether the dual is differentiable, which depends on the original optimization problem [Ber99, Proposition 6.1.1].

Price coordination is particularly useful for solving large, separable-cost problems with a mixture of local agent constraints and coupling constraints. In such problems, the coupling constraints can often be relaxed and replaced with an adjustable price. This approach is commonly referred to as Lagrangian relaxation, or dual decomposition, and its main attraction is that it preserves privacy of the cost function and local constraints. It has also found application in areas such as coordinated model predictive control (MPC) [JSJJ08]. In the worst case, though, such dual ascent methods must employ rather conservative price update rules in order for the coordinator to be able to guarantee convergence to optimality. This typically results in a very high number of iterations (thousands or tens of thousands), especially for poorly-conditioned problems.

One method that may be applicable to maximization of the dual function, and offer advantages over naive Lagrangian relaxation, is the Cutting Plane method, in which the (dual) function is approximated by supporting hyperplanes that accumulate over the iterations of the algorithm [Kel60]. The maximizer of this outer-approximated function is used as the next dual iterate. However, this method is problematic in that the maximizer may be unbounded until sufficient hyperplanes in certain orientations have accumulated. Modifications to the method such as ACCPM [Son86] or bundle methods [HUL96] may improve the behaviour of the approximation, and these find uses in various applications. However they still rely on an oracle that returns the dual value (“d(\lambda)”), and may therefore not be realistic for applications to pricing mechanisms, in which the market participants can
only be expected to report the cost-minimizing actions themselves ("$x^*(\lambda)$").

An alternative approach is to improve the way candidate dual values are updated, since for poorly-conditioned problems the steepest dual-ascent direction (which would normally be the direction in which the dual is modified) gives a change that points almost at right angles to the dual optimum ([Lem01], §4.1). Two possibilities proposed by Shor are the Ellipsoid Algorithm, which dilates the dual space according to obtain a better-conditioned ascent problem [SKR85], and the $r$-Algorithm [SZ74].

A more drastic way of overcoming the problem of slow convergence is to add a (usually quadratic) regularization or "prox" term to the objective function, in order to bring strict convexity to the problem (these methods are reviewed in [PB13]). One promising method employing such a penalty function is the Alternating Direction Method of Multipliers (ADMM) [GM75, GM76], recently revived in the light of new computational possibilities [BPC+10]. ADMM can be formulated to preserve the same privacy advantages as standard dual decomposition, but with dramatically better convergence properties. However it relies on passing a quadratic penalty to market participants in addition to the conventional price signal, which may not be considered appropriate in a market context. Particularly problematic is the fact that at convergence this penalty tends to zero, meaning that participants with knowledge of this fact have no real incentive to respond to it during intermediate iterations of the mechanism. It remains to be seen whether incentive-compatible mechanisms to discover efficient prices in an acceptable number of iterations can be designed without such constructs.

1.3.4 Novel electricity pricing mechanisms

There have been significant efforts to propose methods that improve the efficiency of power systems featuring either large amounts of intermittent renewable energy, or a significant ability to manipulate flexible devices such as household appliances or plug-in electric vehicles. The volume of work in this area is large, and many authors have described similar insights and case studies illustrating the role of energy storage in balancing wind power fluctuations, either from the point of view of a market participant or the system operator. However, the number that have proposed fundamental improvements to pricing and market coordination is more limited.
Although time-of-use pricing is a very old idea, the idea of dynamic spot pricing for electricity in which prices are computed according to the state of the network is newer, first surfacing in the 1970s and 1980s [Vic71, STK+80, CBS82]. Consumers receive a price that responds to the real cost of producing electricity at a given time of day, and it is argued that this allows their true preferences, or price elasticity characteristics, to be accounted for when supplying power. This in turn allows consumers who place a particularly high value on consuming electricity at a given time to do so, while others may deactivate some loads if the cost savings outweigh the inconvenience. An additional advantage over the “flat rate” model ([CBS82], §2), is that when there is a shortage of generation a utility company without a real-time pricing capability must indiscriminately ration power, cutting a mixture of both high- and low-priority loads. For real-time pricing to work according to this theory the prices must be communicated to market participants in real time without cost – this is a strong assumption, but one that has in recent years become increasingly plausible with the advent of cheap real-time metering and billing technology.

Alvarado and others considered whether real-time prices alone could ever be used to coordinate electricity generation [Alv03]. Jokić et al. proposed a distributed real-time price controller that dynamically acted on line overloads and deviations of the frequency from nominal in order to stabilize the system state at an optimal solution to the OPF problem [JLvdB09]. The approach is loosely related to continuous-time dual ascent methods (see Section 1.3.3) in that price is used as a control input that varies in response to system constraint violations, though the scheme additionally incorporates the dynamics of generators responding to these prices.

Bitar et al. have offered pricing mechanisms that allow users of electric vehicles to express their willingness to pay for their vehicle to be charged more promptly when connected to the mains [BL12]. This allows vehicle charging to be carried out more efficiently with constrained resources, e.g. limited substation capacity. The related problem of finding optimal policies governing how an aggregator of such vehicles should buy power and reserves for their charging under uncertainty was considered by Subramanian et al. [STB+12].

Other recent proposals are specific to the problem of integrating random wind energy. A common criticism of feed-in tariffs for renewables is that while they have successfully encouraged investment in those technologies, they have brought about a situation where renewables do not carry responsibility for the security problems they cause [KNB08]. In response to this, several authors have discussed
how wind energy producers should optimally participate in free electricity markets if they are instead exposed to balancing prices (as opposed to benefiting from feed-in tariffs) [Bre04, PCK07, BRK+12]. Baeyens et al. showed that it is in the interests of multiple such producers to operate jointly to reduce their expected imbalance costs [BBKP13].

More fundamental changes to the way electricity markets should operate have been proposed. Taylor et al. proposed consolidating reserve products that currently operate on different time-scales into a single product whose value is characterized by the co-state of a linear-quadratic regulation problem [TNCP13]. Varaiya and colleagues have proposed an approach to the acquisition of power by a system operator inspired by stochastic optimal control. They derived optimal policies for refining the purchase of power up to the point of delivery, which have a convenient analytic form under certain assumptions on the uncertainty [VWB11, RBWV12]. This idea was accompanied by the even more radical approach of allowing intermittent wind farm output to be sold as packages of “unreliable power” [VWB11, BPK+12] to reward consumers who are less sensitive to power availability.

Other authors have considered how the electricity market may be operated in a totally decentralized way using price negotiation without a central coordinator. These efforts are of relevance to Part III of this dissertation. An early example is Conejo and Aguado’s work, which derived a model in which market participants communicating only along the transmission lines were shown to reach a static OPF solution under convexity assumptions over a simplified network model [CA98]. More recently, the same principle has been applied using ADMM in order to take advantage of its good convergence properties [KCLB12], though at the expense of some modification to the original market interpretation.
Part II

Policy-Based Reserves for Power Systems
Chapter 2

Affine Reserve Policies

This chapter introduces the concept of affine reserve policies for accommodating large, fluctuating renewable in-feeds in power systems. The approach uses robust optimization with recourse to determine operating rules for power system entities such as generators and storage units. These rules, or policies, establish several hours in advance how these entities are to respond to errors in the prediction of loads and renewable in-feeds once their values are discovered. Affine policies consist of a nominal power schedule plus a series of planned linear modifications that depend on the prediction errors that will become known at future times. The chapter then describes how to choose optimal affine policies that respect the power network constraints, namely matching supply and demand, respecting transmission line ratings, and the local operating limits of power system entities, for all realizations of the prediction errors. Crucially, these policies are time-coupled, exploiting the spatial and temporal correlation of these prediction errors. Affine policies are compared with existing reserve operation under standard modelling assumptions, and operating cost reductions are reported for a multi-day benchmark study featuring a poorly-predicted wind in-feed. The chapter material is primarily taken from [WGMM13].

Section 2.1 gives a motivation for feeding statistical information on future uncertainty into power system decision-making. Section 2.2 describes the modelling assumptions used. Section 2.3 derives the policy optimization problem and justifies the restriction to affine policies. Section 2.4 reports on an illustrative case study, in which operating costs are reduced in the presence of a random wind in-feed. Section 2.5 discusses the results.
2. Affine Reserve Policies

In the remainder of this part of the dissertation, Chapter 3 will discuss the implications of incorporating reserve policies for the market and technical operations respectively, and Chapter 4 will describe three extensions to the mathematical formulation that bring the concept closer to incorporation with current operating practice.

2.1 Introduction

A key challenge in incorporating highly variable intermittent renewable energy sources into power systems is the need to maintain system integrity while making best use of the energy they provide, which comes at zero marginal cost. It is widely agreed that in the next few decades, as the share of wind power becomes very large, current techniques for accommodating wind variability will become sufficiently expensive that alternatives will be sought [AAD+07, New10].

Running power systems with very high wind penetration and without excessive frequency control costs, or resorting to curtailment of renewable output, requires intelligent use of the best available forecasts, at all times. In particular, any successful method for dealing with the high variability of renewables on intra-day time-scales (the only time-scales over which prediction errors are reasonably small [GSH07]) will require the following:

1. A forecast of future intermittent energy injections available at the time when control decisions are made.
2. Rules for acting on errors in this forecast when they are discovered.
3. Forecast error probability distributions and their correlations in both time and space over the grid.

Theoretical attention to these points has grown in the last few years as the share of wind power in several countries has grown [GBAR05, BG08, MCPR09, XHPR11, VMLA13]. In Morales et al. [MCPR09] a unit commitment integer programming problem was solved first, and then reserve margins were selected in a second stage based on the requirement that the actual reserve deployment be feasible for all the scenarios considered. These ideas have since been developed to provide probabilistic guarantees on transmission constraint satisfaction using a limited number of scenarios [VMLA13].
This chapter primarily considers how reserves could be operated more efficiently on
daily time-scale around any unit commitment decisions that have already been made,
though an extension to incorporate UC is later described in Section 4.1. In contrast with
existing literature, the approach employs a robustness formulation of the problem where
the bounds on the uncertainty are assumed to have been chosen according to probabilistic
criteria or are inherent, e.g. arising from wind farm capacities. The optimization results in
time-coupled policies, which are rules agreed in advance governing how individual power
system entities will respond to prediction errors affecting power system operation. Policies
therefore constitute a reserve mechanism, a concept that represents the main contribution
of this chapter. Reductions in the cost of operating reserves when policies are used in the presence
of a large uncertainty are reported, in comparison to reserve rules that do not make
use of policies.

This work was inspired by results on disturbance feedback policies from the control
literature. However, these are pre-dated by the concept of linear decision rules
(LDRs) from operations research, where current states, past data or future predictions
are combined linearly in order to make an operational decision [HMS55, Sil79].
Typically, though, LDRs were unable to deal rigorously with operating constraints,
and were not studied in much detail after the 1970s [KWG09].

In the last decade, however, LDRs have been revived as a means of solving constrained
optimization problems where the minimizer is allowed to be a function of the data
uncertainty [Gus02, BTGGN04, KWG09]. This has led to some new applications, for example
in portfolio optimization [RK12]. These solution methods were also shown to be applicable
to robust predictive control [Lof03, GKM06], where control policies with various dependences
on the disturbance have been studied as a means of respecting state and input constraints under uncertain system
dynamics. A recent application of this is intelligent building control [OPJ+12]. See the review in Chapter 1 for a more
detailed review of robust optimization leading to the state of the art.

Although optimal short-term operation of power systems, including reserves, has
been studied in various ways for decades [BLSJP83, WSK+95] as a variant of the
standard optimal power flow problem, affine policies have not until now been exploited
for real-time decision making in electricity provision under uncertainty, despite
their attractiveness for incorporating forecasts into power system operations.
To this end, systematic ways of using future wind prediction error measurements
are presented in order to reduce the average costs of reserve provision. The transmission network considered is restricted to the standard linearized model, since the exact robust problem for the full AC network appears to be extremely difficult to characterize, and may in any case be computationally prohibitive. The linearization used has been shown in real applications to produce good approximations of true AC power flows, and some post-hoc adjustment to the solutions is envisaged, in order to accommodate Ohmic losses and other unmodelled effects.

2.2 Power system model

This chapter considers the problem of optimal operation of an electrical network to satisfy loads in the presence of uncertainty. The uncertainty to be accommodated manifests itself in the form of random power in-feeds from renewables and fluctuating load requirements. Operating rules that apply only for a finite time into the future will be chosen, on the assumption that new rules will be determined before the chosen rules expire. This finite time, or planning horizon, is divided into $T$ discrete time steps, corresponding to the trading intervals (of length 5 minutes to 1 hour) over which electricity is traded on modern intra-day markets [Sto02]. The length of time horizon considered relevant to this work is up to 24 hours, after which predictions of renewable infeed are assumed to become too poor to incorporate into sophisticated decision-making rules, and unit commitment decisions are not yet fixed. A small worked example using the model outlined below can be found in [WGMM12b].

2.2.1 Participant model

Consider the actions of $N$, generic participating devices, connected to a transmission grid. Each participant $i$, for example a generator, load, or storage unit, injects power into or extracts power from a fixed location on the network, in two forms (one or both of which may be present for a given participant):

- **Inelastic power flows**, which cannot be influenced by control signals.
- **Elastic power flows**, which are determined by the result of an optimization over possible control actions.
2.2. Power system model

Inelastic power flows

The inelastic, or exogenous, injection or extraction of power for each participant $i$ is modelled as $r_i + G_i \delta$, with positive values denoting a net power injection. Its two components are a nominal prediction $r_i \in \mathbb{R}^T$ plus a linear function $G_i \in \mathbb{R}^{T \times N_i T}$ of entries of a random forecast error vector $\delta \in \mathbb{R}^{N_i T}$, whose value is to be discovered in the future. It has the form $\delta = [\delta_0', \ldots, \delta_{T-1}']'$, where each $\delta_k \in \mathbb{R}^{N_i}$. In other words, $N_i$ is the number of elements in the disturbance vector at a given time, and this vector is mapped to the exogenous power flows in the system at that time. If the prediction error $\delta$ turns out to be zero, then the net power injection at step $k$ will simply be $[r_i]_k$.

The forecast error $\delta$ is assumed to belong to a compact set $\Delta := \{\delta \mid S\delta \leq h\}$ with $h \in \mathbb{R}^q$ (though a formulation for a norm-bounded set such as an ellipsoid will also be given), which contains the origin and whose relative interior is non-empty. Although the error is random, it is assumed that adequate estimates of the mean prediction error $\mathbb{E}[\delta]$ and the second moment $\mathbb{E}[\delta\delta'] \in \mathbb{R}^{N_i T \times N_i T}$ are available. No other restrictions are placed on the probability distribution.

Elastic power flows

Elastic power flows are governed by a participant’s dynamics in conjunction with some pre-defined costs. They are described in standard state space form ([Mac02], Section 2.1). At time $k$, each participant $i$ has internal state $x^i_k \in \mathbb{R}^{n_i}$, where $n_i$ is the state dimension, and is governed by linear time-invariant dynamics, so that given an input $u^i_k$ at time $k$ the state at time $k+1$ is given by $x^i_{k+1} = \tilde{A}_i x^i_k + \tilde{B}_i u^i_k$, where $\tilde{A}_i \in \mathbb{R}^{n_i \times n_i}$ and $\tilde{B}_i \in \mathbb{R}^{n_i}$. The first element $[x^i_k]_1$ of the state vector $x^i_k$ is assumed to represent the current power injection at the relevant node of the transmission network, and other elements are used to model internal dynamics or memory of previous states. The scalar input $u^i_k \in \mathbb{R}$ controls the net power injection of the participant at time $k + 1$.

Assigning the current time the value $k = 0$, a vector of future states for participant $i$, $x^i := [x^i_1' \ldots x^i_T']' \in \mathbb{R}^{n_i T}$, can be written as a function of the input sequence $u^i := [u^i_0' \ldots u^i_{T-1}']' \in \mathbb{R}^T$ and the current state $x^i_0$:

$$x^i = A_i x^i_0 + B_i u^i,$$

(2.1)
2. Affine Reserve Policies

where

\[
A_i := \begin{bmatrix}
\tilde{A}_i \\
\tilde{A}^2_i \\
\vdots \\
\tilde{A}^T_i
\end{bmatrix}, \quad B_i := \begin{bmatrix}
\tilde{B}_i & 0 & \cdots & 0 \\
\tilde{A}_i \tilde{B}_i & \tilde{B}_i & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
\tilde{A}^{-1}_i \tilde{B}_i & \cdots & \tilde{A}_i \tilde{B}_i & \tilde{B}_i
\end{bmatrix}.
\]

The vector of outputs \(C_i x_i\) as seen by the network is just the power injected or consumed by the participant. Therefore each matrix \(C_i \in \mathbb{R}^{T \times n_i T}\) selects only the first element of the state vector at each time, i.e. \(C_i = I_T \otimes \tilde{C}_i\), where \(\tilde{C}_i = [1 \ 0_{1 \times (n_i - 1)}]\).

Costs

The function \(J_i : \mathbb{R}^{n_i T} \times \mathbb{R}^T \rightarrow \mathbb{R}\) is used to define costs for the states and inputs along the time horizon. The common assumption [MEHA99] is made that costs can be modelled using a quadratic function,

\[
J_i(x^i, u^i) := f_{x^i}^T x^i + \frac{1}{2} x'^i H^x_i x^i + f_{u^i}^T u^i + \frac{1}{2} u'^i H^u_i u^i + c_i,
\]

where the Hessian matrices \(H^x_i\) and \(H^u_i\) are assumed to be positive semi-definite, in order for the optimization problem defined in Section 2.3 to be convex. Linear components \(f_{x^i}\) and \(f_{u^i}\) are of the form \(1_{T \times 1} \otimes \tilde{f}_{x^i}\) and \(1_{T \times 1} \otimes \tilde{f}_{u^i}\) respectively, where \(\tilde{f}_{x^i} \in \mathbb{R}^{n_i}\) and \(\tilde{f}_{u^i} \in \mathbb{R}\). Similarly the quadratic components are given by \(H^x_i := I_T \otimes \tilde{H}^x_i\) and \(H^u_i := I_T \otimes \tilde{H}^u_i\), where \(\tilde{H}^x_i \in \mathbb{R}^{n_i \times n_i}\) and \(\tilde{H}^u_i \in \mathbb{R}\). Coupling costs between time steps can be represented by augmenting the state vector to include a memory of prior states in the added states. Constant \(c_i := T \tilde{c}_i\) allows for a constant stage cost \(\tilde{c}_i\).

Constraints

The set \(Z_i\) consists of permissible combinations of state and input sequences \(x^i\) and \(u^i\) for participant \(i\), which may in some cases be additionally constrained by \(\delta\). It is a compact set defined by \(l_i\) linear inequalities (i.e. a polytope) and takes the form

\[
Z_i := \left\{ \begin{bmatrix} x^i \\ u^i \\ \delta \end{bmatrix} \mid T_i x^i + U_i u^i + V_i \delta \leq w_i \right\},
\]
where \( T_i \in \mathbb{R}^{l_i \times n_i}, U_i \in \mathbb{R}^{l_i \times T}, V_i \in \mathbb{R}^{l_i \times N \delta} \) and \( w_i \in \mathbb{R}^{l_i} \). Of course, \( x^i \) and \( u^i \) are related by the dynamic equation \( (2.1) \), and \( T_i, U_i, V_i, \) and \( w_i \) may also depend on the current state \( x_{0i}^i \); these dependences are left out of the notation above for clarity. A generator with generation set-points limited to the range \([p_{\text{min}}, p_{\text{max}}]\), for instance, could be modelled with \( T_i = 0, U_i = [IJ], V_i = 0, w_i = [p_{\text{max}} - 1 - p_{\text{min}} - 1] \). Note that since \( x^i \) and \( u^i \) are trajectories rather than state or input vectors corresponding to a single time, a wide range of constraints coupling states and inputs can be modelled. For example, ramp rates may be imposed on generators, and empty/full constraints may be imposed on storage units.

Usually \( V_i = 0 \), unless the uncertainty feeds directly into the participant’s operating constraints. An example of this would be a curtailable wind farm whose maximum power availability at any given time is uncertain, and whose power output could be varied at any time between zero and this upper limit. In this study, though, wind curtailment is assumed to be undesirable.

Binary decision variables, which would be needed to model start-up and shut-down (unit commitment) decisions for generators, are for the sake of clarity not considered in this section, since it is assumed that such decisions have been made at an earlier stage. However a mathematical formulation and results with this extension will be described in Section 4.1, in which the formulation is extended using a mixed logical-dynamical system description [BM99].

### 2.2.2 Network model

The network model is a standard linearized approximation of a high-voltage transmission grid [CWW00], in which lines are lossless, voltage magnitudes are constant, and line flows are proportional to the phase differences (assumed to be small) between nodal voltages. Let each participant be connected to one of \( N_n \) network nodes, and let \( L \) be the number of lines connecting these nodes.

The network imposes two constraints on power system operation. The first is that the net power injection, comprising the sum of inelastic flows \([r_i + G_i\delta]_k\) and elastic flows \([C_i x^i]_k\), has to be zero at all times \( k = 1, \ldots, T \). This can be modelled using an equality constraint with \( T \) rows:

\[
\sum_{i=1}^{N_p} (r_i + G_i\delta + C_i x^i) = 0. \tag{2.4}
\]
The second constraint is that line currents cannot exceed the respective line ratings anywhere on the network, at any time. Under the preceding assumptions, this constraint is linear in the power injections, as long as the net power injection into the network is zero (i.e. condition (2.4) holds) [CWW00]. It can be represented by the vector inequality

\[
\sum_{i=1}^{N_p} \Gamma_i (r_i + G_i \delta + C_i \delta^i) \leq p.
\] (2.5)

This has \(2LT\) rows, one for each flow direction, for each line, at each time. Each matrix \(\Gamma_i \in \mathbb{R}^{2LT \times T}\) maps the power output of the node to which participant \(i\) is attached to contributions to line flows. Each \(\Gamma_i\) can be constructed from the network line impedances base on the technique used to arrive at equation (III.4) in [CWW00]; see the Appendix for the derivation, which involves eliminating the nodal voltage phase angles from the power flow equations. Each matrix \(\Gamma_i\) has a block diagonal structure, since power outputs at one time cannot affect line flows at another time. The vector \(\bar{p} \in \mathbb{R}^{2LT}\) contains the appropriate stack of line rating values.

### 2.3 Optimal affine reserve policies

Current reserve mechanisms use a cascaded loop structure, where the fastest (primary) controller stabilizes the grid frequency, a minutes-scale (secondary) controller corrects it back to its reference, and slower, separately-purchased tertiary reserves re-dispatch generators in order to free up the margins within which the faster control operates [Sto02]. This reserve action is only a real-time response to the error as it unfolds, and makes no systematic use of what is expected to happen in future. Intra-day electricity markets in many countries are currently experiencing dramatically increasing trade volumes [Web10], and this increase can be seen as an attempt to adjust the short-term economic operation of the power system in the light of new forecast information. Different countries operate these markets in diverse ways (an overview of the various mechanisms used to acquire reserves and short-term power commitments in European countries can be found in [ENT12]). Such trading actions take little systematic account of time-coupled costs and constraints imposed on the market participants.

This section describes a more systematic predictive mechanism that explicitly takes
2.3. Optimal affine reserve policies

account of short-term uncertainties with the aim of reducing the expected running costs of the power system over the time horizon.

2.3.1 Finite horizon optimization

Consider the problem of minimizing expected running costs

\[ \sum_{i=1}^{N_p} \mathbb{E}[J_i(x^i, u^i)] \]

over a horizon of length \( T \), subject to the local constraints (2.3) and network constraints (2.4) and (2.5). This is done by choosing a sequence of control inputs \( u^i \) for each participant \( i \) that can vary with \( \delta \). The best causal response to prediction errors is desired, a policy \( u^i = \pi_i(\delta) \), where \( \pi_i : \mathbb{R}^{N_p T} \rightarrow \mathbb{R}^T \) is to be chosen before the error is known. “Causal” means that \( u^i_m \) can depend only on the measurements of \( \delta_0, \delta_1, \ldots, \delta_m, \delta_{m+1} \). That is, it is assumed that \( \delta_{m+1} \), the sub-vector of \( \delta \) pertaining to time \( m+1 \), is revealed just before input \( u^i_m \) is applied. Obviously, a dependence on any of \( \delta_{m+2}, \ldots, \delta_{T-1} \) would violate causality because \( u^i_m \) would be a function of information unavailable at time \( m \).

Substituting \( u^i = \pi_i(\delta) \) into the state update equation (2.1) and eliminating \( x^i \), the following finite horizon optimization problem is obtained:

\[
\min_{\text{Causal } \pi_i} \sum_{i=1}^{N_p} \mathbb{E}[J_i(A_i x^i_0 + B_i \pi_i(\delta), \pi_i(\delta))] \quad (2.6a)
\]

s.t. \( \sum_{i=1}^{N_p} r_i + G_i \delta + C_i(A_i x^i_0 + B_i \pi_i(\delta)) = 0, \forall \delta \in \Delta \), \( (2.6b) \)

\[
\sum_{i=1}^{N_p} \Gamma_i(r_i + G_i \delta + C_i(A_i x^i_0 + B_i \pi_i(\delta))) \leq p, \forall \delta \in \Delta, \quad (2.6c)
\]

\[
\begin{bmatrix}
A_i x^i_0 + B_i \pi_i(\delta) \\
\pi_i(\delta)
\end{bmatrix} \in Z_i, \forall \delta \in \Delta. \quad (2.6d)
\]

Constraints (2.6b) and (2.6c) are the expanded forms of (2.4) and (2.5) after substituting definitions of \( x^i \) and \( u^i \), so that the optimization is expressed only in terms of \( \pi_i \).

This problem is intractable due to the wide variety of candidate functions \( \pi_i \) that could satisfy the constraints. The policies are therefore restricted from now on to the affine form

\[ u^i = D_i \delta + e_i, \quad (2.7) \]
2. Affine Reserve Policies

Figure 2.1: Affine reserve policy, consisting of a nominal schedule $e_i$ (left) plus a matrix $D_i$ (right) which multiplies yet-to-be-revealed disturbances in a causal manner.

so that participant $i$’s power schedule $u_i^t$ is parametrized by a nominal schedule $e_i = [e_i^0 \ldots e_i^{T-1}]'$ plus a linear variation $D_i$ with future prediction errors. In order for the use of future disturbances to be causal, $D_i$ takes the lower-triangular form

$$D_i = \begin{bmatrix}
[D_i]_{0,0} & 0 & \cdots & 0 \\
[D_i]_{1,0} & [D_i]_{1,1} & \ddots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
[D_i]_{T-1,0} & \cdots & [D_i]_{T-1,T-2} & [D_i]_{T-1,T-1}
\end{bmatrix}$$

where $[D_i]_{l,m} \in \mathbb{R}^{1 \times N_d}$ is the response of input $u_i^l$ to error $\delta_{m+1}$. The presence of non-zero blocks above the diagonal would violate causality, because the values of as-yet-unknown errors would contribute to the control rule. The form considered is shown in Fig. 2.1.

Constraint (2.6d) and the causality requirement are rewritten for compactness as a set of admissible policies $(D_i, e_i)$ parametrized by the current state $x_i^0$:

$$F_i(x_i^0) = \left\{ \begin{array}{c}
(D_i, e_i) \\
[D_i]_{l,m} = 0, \quad \forall m > l \\
Ax_i^i + B_i(D_i\delta + e_i) \\
D_i\delta + e_i \\
\delta \in \Delta, \forall \delta \in \Delta
\end{array} \right\}$$

This leads to the following rewriting of problem (2.6) in terms of nominal schedules
\( \{ e_i \}_{i=1}^{N_P} \) and policy matrices \( \{ D_i \}_{i=1}^{N_P} \):

\[
\min_{(D_i, e_i) \in F_i(x_0^i)} \sum_{i=1}^{N_P} \tilde{J}_i(x_0^i, D_i, e_i) \tag{2.8a}
\]

\[
\text{s.t. } \sum_{i=1}^{N_P} r_i + G_i \delta + C_i(A_i x_0^i + B_i(D_i \delta + e_i)) = 0, \forall \delta \in \Delta, \tag{2.8b}
\]

\[
\sum_{i=1}^{N_P} \Gamma_i(r_i + G_i \delta + C_i(A_i x_0^i + B_i(D_i \delta + e_i))) \leq \bar{p}, \forall \delta \in \Delta. \tag{2.8c}
\]

The objective function has been redefined as \( \tilde{J}_i(x_0^i, D_i, e_i) := \mathbb{E}[J_i(x^i, u^i)] \) to reflect its dependence on \( x_0^i, D_i, \) and \( e_i \).

The assumption of a positive semidefinite quadratic form \( \langle 2.2 \rangle \) for \( J_i(x^i, u^i) \) allows the expectation over \( \delta \) to be minimized straightforwardly in comparison to the case of an arbitrary non-linear cost function, since only the moments \( \mathbb{E}[\delta] \) and \( \mathbb{E}[\delta^2] \) are needed. Substitution from \( \langle 2.2 \rangle \) gives the following representation of the objective, which is convex:

\[
\tilde{J}_i(x_0^i, D_i, e_i) = \mathbb{E}[J_i(A_i x_0^i + B_i(D_i \delta + e_i), D_i \delta + e_i)]
\]

\[
= f_i^x x_0^i + B_i e_i + f_i^w e_i + x_0^i A_i^T H_i^w B_i e_i 
+ \frac{1}{2} x_0^i A_i^T H_i^w A_i x_0^i + \frac{1}{2} e_i (B_i^T H_i^w B_i + H_i^w) e_i 
+ (f_i^y B_i + f_i^w + x_0^i A_i^T H_i^w B_i + e_i^T B_i^T H_i^w B_i + e_i^T H_i^w)) D_i \mathbb{E}[\delta]
+ \frac{1}{2} (D_i^T (B_i^T H_i^w B_i + H_i^w) D_i, \mathbb{E}[\delta^2]) + e_i
\]

The notation \( \langle A, B \rangle \) is used to represent the trace of \( A'B \). Since in many cases the reference predictions \( r_i \) will be chosen with \( \mathbb{E}[\delta] = 0 \), the corresponding term above often cancels. Note that although the constant term \( c_i \) makes no difference to the solutions \( D_i \) and \( e_i \), it is needed in order to compare the costs of different approaches.

### 2.3.2 Equivalent tractable reformulation

Problem \( \langle 2.8 \rangle \) cannot be solved directly because constraints \( \langle 2.8b \rangle \) and \( \langle 2.8c \rangle \), as well as the definition of \( F_i(x_0^i) \), apply for all \( \delta \in \Delta \) and are therefore the intersection of an infinite number of constraints. To obtain a numerical solution they must be written in an equivalent finite form.
If $\Delta$ is assumed to contain the origin and have non-empty interior, implying $\delta$ is not trivially constrained in any dimension, it can be shown that constraint (2.8b) is satisfied if and only if the following conditions hold:

\[
\sum_{i=1}^{N_p} (r_i + C_i A_i x_{0i} + C_i B_i e_i) = 0, \quad (2.9a)
\]

\[
\sum_{i=1}^{N_p} (G_i + C_i B_i D_i) = 0. \quad (2.9b)
\]

Constraint (2.8c) and the sets $F_i(x_{0i})$ can be written using a result due to Guslitser and others [Gus02, BTGGN04]. Recalling that $\Delta = \{ \delta \mid S\delta \leq h \}$, the following equivalences hold for (2.8c):

\[
\sum_{i=1}^{N_p} \Gamma_i (r_i + G_i \delta + C_i (A_i x_{0i} + B_i (D_i \delta + e_i))) \leq \overline{p}, \quad \forall \delta \in \Delta
\]

\[
\Rightarrow \max_{\delta \in \Delta} \sum_{i=1}^{N_p} \Gamma_i (G_i + C_i B_i D_i) \delta + \sum_{i=1}^{N_p} \Gamma_i C_i B_i e_i \leq \overline{p} - \sum_{i=1}^{N_p} \Gamma_i (r_i + C_i A_i x_{0i})
\]

\[
\Rightarrow \exists Z: Z'h + \sum_{i=1}^{N_p} \Gamma_i C_i B_i e_i \leq \overline{p} - \sum_{i=1}^{N_p} \Gamma_i (r_i + C_i A_i x_{0i}), \quad \sum_{i=1}^{N_p} \Gamma_i (G_i + C_i B_i D_i) = Z'S, \quad \text{and } Z \geq 0 \text{ element-wise.}
\]

An extra matrix variable $Z$ is introduced for the last equivalence, which uses strong duality in linear programming (see [GKM06], Example 7).

Similarly, it can be shown that the sets $F_i(x_{0i})$ can be rewritten in finite form as follows, starting from definition (2.3) and introducing extra matrix variables $Y_i$ of appropriate dimension:

\[
F_i(x_{0i}) = \left\{ (D_i, e_i) \mid \begin{array}{l}
[D_i]_{l,m} = 0, \quad \forall m > l \\
\exists Y_i \geq 0: (T_i B_i + U_i) D_i + V_i = Y_i'S,
\end{array} \right. \\
T_i A_i x_{0i} + (T_i B_i + U_i) e_i + Y_i'h \leq w_i
\]

These changes lead to the following tractable representation of optimization problem (2.8):

\[
\min_{Z \geq 0, (D_i, e_i) \in F_i(x_{0i})} \sum_{i=1}^{N_p} j_i(x_{0i}, D_i, e_i)
\]
In summary, (2.10b) states that the nominal schedules of power output changes $e_i$ should track the base prediction; (2.10c) states that the rules $D_i$ used by the participants should together track any error vector $\delta \in \Delta$; (2.10d) and (2.10e) ensure that line current limits should not be exceeded for any $\delta \in \Delta$.

After solving (2.10) the state and input trajectories $x^i$ and $u^i$ for a particular prediction error $\delta$ can be computed by substituting the solution $(D_i, e_i)$ back into (2.7) and (2.1).

Formulation for an ellipsoidal uncertainty set

A variant of problem (2.10) can be obtained straightforwardly for the case of an ellipsoidal uncertainty set of the form

$$\Delta_{\text{ell}} := \{ \delta | \delta = E_{\text{ell}}d + f_{\text{ell}}, ||d||_2 \leq 1 \},$$

(2.11)

where $E_{\text{ell}} \in \mathbb{R}^{N_S \times l}$ and $f_{\text{ell}} \in \mathbb{R}^{N_S T}$ define the scaling and translation of the set and $d \in \mathbb{R}^l$ is an underlying bounded vector. The derivation from Example 8 of [GKM06] can be adapted as follows (noting that a related derivation is available for any $p$-norm-bounded set $\Delta$ for $p \in [1, \infty]$):

$$\sum_{i=1}^{N_p} \Gamma_i (r_i + G_i \delta + C_i (A_i x^i_0 + B_i (D_i \delta + e_i))) \leq \bar{p}, \forall \delta \in \Delta_{\text{ell}}$$

$$\max_{\delta \in \Delta_{\text{ell}}} \sum_{i=1}^{N_p} \Gamma_i (G_i + C_i B_i D_i) \delta + \sum_{i=1}^{N_p} \Gamma_i C_i B_i e_i \leq \bar{p} - \sum_{i=1}^{N_p} \Gamma_i (r_i + C_i A_i x^i_0)$$
2. Affine Reserve Policies

\[
\begin{align*}
&\left\| E_{\text{cell}} \sum_{i=1}^{N_p} \Gamma_i (G_i + C_i B_i D_i) \right\|_2 + \sum_{i=1}^{N_p} [\Gamma_i (G_i + C_i B_i D_i)] f_{\text{cell}} + \sum_{i=1}^{N_p} \Gamma_i C_i B_i e_i \\
&\leq \bar{p} - \sum_{i=1}^{N_p} \Gamma_i (r_i + C_i A_i x_0^i)
\end{align*}
\]

The new constraint formulation can be posed as a second-order cone program, which can be handled by many solvers. Note that the local constraint set \( F_i(x_0^i) \) should also be modified to provide robustness to the new uncertainty set.

2.3.3 Computational requirements

Problem \((2.10)\) is a quadratic program, which in principle can be solved even where many thousands of variables and constraints are present. However some comment on the size and structural properties of the problem should be made. Each vector \( e_i \) has \( T \) elements, each matrix \( D_i \) has \( N_\delta T^2 \) elements (neglecting the fact that some of these are constrained to be zero), and each matrix \( Y_i \) (introduced by definition of the sets \( F_i(x_0^i) \)) has \( q l_i \) elements, recalling that \( q \) is the number of constraints defining \( \Delta \) and \( l_i \) is the number of constraints defining \( Z_i \). The matrix \( Z \) has \( 2qLT \) elements. Therefore the total number of primal optimization variables is \( N_p (T + N_\delta T^2 + q\bar{l}) + 2qLT \), where \( \bar{l} = \frac{1}{N_p} \sum_{i=1}^{N_p} l_i \). A significant contributor to computational cost is the fact that the optimization is over the lower-triangular parts of the matrices \( D_i \), which together contain roughly \( \frac{1}{2} N_p N_\delta T^2 \) decision variables. The problem size therefore grows quadratically with the time horizon, and already reaches the thousands for modest parameter choices. In contrast, the number of decision variables needed to operate reserves in a way more comparable with existing mechanisms (see Section \(2.4.3\)) grows only linearly with the time horizon.

However, for computational purposes the structure of the problem is as important as the size. The form of \((2.10)\) exhibits a convenient linear coupling between participants \( i = 1, \ldots, N_p \) in both the cost function and the constraint set, and in fact lends itself to solution via a large-scale distributed solution method, such as the recently-revived Alternating Direction Method of Multipliers \([BPC+10]\). A description of this extension is deferred to Section \(4.2\), since for the case study of Section \(2.4\) the resulting optimization problems were still manageable enough for
2.4 Numerical case study

Policy-based reserves were applied to a standard test network that was adapted to include a large share of wind power. Policies were recomputed in a receding horizon fashion (various ways of implementing reserve policies in real time will be described in Section 3.1.4) over a three-day simulated period in order to assess the reserve costs incurred. The test was repeated 50 times with different realizations of the random wind in-feed. The effects of restrictions on the structure of the matrices $D_i$ on the cost of reserves were measured, leading to observations on the cost savings facilitated by recourse. The optimization problems were solved using CPLEX [1107].

2.4.1 System model

The network used was a modification of the 39 bus network described in Appendix A of [Pai89], and shown in Fig. 2.2. This network contains 7 thermal generators, 2 storage units, 19 loads, and 3 wind farms (which replace 3 of the original 10 generators). It was assumed that the generators represent the plants that had been selected for use via an earlier unit commitment decision. Generators have fuel costs represented in $f_i^u$ and $H_i^u$, and ramping costs represented in $H_i^r$. Storage units have a penalty for deviating from their midpoint, represented in $c_i, f_i^s$, and $H_i^s$. The parameters in terms of the definitions in Section 2.2.1 are described in Table 2.1.

The daily pattern of load variation shown with the thick black line in Fig. 2.4 was taken from data for total UK national electrical consumption\(^1\) on 14th September.

\(^1\)available at http://www.bmreports.com
2. Affine Reserve Policies

Figure 2.2: 39 bus test network from [Pai89], with wind in-feed replacing thermal generators at nodes 32, 33, and 34, and with added storage units at nodes 1 and 28.

2012, normalized to the size of each load modelled. Peak load was 6.097 GW, and wind power provided 3 GW at maximum, with an expected energy share of 29.0%. Line flows from bus 16 to 15 and from bus 16 to 17 were restricted to 1000 MW. The load sizes $p_{\text{nom}}$ in Table 2.1 are the nominal values described in [Pai89]. The three-day simulation period was divided into 288 fifteen-minute steps ($\tau = 0.25$ hrs), and the horizon length was $T = 8$.

2.4.2 Uncertainty model

Uncertainties in the system were assumed to originate only in the random wind power availability (loads were assumed to be predicted exactly, though the method could equally be used to account for load uncertainty). The wind farm output was driven by the following first order random process model with saturation:

$$q_{k+1} = \min\{\max\{q_{\text{min}}, q_k + \beta_k\}, q_{\text{max}}\},$$  \hspace{1cm} (2.12)

where $q_k \in \mathbb{R}^{N_{\delta}}$ denotes the state of the uncertainty model, and $\beta_k$ is sampled at each step $k$ from a multivariate normal distribution with variance $\Sigma \in \mathbb{R}^{N_{\delta} \times N_{\delta}}$, 42
Section 2.2.1. The prediction error was defined as $\delta E$ from $q$ truncated to bounds $A$, so that $r_i = G_iE[q]$, where $G_i$ is the same matrix as that described in Section 2.2.1. The prediction error was defined as $\delta := q - E[q]$ so that $E[\delta] = 0$.

<table>
<thead>
<tr>
<th>Table 2.1: Parameters of elastic and inelastic participants</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Thermal generators, $i = 1, \ldots, 7$:</strong></td>
</tr>
<tr>
<td>States: Current output (MW), $A_i = \begin{bmatrix} 0 &amp; 0 \ 1 &amp; 0 \end{bmatrix}$, $B_i = \begin{bmatrix} 0 \ 0 \end{bmatrix}$, $C_i = [0, 1]$, $\hat{f}_i = \begin{bmatrix} 0 \ 0 \end{bmatrix}$, $\hat{H}_i = \begin{bmatrix} \alpha &amp; -\alpha \ -\alpha &amp; \alpha \end{bmatrix}$, $\bar{c} = 0$, $x_0 = \begin{bmatrix} p_0 \end{bmatrix}$</td>
</tr>
<tr>
<td>Constraints, $\forall k$: $0 \leq [x_k]^1 \leq p_{\text{max}}$, $0 \leq [x_k]^2 \leq p_{\text{max}}$, $0 \leq u_k^{i} \leq p_{\text{max}}$</td>
</tr>
<tr>
<td>$i$</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
</tbody>
</table>

<p>| <strong>Storage units, $i = 8, 9$:</strong> |
| States: Current output (MW), $A_i = \begin{bmatrix} 0 &amp; 0 \ 1 &amp; 0 \end{bmatrix}$, $B_i = \begin{bmatrix} 0 \ 0 \end{bmatrix}$, $C_i = [0, 1]$, $\hat{f}<em>i = \begin{bmatrix} 0 \ 0 \end{bmatrix}$, $\hat{H}<em>i = \begin{bmatrix} 0 &amp; 0 \ 0 &amp; 0 \end{bmatrix}$, $\bar{c} = 0$, $x_0 = \begin{bmatrix} p_0 \end{bmatrix}$ |
| Constraints, $\forall k$: $0 \leq [x_k]^1 \leq p</em>{\text{max}}$, $0 \leq [x_k]^2 \leq p</em>{\text{max}}$, $0 \leq u_k^{i} \leq p_{\text{max}}$ |</p>
<table>
<thead>
<tr>
<th>$i$</th>
<th>$s_{\text{max}}$</th>
<th>$\gamma$</th>
<th>$p_{\text{max}}$</th>
<th>$s_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>1000</td>
<td>0.01</td>
<td>200</td>
<td>500</td>
</tr>
<tr>
<td>9</td>
<td>1000</td>
<td>0.01</td>
<td>200</td>
<td>500</td>
</tr>
</tbody>
</table>

| **Wind farms [Node: $G_i$] 32: [2 0], 33: [0 2], 34: [1 1]** |

truncated to bounds $A_\beta b_k \leq b_\beta$. Note that more elaborate wind models exist [PK08], but a simpler model was used here for demonstration purposes.

Defining $q := [q_1 \ldots q_P]$, as the random future evolution of $q$ from current state $q_0$, the nominal predictions of wind farm power output $r_i$ were mapped linearly from $E[q]$, so that $r_i = G_iE[q]$, where $G_i$ is the same matrix as that described in Section 2.2.1. The prediction error was defined as $\delta := q - E[q]$ so that $E[\delta] = 0$. 

43
2. Affine Reserve Policies

Figure 2.3: Example output of correlated uncertainty model \( q_k \), which drives the wind farm outputs in the case study. The solid line represents \([q_k]_1\) and the dotted line \([q_k]_2\).

The prediction error covariance is then \( \mathbb{E}[\delta \delta'] = \mathbb{E}[(q - \mathbb{E}[q])(q - \mathbb{E}[q]')\] \). This was supplied together with the references \( r_i \) and the current system state \( x_0 \) as inputs to optimization problem (2.10).

The uncertainty set \( \Delta \) was recomputed as a function of the current state, to reflect the fact that prediction errors are bounded relative to the nominal predictions by the wind farm power output limits and by the bounds assumed on \( \beta_k \), the change in wind power availability at each step. Its new parameters \( S \) and \( h \) were then supplied to (2.10). In addition, at every simulation time step, an aggregation of 20,000 Monte Carlo runs was used to produce \( T \)-step nominal predictions \( r_i \) for each wind farm, as well as estimates of \( \mathbb{E}[\delta \delta'] \).

The three wind farms in this case study are driven by two temporally and spatially correlated sources of uncertainty, with parameters \( \Sigma = \begin{bmatrix} 2400 & 2000 \\ 2000 & 2400 \end{bmatrix}, \ q_{\text{min}} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \ q_{\text{max}} = \begin{bmatrix} 500 \\ 500 \end{bmatrix}, \ A_\beta = \begin{bmatrix} I \\ 0 \end{bmatrix}, \ b_\beta = 80 \cdot 1 \). The matrices \( G_i \) for the wind farms are of the form \( G_i = I_T \otimes G_i \), and the matrices \( \tilde{G}_i \) are given in Table 2.1. The output of the wind farm at node 34 is the mean of the two at nodes 32 and 33. The purpose of this formulation is to show that random power flows can be driven by a process with dimension lower than the number of nodes affected, either to save computational effort, or because the uncertainty is difficult to model. An example vector time series \( q_k \) generated in this way is shown in Fig. 2.3.
2.4.3 Comparison of reserve costs

For each of the 50 wind realization tests, three parallel models of the power system were driven by the same realizations of the random wind model described above. The observed operating costs under receding horizon control over the simulation period were then compared under the three schemes, which were defined as follows:

1. Prescient case: The values of future disturbances are known at the time the finite-horizon optimization is carried out. Matrices $D_i$ are therefore not needed, and nominal schedules $e_i$ track the power reference perfectly. This scheme, which results in the best attainable receding-horizon cost, is used as a point of comparison for the other two.

2. Flexible-rate reserves: $[D_i]_{l,m} = 0$ for $l \neq m$, for all $i$. This represents the best possible response to uncertainty without time coupling, and the optimization is over the elements of $e_i$ and the diagonal parts of $D_i$.

3. Policy-based reserves: $[D_i]_{l,m} = 0$ for $l < m$, for all $i$. This allows full use of the extra information that will be available at each time step when the reserve is deployed.

The total operation cost was measured from the state and input values $x^i(t)$ and $u^i(t)$ realized by the elastic participants at each time step $t$ over the simulation period,

$$\sum_{i=1}^{N_p} \sum_{t=1}^{288} \left[ \tilde{f}_i x^i(t) + \frac{1}{2} x^i(t) \tilde{H}_i x^i(t) + \tilde{f}_i u^i(t-1) + \frac{1}{2} u^i(t-1) \tilde{H}_i u^i(t-1) + \tilde{c}_i \right].$$

The inputs are indexed by $(t-1)$ because state $x(0)$ is given whereas input $u(0)$ must be chosen and determines $x(1)$, and so on.

An example of the power output traces for the generators is given in Fig. 2.4. The cost results are shown in Table 2.2. A cost of reserves is defined for Schemes 2 and 3 as the operation cost experienced minus the prescient cost (from Scheme 1). This represents the cost incurred in order to accommodate the uncertain wind in-feed. Across the 50 runs, the cost of reserves decreased by an average of 38.4% for full policies (Scheme 3) with respect to the best possible non-recourse reserve scheduler (Scheme 2). Average costs under Scheme 1 were $4.474 \times 10^7$.\[45\]
Figure 2.4: Example of power output traces for one wind realization under Scheme 3. Generator outputs are plotted in stacked form, and the total power output of the two energy storage units is plotted using the thin black line relative to the top of the stack. The wind power injection is the difference between the total load (bold black line) and the sum of storage and generator in-feeds (thin black line).

Table 2.2: Cost comparison for different structural restrictions on $D_i$

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Comp. time</th>
<th>Average cost increase over Scheme 1</th>
<th>Avg. reserve cost vs. Scheme 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Prescient</td>
<td>34 ms</td>
<td>—</td>
<td>0 %</td>
</tr>
<tr>
<td>2. Diagonal</td>
<td>3843 ms</td>
<td>0.825 %</td>
<td>:= 100.0 %</td>
</tr>
<tr>
<td>3. Full policy</td>
<td>4424 ms</td>
<td>0.511 %</td>
<td>61.6 %</td>
</tr>
</tbody>
</table>

for the three-day test, and reserves were found to incur up to 1.10% of additional total power generation costs in the tests under Scheme 2, and up to 0.73% under Scheme 3. Computation times are also reported in Table 2.2 and represent the average time needed to build and solve each finite-horizon optimization problem on an Intel Xeon E5-2670 2.60 GHz CPU. Note that Scheme 1 is far faster than the others because only open-loop schedules $c_i$ need to be computed.

Another series of tests was carried out allowing only the diagonal and immediate sub-diagonal entries in the $D$-matrices to differ from zero. This results in a two-step policy, requiring fewer non-zero optimization variables than Scheme 3. The average savings in the cost of reserves, measured in the same way as for Scheme 3, were 32.4%, indicating that on average most of the 38.4% savings can be obtained from using just a two-step policy rather than a full policy (in this case 8 steps).
2.4.4 Sensitivity analysis

The question arises whether the cost savings reported for affine reserve policies depend on the quantity of wind energy present. To test this the case study was repeated under wind realizations driven by different-sized instances of the random process \( \mathcal{R} \), scaling \( q_{\max} \) and \( b_\beta \) by a factor \( \phi \in [0.2, 1.2] \), and \( \Sigma \) by \( \phi^2 \) since \( \Sigma \) represents the variance of a linearly-scaled quantity. The results, for 50 runs each, are shown in Table 2.3 and plotted in Fig. 2.5 (note that because in this scenario wind power is not curtailed, for \( \phi > 1.2 \) infeasibility arises from the fact that the wind power contribution is larger than the load and cannot be absorbed by storage). The expected proportion of load energy supplied by wind for each scaling factor \( \phi \) is \( \phi \cdot 29.0\% \).

The additional percentage cost of accommodating the uncertainty increased sharply (apparently more than quadratically) as the wind capacity was increased. Using time-coupled policies, around 40% of this could be offset as the wind share grew. Although the percentage saving decreased slightly for increasing \( \phi \), the absolute saving still grew quadratically. Note that the large saving reported for \( \phi = 0.2 \) is measured relative to a tiny number.

---

**Figure 2.5:** Added reserve cost percentages under different wind penetration ratios \( \phi \), averaged over 50 three-day simulation runs.
## 2. Affine Reserve Policies

### Table 2.3: Variation of results with installed wind power capacity

<table>
<thead>
<tr>
<th>Scaling $\phi$</th>
<th>Wind capacity</th>
<th>Average reserve costs Scheme 2</th>
<th>Average reserve costs Scheme 3</th>
<th>Average reduction under Scheme 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.6 GW</td>
<td>0.022 %</td>
<td>0.012 %</td>
<td>64.0 %</td>
</tr>
<tr>
<td>0.4</td>
<td>1.2 GW</td>
<td>0.099 %</td>
<td>0.057 %</td>
<td>43.2 %</td>
</tr>
<tr>
<td>0.6</td>
<td>1.8 GW</td>
<td>0.238 %</td>
<td>0.139 %</td>
<td>42.3 %</td>
</tr>
<tr>
<td>0.8</td>
<td>2.4 GW</td>
<td>0.468 %</td>
<td>0.283 %</td>
<td>39.9 %</td>
</tr>
<tr>
<td>1.0</td>
<td>3.0 GW</td>
<td>0.825 %</td>
<td>0.511 %</td>
<td>38.4 %</td>
</tr>
<tr>
<td>1.2</td>
<td>3.6 GW</td>
<td>1.378 %</td>
<td>0.919 %</td>
<td>33.7 %</td>
</tr>
</tbody>
</table>

## 2.5 Discussion

The potential cost savings from reserve policies are determined by the time coupling of participant costs, constraints, and prediction errors. This is evident from the term $\frac{1}{2} \langle D_i^t(B_i^t H_i^t B_i + H_i^t^*) D_i, \mathbb{E} [\delta \delta^*] \rangle$ in the cost function. Prediction error correlation in space and time results in non-zero off-diagonal entries in $\mathbb{E} [\delta \delta^*]$, which will change the contributions of individual entries of the policy matrices $D_i$ to the expected costs. It is not surprising, then, given the strong correlation of prediction errors found in the numerical case study, and the presence of time-coupled generation costs, that for a given finite-horizon optimization the costs were reduced when reserve policies were enabled.

In receding horizon operation, new power schedules are found at every time step. This can be viewed as a form of recourse that is not taken into account for each finite horizon optimization. This leads to the intuition that some of the apparent value of optimizing over full lower-triangular policies (rather than diagonal ones) may disappear once the optimization is repeated in receding horizon fashion. However the results show that cost savings indeed remain. It was also observed that much of the cost reduction can be gained without using all sub-diagonal entries in the matrices $D_i$.

The interpretation of the reserve cost defined in Section 2.4.3 is important. Scheme 1 represents a lower bound on the cost achievable under any possible receding horizon scheme, but it may in fact be impossible to find causal policies $\pi_i(\delta)$ offering the required robustness to prediction errors with costs approaching this value. Therefore the savings reported are in fact lower bounds on the true available savings. Recent theoretical performance bounds [HGK11, VPGM13] suggest that
the optimality of the reserve policies would depend on the current power system state, but that this relationship is far from straightforward.

A standard set of power system simplifications were employed in order to allow tractable optimization problems to be formulated. For a real implementation, additional ex-post checks (possibly with some iteration), would be needed to confirm feasibility of the flows on the AC transmission grid, in a fashion similar to current system operator practice [ODK11].
Chapter 3

Implementation of Affine Reserve Policies

This chapter discusses the practicalities of implementing affine reserve policies. Section 3.1 discusses the adaptation of existing markets to include time-coupled reserve products, and Section 3.2 discusses technical changes to power system operations. Most of the material is taken from [WMM13], with some parts taken from [WGMM13].

3.1 Market incorporation

This section addresses how reserve policy commitments could be made in a market context. The rules governing reserve markets are fairly fluid, and many market operators may make changes every year. In some markets many different reserve products exist and market participants have significant freedom to devise their own. In others, reserve trading takes place under a more prescriptive regime. See [ENT12] for an overview of these variations in Europe.

When bidding on reserve markets, a participant such as a generation company has to trade off the income from reserves against the opportunity cost of not being able to use the same generators to supply scheduled power, for example on the day-ahead market. This is because selling reserves implies a guarantee that the corresponding generation capacity margin will be available and not used for supplying scheduled power. In some countries, reserve sales are tied to a particular
3. Affine Reserve Policies: Implementation

Figure 3.1: The structure of the matrix containing reserve policy information. Each market participant commits to time sequences described by a matrix, and the figure shows how these matrices are related.

generator, whereas in others participants are free to use whichever generator they like to supply requested reserve power. Regardless of which of these is used, though, there is still an opportunity cost associated with the sale of reserves.

This cost also comes into the decision making process for a participant selling reserve policies. The company offering such a reserve product must be sure that the reserve sequence, which is a function of the uncertain power in-feeds not yet known, can be implemented for any realization of the error (in a given range, declared by the system operator). This necessarily entails more complexity than the status quo, however the optimization problem is relatively straightforward to solve and has the advantage of passing the prediction error statistics directly to the market participant.

As with any traded good, it can be expected that market participants would find it in their best interests to take a rigorous approach to solving the problem of exactly how much value they place on reserve policies, and furthermore would learn over time the best way of participating in any market mechanism introduced.
3.1 Market incorporation

3.1.1 Efficient reserve policy prices

It will now be shown that efficient market prices exist for reserve policies and that these are an exact analogue of standard Locational Marginal Prices (LMPs), which arise in electricity markets from network congestion [Sto02].

A partial Lagrangian of problem (2.10) (keeping the constraints \((D_i, e_i) \in F_i(x_0^i)\) and \(Z \geq 0\) but relaxing all others), \(\mathcal{L}(Z, D_1, e_1, \ldots, D_{N_p}, e_{N_p}, \lambda, \Pi, \nu, \Psi)\), can be formed by introducing the multipliers \(\lambda \in \mathbb{R}^T\) for violation of constraint (2.10b), \(\Pi \in \mathbb{R}^{T \times N_t T}\) for (2.10c), \(\nu \in \mathbb{R}^{2LT}\) for (2.10d), and \(\Psi \in \mathbb{R}^{2LT \times N_t}\) for (2.10e). The Lagrangian thus has the form

\[
L(Z, D_1, e_1, \ldots, D_{N_p}, e_{N_p}, \lambda, \Pi, \nu, \Psi) = \\
\sum_{i=1}^{N_p} \tilde{J}_i(x_0^i, D_i, e_i) + \lambda' \left( \sum_{i=1}^{N_p} r_i + C_i A_i x_0^i + C_i B_i e_i \right) \\
+ \left\langle \Pi, \sum_{i=1}^{N_p} G_i + C_i B_i D_i \right\rangle \\
+ \nu' \left( Z' h + \sum_{i=1}^{N_p} \Gamma_i C_i B_i e_i - \bar{p} + \sum_{i=1}^{N_p} \Gamma_i (r_i + C_i A_i x_0^i) \right) \\
+ \left\langle \Psi, \sum_{i=1}^{N_p} \Gamma_i (G_i + C_i B_i D_i) - Z' S \right\rangle
\]

After some rearrangement, this can be written in separable form as

\[
\mathcal{L}(Z, D_1, e_1, \ldots, D_{N_p}, e_{N_p}, \lambda, \Pi, \nu, \Psi) = \\
\sum_{i=1}^{N_p} (\tilde{J}_i(x_0^i, D_i, e_i) - \lambda e_i - \langle \Pi_i, D_i \rangle) \\
+ \nu' Z' h - \langle \Psi, Z' S \rangle + f(\lambda, \Pi, \nu, \Psi)
\]

where \(f\) is constant with respect to the primal variables \(D_i, e_i,\) and \(Z\). The Lagrange multipliers from optimization problems are commonly interpreted as prices, and here two prices have in effect been defined. The first is a nodal power price

\[
\lambda_i := -B'_i C'_i (\lambda + \Gamma'_i \nu),
\]

consisting of a global component depending on \(\lambda\) and a local component (induced by any line congestion present) depending on \(\nu\). This agrees with the standard
derivation of locational marginal prices (LMPs) in optimal power flow theory. The second is a matrix of reserve policy prices

$$\Pi_i := -B_i'C_i'(\Pi + \Gamma_i\Psi),$$

whose entries are the marginal value of each element of $D_i$. Note that the elements of $\Pi_i$ above the main diagonal are not required, since these are related to entries that would anyway result in non-causal responses to prediction errors, were they to be different from zero.

It can be shown that $C_iB_i = I$ in most cases (see the note below), so that usually $\lambda_i = -(\lambda + \Gamma_i\nu)$ and $\Pi_i = -(\Pi + \Gamma_i\Psi)$. The minus signs are a result of the sign convention used to write constraints (2.10b) and (2.10c). The identical form of $\lambda_i$ and $\Pi_i$ suggests that optimal prices for reserve policies exhibit the same locational variation as LMPs.

Because the Lagrangian is separable, the terms $\lambda_i'e_i$ and $\langle \Pi_i, D_i \rangle$ can be identified as the efficient payments to each participant $i$ for committing to nominal plan $e_i$ and reserve policy $D_i$. The payments represent the money transfers that would result from a market mechanism that solved problem (2.10) efficiently. Under such a scheme, the expected profit made by participant $i$ would be equal to

$$-\tilde{J}_i(x_0^i, D_i, e_i) + \lambda_i'e_i + \langle \Pi_i, D_i \rangle.$$

**Cases where $C_iB_i = I$ for elastic participants:**

Recall that from the definitions for $C_i$ and $B_i$ given in Section 2.2.1,

$$C_iB_i = \left[ I_T \otimes [1 0_{1 \times (n_i-1)}] \right] = \begin{bmatrix} \tilde{B}_i & 0 & \cdots & 0 \\ \tilde{A}_i\tilde{B}_i & \tilde{B}_i & \cdots & \vdots \\ \vdots & \cdots & \cdots & 0 \\ \tilde{A}_i^{T-1}\tilde{B}_i & \cdots & \tilde{A}_i\tilde{B}_i & \tilde{B}_i \end{bmatrix}.$$

As described in Section 2.2.1, it is assumed that the first state (the participant’s net power infeed) is set directly by the input decision at the previous time step.
3.1. Market incorporation

and not as a function of current states. Therefore the first row of $\tilde{A}_i$ contains only zeroes, and $[\tilde{B}_i]_1 = 1$. From this it can easily be shown that $C_iB_i = I_T$, from which the property follows immediately.

3.1.2 Policy-based reserve products

A natural question is what physical commitment a participant makes when agreeing to honour a given policy $(D_i,e_i)$. Clearly the nominal part $e_i$ is the schedule participant $i$ will follow if predictions turn out to have been made with perfect accuracy, i.e. $\delta = 0$. The interpretation of $D_i$ is more subtle and can be termed in two ways. The following analysis is given for a single uncertainty source, i.e. $N_\delta = 1$; the case for $N_\delta > 1$ is analogous.

Firstly, each row $l$ of $D_i$ can be read as the rule the participant must follow to construct its input $u_i^l$, which for the realized errors $\delta_0, \ldots, \delta_l$ determines the power it injects at time $l + 1$, so that $u_i^l = [e_i]^l + \sum_{m=0}^{l}[D_i]_{l,m}\delta_m$.

Secondly, each column $m$ of $D_i$ (reading down from the element on the diagonal) is, in control terminology, the planned impulse response $g(\delta_{m+1})$ of participant $i$ to a unit prediction error $m$ steps from the current time (recalling that $\delta_{m+1}$ is revealed just before $u_i^m$ is applied), so that $g(\delta_{m+1}) = [[D_i]_{m,m},[D_i]_{m+1,m}, \ldots, [D_i]_{T-1,m}]$.

Consider the example of the decision rule governing $u_i^l$, which sets the power participant $i$ supplies during step $l + 1$ of the planning horizon. For given choices of $[D_i]_{l,0}, [D_i]_{l,1}, \ldots, [D_i]_{l,l}$ agreed, the payment (using the price notation from Section 3.1.1) made to participant $i$ for the reserve service would be $\sum_{m=0}^{l}[\Pi_i]_{l,m}[D_i]_{l,m}$ and the payment for scheduled power $[e_i]^l$ would be $[\lambda_i]^l[e_i]^l$. An analogous product could be sold based on the column-wise reading of $D_i$, in which case the participant would be selling an a posteriori response to errors.

Existing reserve mechanisms could be modelled by matrices $D_i$ for which only the main diagonal is populated. In secondary reserves provided by conventional generators, feedback controllers use the frequency deviation, in the form of the Area Control Error (ACE) signal, to adjust their power outputs to follow load mismatches. For a unit deviation from the net load reference, each generator will end up with an offset that depends on its controller parameters. For such a unit load deviation at time $l$, this offset is in the notation used here exactly the matrix entry $[D_i]_{l,l}$. 55
Therefore, a simple way of comparing policy-based reserves to existing mechanisms is to restrict the structure of the matrices $D_i$ accordingly and inspect the results. This was demonstrated in Section 2.4.

### 3.1.3 Clearing mechanisms

Three potential methods for determining how reserve policies are to be split between market participants are now discussed. For reasons explained below, the third of these, namely an optimization carried out by the system operator, appears to be the most plausible.

**Double-sided auctions**

Double-sided auctions are currently used to clear day-ahead and intra-day electricity markets; the theory also extends to the provision of reserve policies. A possible method would be to run an auction for every element of the policy matrix structure shown in Fig. 3.1 apart from the elements above the diagonal, which are zero for everyone. The system operator would buy commitments summing to 1 for the diagonal entries, meaning that the reserve actions of the market participants would sum to counteract the load mismatch as it occurs; and operate auctions clearing at a net demand of 0 for subdiagonal entries, meaning that the participants’ commitments to change operating point would sum to zero. Participants could bid as buyers or sellers into these auctions, at positive or negative prices. These auctions are illustrated schematically in Fig. 3.2.

There is a key difference between current reserve acquisition methods and that proposed in the paragraph above. Currently the system operator buys a specific capacity in MW to cover possible errors in the prediction of net load. In the proposal described in this work, the system operator publishes its bounds and statistics for the prediction error, and the participants bid to counteract a certain fraction of the error, using the parameters the system operator has provided. This modification may be beneficial in that it allows market participants to optimize their actions with respect to the real uncertainty statistics affecting the network, rather than a coarse specification of reserves that may cover scenarios that cannot (or are extremely unlikely to) arise in the hours ahead.
3.1. Market incorporation

Each participant enters accepted bids into its own policy matrix. Blue matrix elements indicate commitments to be settled by auction for each participant.

**Figure 3.2:** Auctions for the two different types of reserve policy commitment. Diagonal entries correspond to a commitment to counteract a fraction of the net load mismatch, and therefore the auction should clear with a total of 1. Subdiagonal entries correspond to commitments to adjust after a mismatch has been counteracted, and therefore the auctions clear with net quantity 0.

**Bilateral trade**

The sub-diagonal entries of matrices diagrammed in Fig. 3.1 should sum over all market participants to zero. In other words, any of these time sequence components that are sold by one participant have to be bought by another, so that when the time sequence is triggered, the responses of the two participants cancel each other out. This may appear to render reserve policies unnecessary, however it is important to note that the operating points of the two participants will be different as a result, which means that costs and constraints for actions such as ramping generators up and down, or emptying and filling energy storage, will evolve differently over time.

This symmetry suggests that one way of structuring a market for reserve policies would be to allow bilateral trades to occur between market participants, for any given step of a sequence.
Online ISO-side optimization

Across different markets the system operator takes differing degrees of authority over the participants’ plants for the purpose of actuating reserves. Online optimization is playing an increasing role in markets such as PJM [Ott10] as a way of reducing costs that cannot be taken care of easily using price signals. A notable example of this is generator switching, the cost of which was traditionally socialized amongst electricity consumers in the PJM interconnection. The mechanism was altered so that rather than letting market participants choose when to switch on and off, the ISO now solves a mixed-integer optimization problem at regular intervals in order to determine the most efficient switching decisions. This led to significant savings for end users, at the cost of requiring more interference from the ISO in daily plant operations.

The fact that centralized system operator optimizations are already used to coordinate reserves in some liberalized markets suggests that it may be possible or attractive to coordinate reserve policies centrally, rather than running auctions or allowing bilateral trades. Reserve markets often suffer from liquidity and market power problems, and this third alternative may help to mitigate these.

The ISO would require a good model of the constraints and costs of the market participants to run a centralized optimization and compute efficient reserve policies. This might require some intervention to force participants to reveal more cost or constraint data in the case that the information communicated via bids (or already held by the ISO for other technical purposes) is insufficient.

3.1.4 Real time operation

For continuous operation of the power system, it is necessary to choose new policies periodically, because the policies only apply for the following $T$ steps at the time when they are chosen. Furthermore the optimal policies are a function of the current state of the system, and it may be attractive to choose new policies early in the light of new forecast information. Therefore a systematic way of choosing policies repeatedly is required. Two possible schemes for such “closed-loop” operation are:

- **Batchwise**: Set policies for a horizon of $T$ steps, let all $T$ steps play out, then choose a new batch of policies once the current ones have expired.
• **Receding horizon**: Let only the first step play out, then immediately update the policies for the next \( T \) steps.

These are shown schematically in Fig. 3.3 A middle road between these two options also exists, namely letting some of the \( T \) steps play out, then discarding the remainder and choosing new policies.

The receding horizon scheme presents another choice – whether to honour previous policies in the new choices for \( D_i \) (respectively \( e_i \)). This would be done by shifting the matrices (resp. vectors) upward and leftward (resp. upward) by one element, then optimizing over the bottom row of elements (resp. element) to obtain new policies. These new policies would be feasible with respect to \( F_i(x_i^0) \) as well as the global constraints [2.10b]-[2.10e] as long as \( \Delta \) is defined such that the bounds on the uncertainty grow monotonically along the prediction horizon (for conciseness this is stated without proof here).

An alternative to this is to reject the previous policies and choose fresh values for \( D_i \) and \( e_i \) at every step. This option is attractive because it allows new (presumably better) policies to be chosen in the light of new information as soon as it becomes available. However its drawback is that due to the repeated renewal process, only the \((0,0)\) block of the matrix \( D_i \) ever gets used, and no policies beyond the first row ever appear to be implemented. This conflicts with the ideas developed in Section 3.1.2, in that contracts for policy-based reserves would be agreed but then never called on. It is important to note, however, that the optimal values of \([e_i]_0\) and \([D_i]_{0,0}\) would be affected strongly by the costs and constraints modelled for steps 1 to \( T - 1 \), even though those later steps would never be realized. This means that under this scheme, any cost savings from employing reserve policies are ultimately to be found in the more intelligent choice of \([e_i]_0\) and \([D_i]_{0,0}\). This effect was reported in Section 2.4.

Rejecting previous policies at each time step makes choosing correct payments to market participants difficult. One simple way of overcoming this would be instead to define the new policies as deviations from shifted versions of the existing previous policies, with payments settled additively. Under such a scheme there would be \( T \) payments for the reserve action undertaken at a given time, resulting from the superposition of policy choices made over the last \( T \) steps.

In the presence of a large renewable in-feed, bounds on the prediction errors may change significantly over the course of a prediction horizon; by the time the prediction horizon has nearly been played out, prediction error bounds for the last few
steps are likely to be much smaller than those assumed at the start of the horizon, when the policies were chosen. Therefore it seems logical to prefer a rolling system of adjustments to policies, of the kind described above, rather than a batch-wise approach.

3.1.5 Decision-making process

Fig. 3.4 shows a generic structure that could be used by system operators to make real-time decisions in the face of rapidly-changing system state and forecast information. The output of a numerical weather prediction model based on a finite-element analysis is high-dimensional, and a description of the uncertainty of the model (obtained, say, by running a series of Monte Carlo simulations with perturbed initial conditions) will also be high-dimensional. In order to formulate a tractable optimization problem, the uncertainty must be described by a vector \( \delta \) of reasonably small size \( N_\delta \). The approach described in this chapter therefore requires the true uncertainty description to be reduced using an appropriate component analysis method. However the exact method is outside the scope of this work; it is a basic assumption that an appropriate uncertainty model, obtained after applying a method such as principal component analysis, is supplied as an input parameter to the optimization solver.

3.2 Technical incorporation

This section explains how electrical reserves may be made more flexible in order to unlock any potential cost savings that time-coupled reserve policies may bring. First the existing cascaded control structure used to regulate grid frequency is described, then an adaptation is proposed in order to incorporate reserve policies while maintaining grid stability. Subsection 3.2.2 describes how secondary reserves can be modified to implement the first step of a reserve policy, and 3.2.3 explains how the remainder of the reserve policy steps can be handled by tertiary reserve commands.
3.2. Technical incorporation

**Figure 3.3:** Batchwise and receding horizon implementations of reserve policy decisions. Policy decisions can either be allowed to expire (batchwise case) or overlaid/overwritten (receding horizon case) after making new choices every period.

**Figure 3.4:** General decision-making optimization scheme for system operators. The blue boxes denote input parameters to the optimization process that evolve in real time.
3. Affine Reserve Policies: Implementation

3.2.1 Existing control structure

The three control layers for power system reserves, illustrated in Fig. 3.5, are as follows (with the newer ENTSO-E terminology [ENT12] given in brackets):

Primary control (frequency containment)

This is the fastest acting category of reserves, acting in less than 30 seconds. Frequency containment ensures that in the event of a contingency or load mismatch the global grid frequency does not continue to diverge from nominal. The primary reserve action is, in control terminology, a proportional feedback controller, where the size of the control input (the change in power $\Delta P_{\text{prim}}(t)$ injected into the grid) at any given time is proportional to the deviation of the current locally-measured grid frequency $f(t)$ from the nominal frequency $f_0$:

$$\Delta P_{\text{prim}}(t) = k_{\text{prim}} [f(t) - f_0]$$

where $k_{\text{prim}}$ is a constant of proportionality, or gain. Note that this feedback is applied negatively as shown in Fig. 3.5 since extra power should be injected when the frequency drops below nominal. The proportional controller is configured to provide a very fast response in the event of a contingency such as a generator failure. However, it will not in general be able to correct the frequency deviation back to zero, meaning there will be a steady-state error in frequency.
Secondary control (frequency restoration)

Secondary reserves act on timescales longer than 30 seconds, and continually correct the grid frequency back to nominal. These reserves act not on a local measurement but on the Area Control Error (ACE) signal $ACE(t)$, which is provided by the system operator and contains information on the power supply mismatch in the local area, as deduced by tie line measurements. The secondary control signal is also called the Automatic Generator Control (AGC) signal $AGC(t)$, and takes the form of a generator setpoint. It is calculated by adding proportional and integral responses to the ACE signal, of the form

$$AGC(t) = k^p_{sec} ACE(t) + k^i_{sec} \int_{-\infty}^{t} ACE(\tau) \, d\tau,$$

where $k^p_{sec}$ and $k^i_{sec}$ are gains that govern the magnitudes of the proportional and integral feedback components respectively. In control terminology, this is a proportional-integral controller.

Tertiary reserves (replacement reserves)

These are activated in order to bring the margins for frequency restoration reserve back to a desired level. The activation of secondary reserves results in changes to generator operating points that may result in them hitting operating limits. In mathematical terms, the action of this control layer cannot be stated as explicitly as for the primary and secondary reserves described above. Tertiary reserve products and their means of activation are diverse, and rules for their provision differ between markets to a much greater extent than for primary and secondary reserves [ENT12].

3.2.2 Adapting AGC to implement the first step of a reserve sequence

Dispatch planning in power systems takes place over discrete time intervals of between 5 minutes and 1 hour [ENT12]; 15 minutes is a typical duration. However, wind in-feeds fluctuate in continuous time without regard to such planning intervals. Therefore when planning a rigorous response to errors in the prediction of these in-feeds it is necessary to reconcile the discrete-time plan with continuous-time reality.
Reserve policies commit participants to counteract fixed fractions of the prediction error during a trading interval. The main technical challenge is therefore to make sure that at the end of each trading interval, the changes in power output of the participants are consistent with the reserve policies and the realized value of the prediction error for that interval. Crucially, this must be achieved in the presence of automatic generator controllers already installed in the network.

Note that although the model in Section 2.2 allows for the possibility of multiple sources of uncertainty, for example from geographically distinct wind farms, for simplicity of explanation it is assumed here that a single source of uncertainty is used to model a change in the system-wide net load. The extension to multiple sources of uncertainty is valid as long as the values of these uncertainties can be estimated in real time.

Let the total amount of unpredicted energy fed into the power system over a trading interval $k$ be denoted $\Xi_k$ (measured in MWh). This quantity is accumulated over the course of the interval, and the total is only known at the end of it. Let the fractional responsibility for countering this additional energy be denoted $\phi_{ik}$ for each participant $i$, as governed by the $k$th column of the policy matrix. For example, $\phi_{ik} = 0.2$ commits participant $i$ to reduce its energy production by 20 MWh to accommodate part of a 100 MWh increase in wind energy during interval $k$. The coefficients $\phi_{ik}$ have been chosen in advance such that the extra in-feeds cancel the unpredicted energy $\Xi_k$:

$$\sum_{i=1}^{N_p} (\phi_{ik} \Xi_k) = \Xi_k$$

(3.3)

where $N_p$ is the number of participants. Furthermore, it is required that the final offset of a given participant $i$’s operating point is equal to the average of the additional power required from it over the interval. Defining the instantaneous power offset for participant $i$ at time $t$ after the start of interval $k$ as $\theta_{ik}(t)$, this requirement can be written as

$$\theta_{ik}(T) = \frac{\phi_{ik} \Xi_k}{T},$$

(3.4)

where $T$ is the length of the interval. The most natural way of meeting requirements (3.3) and (3.4) is to design the control action so that this share is tracked as it evolves over the whole interval; otherwise there may be a need to make a sudden compensation at the end to account for an accumulated mismatch.
Let the instantaneous unpredicted power in-feed at time $t$ after the start of interval $k$ be denoted $\xi_k(t)$ (measured in MW) so that
\[ \int_0^T \xi_k(t) \, dt = \Xi_k, \]
and similarly
\[ \int_0^T \theta_{ik}(t) \, dt = \phi_{ik} \Xi_k. \]
Then at a given time, it can be specified that
\[ \theta_{ik}(t) = -\phi_{ik} \xi_k(t). \tag{3.5} \]
Since $\sum_{i=1}^{N_p} \phi_{ik} = 1$ it is possible to ensure that $\sum_{i=1}^{N_p} \theta_{ik}(t) = \xi_k(t)$, as long as the error $\xi_k(t)$ can be estimated in real time. The control rule (3.5) could be followed by setting parameters on the automatic controllers, for example by altering the AGC set-point that each participant receives, so that the integral component ramps up according to the fraction of the error that the participant is to compensate. Note that manipulations of AGC have been proposed elsewhere \cite{VMLA13} as a way of improving the responsiveness of power systems. Non-generator market participants, such as storage units or aggregated demand response, could participate in the same mechanism as long as they are capable of manipulating their power output in response to an AGC-like signal transmitted by the system operator.

One question that arises is whether the presence of primary control action in some generators might interfere with the AGC adaptation described here, since according to the block diagram on Fig. 3.5 primary control adds its own offset to the set-point of generators. On average, primary control action tends to be neither positive nor negative, and the quantities of energy provided through primary reserves also tend to be small. This implies that the primary control action should rarely interfere with the control strategy described above.

### 3.2.3 Adapting tertiary reserves to implement the remaining steps of a reserve sequence

The proposal above addressed how to respond to a prediction error during the interval in which it is measured, thereby implementing the first step of the planned reserve sequence. The remaining steps of the reserve sequence must then be
implemented, which can be achieved straightforwardly by adding an offset to the power set-point of each participant. This would enter as a change to the “Nominal power set-point” at the left-hand side of Fig. 3.5.

The optimization routine that chooses the reserve policies also takes care of freeing up generation capacity and therefore constitutes a dynamic re-dispatch rule. Since this is the key function of tertiary reserves (along with switching decisions for non-spinning reserves), it could be said that part of the traditional tertiary reserve function is covered as part of the reserve policy concept.
Chapter 4

Extensions

The results presented above were based on simplifying assumptions concerning the system model primarily that no integer decisions were to be coupled with the choice of policy. In addition, no solution was offered to the problem that the recourse variables $D_i$ cause problem (2.10) to become “large” even for rather modest numbers of participating devices.

This section details some extensions that overcome these limitations. Section 4.1 describes a combined unit commitment formulation that turns out to be particularly attractive. It includes material from [WHGM14]. Section 4.2 reports on computational tests using distributed optimization to scale the problem up to large size. Section 4.3 sketches a double-sided extension that accommodates the widely-recognized distinction in the costs of providing up-spinning and down-spinning reserves in real power systems.

4.1 An MLD recourse formulation of robust unit commitment

One possible attraction of affine reserve policies is that they may beneficially influence the unit commitment problem if the two decisions are made in a coupled manner. This is because if uncertainty can be accommodated less conservatively, it may be possible to avoid switching expensive generators on. To see this effect, it is necessary to augment problem (2.10) by introducing integer decision variables
that govern the generator switching actions.

4.1.1 Problem reformulation

The elastic participants described in Section 2.2.1 are now separated into switchable and non-switchable units. Each non-switchable participant $i$ is modelled using linear time-invariant dynamics, where the continuous state of participant $i$ at time $k$ is denoted $x^{c,i}_k$ with

$$x^{c,i}_{k+1} = \tilde{A}^{c,i}x^{c,i}_k + \tilde{B}^{c,i}u^{c,i}_k. \quad (4.1)$$

For each switchable participant $i$, the following dynamics incorporating a discrete on/off state $x^{d,i}_k$ are defined:

$$x^{c,i}_{k+1} = \begin{cases} \tilde{A}^{c,i}x^{c,i}_k + \tilde{B}^{c,i}x^{d,i}_{k+1} & \text{if } [x^{d,i}_{k+1}]_1 = 1 \\ \tilde{A}^{c,i}x^{c,i}_k & \text{if } [x^{d,i}_{k+1}]_1 = 0 \end{cases} \quad (4.2a)$$

$$x^{d,i}_{k+1} = \tilde{A}^{d,i}x^{d,i}_k + \tilde{B}^{d,i}u^{d,i}_k. \quad (4.2b)$$

For both switchable and non-switchable participants, the first element of the continuous state $[x^{c,i}_k]_1$ represents the power output at time $k$ while the remaining elements are used to model internal dynamics or to include prior states.

For switchable participants, a discrete state vector $x^{d,i}_k$ is employed, whose first element $[x^{d,i}_k]_1$ represents the on/off-state at time $k$ while other elements of $x^{d,i}_k$ are used to include prior states (this may be necessary, for example, to constrain or penalize generator switching behaviour over time). The scalar, binary input $u^{d,i}_k$ determines the on/off-status at step $k + 1$, with a value of 1 indicating an on state. The participant can inject or extract power at the next time step $k + 1$ only if it will be on, i.e., $[x^{d,i}_{k+1}]_1 = 1$. Note that $x^{d,i}_k$ is only affected by binary inputs, thus can be modelled as a continuous variable.

The logical constraint (4.2a) can be written more compactly as a bilinear equality:

$$x^{c,i}_{k+1} = \tilde{A}^{c,i}x^{c,i}_k + \tilde{B}^{c,i}x^{d,i}_{k+1} u^{c,i}_k, \quad (4.3)$$

$$x^{d,i}_{k+1} = \tilde{A}^{d,i}x^{d,i}_k + \tilde{B}^{d,i}u^{d,i}_k. \quad (4.3)$$

However, for optimization purposes this bilinearity is inconvenient. This is resolved by using the standard MLD system representation [BM99], which converts the model into one involving integers and easier-to-handle linear constraints. This requires an additional auxiliary variable $z^i_k := x^{d,i}_k u^{d,i}_k$, to which appropriate extra
4.1. An MLD recourse formulation

constraints are applied using the so-called “big-M” reformulation (see [BM99]). The vector $z^i_k$ is zero for times when the participant will be off at step $k + 1$, and equal to $u^c_k$ when it will be on.

Substituting the auxiliary variables into the dynamics and concatenating the continuous and discrete state, the system dynamics can be described in the following way:

$$
\begin{bmatrix}
x^c_{k+1} \\
x^d_{k+1}
\end{bmatrix} =
\begin{bmatrix}
\tilde{A}^c_i & 0 \\
0 & \tilde{A}^d_i
\end{bmatrix}
\begin{bmatrix}
x^c_k \\
x^d_k
\end{bmatrix} +
\begin{bmatrix}
0 \\
\tilde{B}^c_i
\end{bmatrix}
\begin{bmatrix}
u^c_k \\
z^i_k
\end{bmatrix},
$$

subject to the aforementioned additional linear constraints enforcing consistent behaviour of the auxiliary variable $z^i_k$:

$$
\tilde{E}^c_z z^i_k \leq \tilde{E}^c_i u^c_k + \tilde{E}^d_i u^d_k + \tilde{E}^0_i.
$$

For non-switchable (“always-on”) loads the definition $z^i_k := u^c_k$ is made for consistency.

As before, the stacked finite-horizon representation is defined for the purposes of optimization. Defining the state vector $x^i_k$ as $[x^c_{k1}', x^d_{k1}']'$ for switchable loads and as $x^c_{k1}'$ for non-switchable loads, the state of a switchable participant evolves according to

$$
x^i_k = A^i x^i_0 + B^d_i u^d_k + B^c_i z^i_k,
$$

and the state of a non-switchable participant according to

$$
x^i_k = A^i x^i_0 + B^c_i z^i_k,
$$

where the following stacked vectors have been defined: $x^i := [x^c_1', \ldots, x^c_T']'$, $u^d := [u^d_{01}', \ldots, u^d_{T-1}']'$, and $z^i := [z^c_1', \ldots, z^c_{T-1}']'$. Matrices $A^i$ and $B^d_i$ are defined as

$$
A^i =
\begin{bmatrix}
\tilde{A}^c_i & \tilde{A}^d_i & \cdots & 0 \\
\tilde{A}^c_i & \tilde{A}^d_i & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
\tilde{A}^c_i & \tilde{A}^d_i & \cdots & \tilde{A}^c_i
\end{bmatrix},
B^d_i =
\begin{bmatrix}
\tilde{B}^d_i & 0 & \cdots & 0 \\
\tilde{A}^c_i \tilde{B}^d_i & \tilde{B}^d_i & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
\tilde{A}^c_i \tilde{B}^d_i & \tilde{B}^d_i & \cdots & \tilde{A}^c_i \tilde{B}^d_i
\end{bmatrix},
$$

and $B^c_i$ is defined analogously to $B^d_i$. 69
The power flows seen by the network are given by $C_i x_i$, where $C_i$ selects the first continuous state at each time step, $C_i = I_T \otimes [1, 0, \ldots, 0]$.

The costs of a given state and input trajectory for participant $i$ are modelled using a function $J_i : \mathbb{R}^{T_{ni}} \times \mathbb{R}^T \times \mathbb{R}^T \times \mathbb{R}^T \to \mathbb{R}$ of the form

$$J_i(x_i, u^c_i, z_i, u^d_i) = f_i^{xx'} x_i + \frac{1}{2} x_i' H^x_i x_i + f_i^{uc} u^c_i + \frac{1}{2} u^c_i' H^{uc} u^c_i + f_i^{ud} u^d_i + c_i,$$

(4.8)

where it is noted that $x_i$ depends on $z_i$. The Hessians and linear coefficients are defined analogously before (see [WHGM14] for details).

The description of the local constraint set $Z_i$ must be augmented to account for discrete variables:

$$Z_i := \left\{ \begin{bmatrix} x^i \\ u^c_i \\ u^d_i \\ z_i \\ \delta \end{bmatrix} \left| \begin{array}{l} T_i x^i + U^c_i u^c_i + U^d_i u^d_i + V_i \delta \leq w_i \\ E^z_i z_i \leq E^c_i u^c_i + E^d_i u^d_i + E^0_i \\ u^d_i \in \{0, 1\}^T \end{array} \right. \right\}.$$

(4.9)

The binary inputs $u^d_i$ are to be chosen open-loop, and the continuous input $u_i$ is once more restricted to the affine function

$$u^c_i = e_i + D_i \delta,$$

(4.10)

Correspondingly, for $z_i$ the terms

$$z_i = z^c_i + Z^D_i \delta$$

(4.11)

are introduced. Note that while it would be attractive to determine a policy for the binary inputs, i.e. a causal function mapping the uncertainty $\delta$ to binary decisions $u^i$, this problem is far harder to solve, especially once computational cost is considered [BG13].

Such a policy must satisfy the constraints (4.5) and (4.9) for all values of the prediction error $\delta$. Let the set of feasible policies for participant $i$ given current
state \( x_0^i \) be denoted

\[
\begin{align*}
F_i(x_0^i) := \begin{cases}
D_i, & \text{if } E_i^{c,e} x_i^e \leq E_i^{d,e} u_i^d + E_i^{0,e} \\
E_i^{c,D}[Z_i^D]_n \leq E_i^{d,D}[D_i]_m + E_i^{0,D} & \text{if } A_i x_0^i + B_i^e u_i^d + B_i^z (z_i^e + Z_i^D \delta) + e_i + D_i \delta, \\
& u_i^d, z_i, \delta, \in Z_i, \forall \delta \in \Delta
\end{cases}
\end{align*}
\] (4.12)

The matrices \( E_i^{*,e} \) and \( E_i^{*,D} \) are built from individual stage-wise constraints of the form \( (4.5) \).

The cost function for participant \( i \) is redefined to take its dependence on the optimization variables \( u_i^d, e_i, D_i, z_i^e, Z_i^D \) and the current state \( x_0^i \) into account:

\[
\tilde{J}_i(x_0^i, D_i, e_i, z_i^e, Z_i^D, u_i^d) := E[J_i(A_i x_0^i + B_i^e u_i^d + B_i^z (z_i^e + Z_i^D \delta), e_i + D_i \delta, u_i^d)]
\]

This can be expanded to give a function that is quadratic in the optimization variables and linear in the error statistics \( E[\delta] \) and \( E[\delta \delta'] \). In contrast to that derived in Section 2.3, this function now depends on the discrete input \( u_i^d \) and the auxiliary variables \( (z_i^e, Z_i^D) \).

Using the same arguments that were used to derive problem \( (2.10) \), the finite-horizon problem accounting for unit commitment decisions can be written in finite form as

\[
\begin{align*}
\min_{W \geq 0, (D_i, e_i, x_i^0, z_i^e, Z_i^D, u_i^d) \in \mathcal{F}(x_0^i)} & \sum_{i=1}^{N_p} \tilde{J}_i(x_0^i, D_i, e_i, z_i^e, Z_i^D, u_i^d) \\
\text{s.t.} & \sum_{i=1}^{N_p} [r_i + C_i(A_i x_0^i + B_i^z z_i^e)] = 0, \\
& \sum_{i=1}^{N_p} (G_i + C_i B_i^z Z_i^D) = 0, \\
& \sum_{i=1}^{N_p} \Gamma_i [r_i + C_i(A_i x_0^i + B_i^z z_i^e)] - \tilde{p} \leq W^T h, \\
& \sum_{i=1}^{N_p} \Gamma_i (G_i + C_i B_i^z Z_i^D) = W^T S.
\end{align*}
\] (4.13a-e)
The form is very similar to that of problem (2.10), albeit with some changes of variable and a new definition for the sets $\mathcal{F}_i(x_0^i)$. The variable $Z$ from (2.10) has been renamed $W$ here due to the introduction of the auxiliary variables $z^e_i$ and $Z^p_i$. The problem is now a conventional mixed-integer quadratic program (MIQP). Only the coupling constraints have been written out explicitly; the sets $\mathcal{F}_i(x_0^i)$ and constraint $W \geq 0$ are left implicit because they do not couple across participants.

4.1.2 Numerical example

The use of the method described above will now be demonstrated for a case study in which energy storage and generation units must satisfy an uncertain net load over a finite planning horizon. It is compared to a method that does not use recourse, meaning that the reserve action in response to the discovery of prediction errors is not allowed to be time-coupled across the planning horizon. It is shown that time-coupled affine reserve policies introduce extra degrees of freedom that reduce the need to switch generators, and therefore reduce costs.

The example considers how slow-ramping cheap generation, expensive but fast-acting “peaking” generation, and energy storage devices should collectively satisfy an uncertain net load. All three categories of device take part in the reserve mechanism that counteracts the uncertainty. Although the net load is poorly predicted, it is known that it will increase suddenly within the planning horizon, within the bounds shown in Fig. 4.1. Such a situation often arises, for example, when a passing weather front causes a sharp drop in wind power availability, meaning that conventional generators must make up the resulting power shortfall.

The slow-ramping cheap generation is modelled as a single lumped unit that stays switched on throughout, but which cannot ramp up its power output fast enough
4.1. An MLD recourse formulation

Figure 4.2: Test system use for coupled unit commitment and policy optimization.

to compensate for the sudden drop in wind power. The difference is tracked by a lumped energy storage capability that represents, for example, the flexibility provided by demand response in the system, or batteries, or hydro-storage. However, since the energy storage has only limited capacity, it may not be able to track the difference alone. When this is the case, the expensive peaking generation is activated as a last resort (or rather, the unit commitment routine switches a lumped unit representing expensive generation on). The parameters from which the participant models were built are given in Table 4.1. Note that the cost of peaking generation is such that it will only be activated when no other solution exists; the fuel costs are chosen to reflect typical daily electricity price fluctuations. The setup is shown schematically in Fig. 4.2.

The number of steps in which the peaking generation needs to be activated during the planning period are measured for various ramping limits on the cheap generation, and for various energy storage capacities. Two cases are studied, as shown in Fig. 4.3:

1. Matrices $D_i$ restricted to a diagonal structure, which represent the best achievable performance where the reserve response is proportional to the instantaneous load mismatch, but cannot be time-coupled.

2. Matrices $D_i$ that can have any lower-triangular structure (recalling that elements above the diagonal would produce non-causal behaviour). These are “full” affine reserve policies, whose response can be a function of any load mismatch measured up to that point.

A time horizon of $T = 10$ hours was used, and transmission line constraints were assumed not to be reached (i.e., a “copper plate” network was assumed). The MIQPs were solved using CPLEX [1107], and the mean computation time was 223 ms with s.d. 144 ms in case 1, and 755 ms with s.d. 725 ms in case 2. Both cases
Table 4.1: Participant parameters

<table>
<thead>
<tr>
<th>Participant</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseload gen.</td>
<td>Linear fuel cost: 40.0 $/MW each hour</td>
</tr>
<tr>
<td></td>
<td>Quadratic fuel cost: 0.05 $/MW(^2) each hour</td>
</tr>
<tr>
<td></td>
<td>Power output (always on): 0 to 100.0 MW</td>
</tr>
<tr>
<td></td>
<td>Maximum ramp rate: variable (R) MW/h</td>
</tr>
<tr>
<td></td>
<td>Initial condition: Output 9.0 MW</td>
</tr>
<tr>
<td>Peaking gen.</td>
<td>Linear fuel cost: 70.0 $/MW each hour</td>
</tr>
<tr>
<td></td>
<td>Quadratic fuel cost: 0.10 $/MW(^2) each hour</td>
</tr>
<tr>
<td></td>
<td>Switching cost: $5,000</td>
</tr>
<tr>
<td></td>
<td>Additional running cost: $1,000 each hour</td>
</tr>
<tr>
<td></td>
<td>Power output range (on): 0 to 100.0 MW</td>
</tr>
<tr>
<td></td>
<td>Initial condition: Switched off, output 0 MW</td>
</tr>
<tr>
<td>Energy storage</td>
<td>Energy capacity: variable (E) MWh</td>
</tr>
<tr>
<td></td>
<td>Maximum power output/input: ±60.0 MW</td>
</tr>
<tr>
<td></td>
<td>Efficiency: 1 (pumping or generating)</td>
</tr>
<tr>
<td></td>
<td>Initial condition: Energy level 0 MWh</td>
</tr>
<tr>
<td>Net load</td>
<td>Nominal forecast for (T = 10) steps:</td>
</tr>
<tr>
<td></td>
<td>12.0, 12.0, 12.1, 15.8, 36.0, 56.2, 59.9, 60.0, 60.0, 60.0 MW</td>
</tr>
<tr>
<td></td>
<td>Uncertainty: (\delta_k \in [-k, +k]) MW at step (k)</td>
</tr>
</tbody>
</table>

contained 30 binary and 4260 continuous variables, with structural constraints trivially restricting some of these variables to particular values.

Fig. 4.4 shows the resulting number of times steps during the planning horizon in which the peaking plant had to be used, for both cases. Case 2 shows a reduced reliance on the peaking plant for a rather large range of energy storage sizes and ramp rate limits. This comes from the choice of more intelligent rules governing how the storage reacts to the uncertainty as its values are revealed over the planning horizon, and the beneficial effect is particularly prominent for intermediate storage sizes and ramp rates. For very large storage units and ramp rates (the top-right region of Fig. 4.4 (c)), the peaking plants are not required to ramp up the generation total in either case, and for very small storage or slow ramping limits (the bottom-left region) their use is in both cases unavoidable. Fig. 4.5 shows the corresponding solution costs.
4.1. An MLD recourse formulation

The operational changes brought about by time-coupled reserves are illustrated in Fig. 4.8, which shows the optimal reserve policies for cases 1 and 2 for an energy storage size of 40.0 MWh and a ramp rate limit of 8.0 MW/h.

In case 1, the energy storage unit must be operated relatively conservatively (the nominal energy level must be kept relatively near the middle of its capacity), because its reserve rule, shown in Fig. 4.8, can only act on the prediction error revealed at that time step. Because of the integrating effect of energy storage dynamics, the uncertainty range of the storage energy level can only broaden over the planning period. In contrast, case 2 allows the storage unit to take time-coupled actions so that it makes better use of the available storage capacity under uncertainty. This ultimately helps to reduce the use of the expensive generation.

Figs. 4.6 and 4.7 show the nominal plans and range of possible power outputs implied by the reserve policies chosen, for cases 1 and 2 respectively. Note that in case 2 the energy storage may provide a larger range of output thanks to the extra degree of operating freedom the time-coupled decision rule has introduced. It should be emphasized that in case 2, it is not possible for each device to provide an output that follows the upper (or lower) boundary of the shaded range; rather, the upper (lower) boundary at each time step is the highest (lowest) value that any sequence $\delta \in \Delta$ can cause at that particular time under the time-coupled decision rule.

4.1.3 Discussion

It was shown that coupling the choice of time-coupled reserve policies into a unit commitment optimization could reduce the activation of expensive generators. The extra degree of freedom that planned recourse provides appears to be particularly
interesting for reserve provision by energy storage units. As the case study showed, it was possible to control the uncertainty range of future energy storage levels much more tightly than would otherwise have been possible.

The formulation outlined above does not introduce any more integer variables than existing unit commitment problems, and therefore the scaling issues encountered...
4.2. Large-scale solutions using ADMM

Figure 4.5: Expected finite horizon cost (fuel plus switching costs). As with Fig. 4.4, the first plot shows case 1, the second shows case 2, and the third shows the difference between the two. The cost reduction was in the range $[0, 0.960 \times 10^4]$. 

by the new approach and others are similar. A useful and obvious future step would be to apply some of the existing scaling methods to solve real-sized problems less conservatively.
Figure 4.6: Case 1: Nominal plans and bounded deviations that could potentially arise under uncertainty, for the parameters $E = 40$ MWh, $R = 8$ MW/h.
4.2. Large-scale solutions using ADMM

Figure 4.7: Case 2: Nominal plans and bounded deviations that could potentially arise under uncertainty, for the parameters $E = 40 \text{ MWh}$, $R = 8 \text{ MW/h}$. 

79
Figure 4.8: Feedback policy matrices in cases 1 and 2, for $E = 40$ MWh, $R = 8$ MW/h. Each shaded set of bars represents the entries of the corresponding matrix $D_i$ for baseload generation (black), peaking generation (grey) and storage (white). The sub-diagonal elements in case 2, which are not permitted in case 1, describe how prediction errors should continue to affect the energy storage unit’s behaviour for time steps after they have been revealed.
4.2 Large-scale solutions using ADMM

Section 2.3.3 described the scaling properties of the affine policy optimization problem, noting that for large problems the number of variables can potentially become prohibitive for a centralized optimization. One solution to this is to use a distributed optimization method to break the problem down into smaller, coupled subproblems. The Alternating Direction Method of Multipliers (ADMM, see Section 5.2.2) was applied to this problem in order to investigate its suitability for a large-scale implementation.

The decomposition is achieved by considering problem (2.10) and duplicating variables $D_i$ and $e_i$ for $i = 1, \ldots, N_p$ into local variables held by the participating devices, and global variables held by a coordinator. Borrowing the $(x, z)$ notation from [BPC$^{+10}$] let the local copies be denoted $D_i^x$ and $e_i^x$, and the global copies $D_i^z$ and $e_i^z$. It is not necessary to duplicate matrix $Z$. Then the following problem has the same solution as (2.10), since the additional quadratic penalty in the objective function is zero for any feasible solution:

$$
\min_{Z \geq 0, \ (D_i^x, e_i^z) \in F_i(x_i^0), (D_i^z, e_i^z)} \sum_{i=1}^{N_p} \tilde{J}_i(x_i^0, D_i^x, e_i^z) + \frac{\rho_D}{2} \|D_i^x - D_i^z\|^2_2 + \frac{\rho_e}{2} \|e_i^x - e_i^z\|^2_2
$$

s.t. $\sum_{i=1}^{N_p} (r_i + C_i A_i x_i^0 + C_i B_i e_i^z) = 0$, \hspace{1cm} (4.14b)

$\sum_{i=1}^{N_p} (G_i + C_i B_i D_i^z) = 0$, \hspace{1cm} (4.14c)

$Z'h + \sum_{i=1}^{N_p} \Gamma_i C_i B_i e_i^z \leq \bar{p} - \sum_{i=1}^{N_p} \Gamma_i (r_i + C_i A_i x_i^0)$, \hspace{1cm} (4.14d)

$\sum_{i=1}^{N_p} \Gamma_i (G_i + C_i B_i D_i^z) = Z'S$, \hspace{1cm} (4.14e)

$D_i^x - D_i^z = 0, \quad i = 1, \ldots, N_p$, \hspace{1cm} (4.14f)

$e_i^x - e_i^z = 0, \quad i = 1, \ldots, N_p$. \hspace{1cm} (4.14g)

where $\rho_D, \rho_e > 0$ are penalty parameters. Note that the cost function terms $\tilde{J}_i(x_i^0, D_i^x, e_i^z)$ and constraint sets $F_i(x_i^0)$ have been written in terms of the local
copies, which have been superscripted with \( x \). The coupling constraints \((4.14b)\) to \((4.14e)\) have been written in terms of global copies superscripted with \( z \).

If constraints \((4.14f)\) and \((4.14g)\) are now relaxed and penalized with the Lagrange parameters \( \Lambda_i \) and \( \lambda_i \) respectively, the following augmented partial Lagrangian function can be written:

\[
L_\rho(x, z, \lambda) := \sum_{i=1}^{N_p} \tilde{J}_i(x^i_0, D^i_x, e^i_x) + \langle \Lambda_i, D^i_x - D^i_z \rangle + \lambda_i' (e^i_x - e^i_z)
+ \frac{\rho_D}{2} ||D^i_x - D^i_z||^2 + \frac{\rho_e}{2} ||e^i_x - e^i_z||^2
\]

\((4.15)\)

In the above, \( x \) is used as shorthand to denote the local variables \( \{ (D^i_x, e^i_x) \}_{i=1}^{N_p} \), and \( z \) is used to denote the global variables \( Z \) and \( \{ (D^i_z, e^i_z) \}_{i=1}^{N_p} \). The symbol \( \lambda \) (with or without a superscript \( k \) to denote the value at a particular iteration) is used to denote the multipliers \( \{(\Lambda_i, \lambda_i)\}_{i=1}^{N_p} \). The ADMM algorithm proceeds according to the steps described in Section 5.2.2:

0. Initialize iteration counter \( k = 0 \), and set \( \Lambda^0_i, \lambda^0_i \), and \( z \) to some initial guess.

1. \( x^{k+1} \leftarrow \arg \min_x L_\rho(x, z^k, \lambda^k) \). This is achieved by solving, for each participant \( i \), the following:

\[
\min_{(D^i_x, e^i_x) \in F_i(x^i_0)} \tilde{J}_i(x^i_0, D^i_x, e^i_x) + \langle \Lambda^k_i, D^i_x - D^i_z \rangle + \lambda^k_i' (e^i_x - e^i_z)
+ \frac{\rho_D}{2} ||D^i_x - D^i_z||^2 + \frac{\rho_e}{2} ||e^i_x - e^i_z||^2
\]

s.t. constraints \((4.14b)\) to \((4.14e)\)

Note that if no line constraints are present, the variable \( Z \) is not needed, and this problem becomes a quadratic minimization with affine equality constraints \((4.14b)\) and \((4.14c)\). Problems of this type can be solved using the explicit solution, a pre-computed projection operation.
3. Check convergence using a criterion based on the residuals $||x - z||$. If not yet converged, perform $\Lambda_{i}^{k+1} \leftarrow \Lambda_{i}^{k} + \rho_{D}(D_{i}x_{i}^{k+1} - D_{i}z_{i}^{k+1})$ and $\lambda_{i}^{k+1} \leftarrow \lambda_{i}^{k} + \rho_{e}(e_{i}^{z,k+1} - e_{i}^{z,k+1})$ for all $i$, increment $k$, and return to step 1.

A possible “market-like” interpretation of this along the lines of the price negotiations described in Chapter 5 is as follows. Step 1 is a separable minimization problem across participants 1, ..., $N_{p}$ that trades off local costs against an “augmented price” term. Step 2 is a problem solved by the operator to find a solution near to those proposed by market participants that is compatible with the network constraints. Step 3 updates the price parameters in order to bring the results of steps 1 and 2 closer together, i.e., to make sure that the participants’ plans can be implemented on the network.

The separable nature of step 1, and the relatively simple form of step 2, together imply that an arbitrarily large-scale version of problem (2.10) can be solved. In particular, if step 1 is totally parallelized across participants $i$ and those subproblems are tractable, then the only possible limitation to the approach is the projection in step 2, which may become unmanageable if the combined dimension of $D_{i}$, $e_{i}$, and $Z$ grows to be too large. In such cases, though, further decomposition possibilities along the lines of [KCLB12] are likely to be viable.

The convergence performance of ADMM is known to depend somewhat on the choice of parameter $\rho$ used to augment the Lagrangian of the original optimization problem. Fig. 4.9 shows the number of iterations needed for the coupling constraints (2.10b) and (2.10c) to be satisfied with an acceptable relative tolerance for various choices of $\rho_{D}$ and $\rho_{e}$, along with a more elaborate scheme in which the $\rho$ parameter values are controlled dynamically. Fig. 4.10 shows typical convergence behaviour. The performance appears to be rather good for many parameter choices, in that the number of iterations needed rarely exceeds a few hundred. This compares favourably with standard dual ascent methods, which typically require thousands of iterations. However, it is important to note that ADMM does not have the same convenient market mechanism interpretation as standard dual decomposition. Discussion of these issues is deferred to Part III of this dissertation.

\[1\] The notion of a “best possible choice” of parameter under simplified problem assumptions was investigated in [TGS+13].
Figure 4.9: Number of iterations needed by ADMM algorithm to converge to a relative tolerance of $10^{-3}$ for various-sized instances of problem (2.10) [Hoh13].

Figure 4.10: Residuals for constraints (2.10b) (“nominal power match”) and (2.10c) (“disturbance power match”) for a typical problem instance [Hoh13].
4.2.1 Application to the unit commitment problem

It has been reported \cite{BPC+2010} that ADMM may have useful properties for solving non-convex problems, including those such as the unit commitment problem \eqref{eq:unit-commitment} that feature binary variables. A potentially attractive approach, which was trialled in the thesis \cite{Hoh2013}, is to duplicate each binary variable in such a way that the minimization over one copy becomes analytic, while the other copy is relaxed to the interval \([0, 1]\). The quadratic \(\rho\) penalty has the effect of driving the relaxed copy to either 0 or 1, so that over the whole problem a relatively large number of the binary variables fully or nearly converge to feasible values. A rounding procedure is then needed to allocate feasible values to the “un-rounded” variables.

4.3 Double-sided reserves

In reality most power system operators distinguish between upward and downward regulation reserves, which act when the net load turns out to be respectively higher and lower than predicted. However the formulation given in this chapter has not yet addressed this possibility, and as a result requires each market participant to make the same margins for upward and downward regulation available. This means the same linear response to the realized uncertainty should be provided whether the regulation needed is positive or negative, which often results in conservative behaviour.

This issue will now be addressed by extending the optimization formulation so that the response can differ for upward and downward disturbances. This allows asymmetric margins and costs to be taken into account. For example, generators and wind turbines running at maximum possible output could provide downward regulation very cheaply without being forced to allow a margin for upward regulation. It may in these cases be more desirable to use some other unit to provide upward regulation.

Problem \eqref{eq:system-state} can be rewritten in order to take this into account, and a sketch of a derivation will now be given for a box uncertainty set, inspired by a similar strategy employed for affinely-adjustable RO in \cite{CZ2009} and more specifically to robust MPC in \cite{WOS2010}. The uncertainty set has until this point been modelled as \(\Delta := \{\delta \in \mathbb{R}^{N_\delta} \mid S\delta \leq h\}\). Now assume that this set describes a box in \(\mathbb{R}^{N_\delta}\) with the faces aligned with the principal axes. Then the set can be described as
\[ \Delta := \{ \delta^+ + \delta^- \mid \delta^+ \in \Delta^+, \delta^- \in \Delta^- \}, \]
where \[ \Delta^+ := \Delta \cap \mathbb{R}^N_{\geq 0} \] and \[ \Delta^- := \Delta \cap \mathbb{R}^N_{\leq 0}. \]

A new higher-dimensional uncertainty
\[ \tilde{\delta} := [\delta^+ \delta^-] \in \tilde{\Delta} \subset \mathbb{R}^{2N_\delta} \]
can now be used to formulate a problem that distinguishes between the positive and negative components of \( \tilde{\delta} \), with separate feedback matrices \( D_i^+ \) and \( D_i^- \) governing each participant \( i \)'s response to positive and negative realizations. The dimension of the uncertainty set is now \( 2N_\delta \), and the corresponding higher-dimensional uncertainty set is given by
\[ \tilde{\Delta} = \{ [\tilde{\delta}^+ \tilde{\delta}^-] \mid \tilde{\delta}^+ \in \Delta^+, \tilde{\delta}^- \in \Delta^-, \delta^+\delta^- = 0 \}. \]

The complementarity constraint, ensuring that \( \delta^+ \) and \( \delta^- \) cannot be non-zero simultaneously, renders the uncertainty set \( \tilde{\Delta} \) non-convex. This would at first sight appear to prevent the robustness reformulation leading to (2.10) being used. Fortunately though, as noted in [WOS10, GWK10], \( \tilde{\Delta} \) can simply be replaced by its convex hull, as illustrated here in Fig. 4.11; it can easily be shown that this yields the same solution as would have been obtained had the complementarity constraint been enforced. In other words, the constraints of the upward and downward reserve policy problem are the same regardless of whether the formulation accounts for the fact that upward and downward disturbances cannot happen simultaneously.

![Figure 4.11: One dimensional disturbance \( w \); segregation into set of positive and negative disturbances \( v^1 \) and \( v^2 \); convex hull of higher-dimensional disturbance set [WOS10]. Image © IEEE 2010.](image-url)

The practical implementation of such a segregated approach would of course rely on estimates of \( \mathbb{E}[\tilde{\delta}] \) and \( \mathbb{E}[\tilde{\delta}^+\tilde{\delta}^-] \), which due to their higher dimension require a more detailed model than that needed for the original disturbance \( \delta \). That said, half of the entries of \( \mathbb{E}[\tilde{\delta}^+\tilde{\delta}^-] \), corresponding to products across elements of \( \delta^+ \) and \( \delta^- \), can be set to zero since \( \delta^+ \) and \( \delta^- \) are never simultaneously non-zero.
Part III

Distributed Optimal Power Flow and Market Mechanisms
Chapter 5

Distributed Solutions to Welfare Maximization Problems

Distributed optimization is a means of solving a large optimization problem by decomposition into a coordinated set of multiple smaller problems. This chapter is concerned with the treatment of a short-term electricity market along these lines, where the larger optimization problem is a welfare maximization whose solution is in the interests of society as a whole, and the smaller problems represent profit or utility maximizations carried out by participants.

Section 5.1 states a generic welfare maximization problem and reviews the theory of efficient prices. Section 5.2 describes how an auction or a variety of iterative methods could be used to solve such a problem, referring to short-term electricity markets as an example. In the remainder of this part of the dissertation, Chapter 6 applies iterative pricing methods to multi-period power flow problems.

5.1 Welfare maximization problems

Markets exist in order to facilitate mutually-beneficial transactions between participants. This section presents the concept of a welfare maximization problem, and shows how efficient prices, quantities, and profits arise from a solution to this underlying problem. The discussion that follows assumes that all traded goods are commodities. A commodity is an interchangeable good for which small qualitative differences are not important to suppliers or consumers.
5.1.1 Problem statement

In economic theory, markets should bring about an adequate solution to an underly-
ging optimization problem whose objective is to maximize the total welfare of a sys-
tem of participants. Such a (social) welfare maximization problem can be written in the generic form

\[
\begin{align*}
\min_{x_1, \ldots, x_N} & \sum_{i=1}^{N} f_i(x_i) \\
\text{subject to:} & \sum_{i=1}^{N} A_i x_i = b, \\
& x_i \in \mathcal{X}_i, \quad i = 1, \ldots, N,
\end{align*}
\]

(5.1)

where each \( x_i \in \mathbb{R}^{n_i} \), \( f_i : \mathbb{R}^{n_i} \to \mathbb{R} \), and \( \mathcal{X}_i \subseteq \mathbb{R}^{n_i} \). The functions \( f_i \) commonly represent costs or (negated) utilities to participants indexed by \( i \). The coupling constraint has \( m \) rows corresponding to the number of markets that must clear: \( b \in \mathbb{R}^m \) and \( A_i \in \mathbb{R}^{m \times n_i} \). Note that although \( m \) may correspond to the number of different commodities traded, a participant may have additional decision variables to optimize over, so that \( n_i \neq m \).

Solving problem (5.1) is equivalent to minimizing the total costs of producing, and/or maximizing the total utility of consuming, resources \( x_i \) under some constraint ensuring that supply and demand are consistent.

In some applications, including power system optimization, there may be additional inequality constraints coupling the participant actions. For clarity, though, the discussion in this section will be limited to problem (5.1), since the extension to additional inequality constraints is straightforward.

5.1.2 Elastic and inelastic supply and demand

The functions \( f_i \) described above can be used to represent elastic supply and demand, or in other words supply and demand that can be influenced in some way. Inelastic supply and demand levels \( x^{inel}_i \) can be viewed as special cases in which failure to incorporate the supply, or satisfy the demand, would incur unacceptable system-wide costs. In the case where “unacceptable” is taken to mean infinite, the situation can be modelled using an extended-value cost function of the form

\[
f_i(x_i) = \begin{cases} 
\text{constant} & \text{for } x_i = x^{inel}_i \\
+\infty & \text{otherwise.}
\end{cases}
\]

90
Solving an optimization problem of such a form is equivalent to satisfying the constraint that $x_i = x_i^{\text{inel}}$ in order to avoid an infinite cost, and the value of the constant becomes irrelevant.

Inelastic supply and demand functions can therefore be represented in (5.1) by altering the right-hand side of the constraint $\sum_{i=1}^{N} A_i x_i = b$.

### 5.1.3 The central planner’s solution

One way of solving problem (5.1) is to ask participants for their functions $f_i$ and constraint sets $X_i$, and solve a centralized optimization problem on their behalf to obtain a welfare-maximizing solution $x_i^*$ for each participant. This solution is often known as the central (or social) planner’s solution, since the planner must then compel the participants to implement the result.

Note that transfers of money within a system of agents do not affect total welfare, only the way it is distributed amongst the agents. Therefore it is not necessary for the central planner to consider payments between participants for the purpose of solving the problem; these can be determined after the result is known.

### 5.1.4 Profit maximization and price coordination

The central planner’s solution appears to be convenient but has the following practical drawbacks:

1. The cost functions $f_i$ may be private to participants $i$, or may be too elaborate or subtle to convey concisely to a central planner.

2. Compelling participants to implement a solution $x_i^*$ implies ownership or control of the participants, something that may not be practical or desirable.

3. A practical criticism of centrally planned economies is that they discourage innovation on the part of participants, since the participants do not know whether the central planner will allow innovators to retain the benefits.

4. The scale of the optimization problem, meaning large $N$ or $n_i$, may make it difficult for the central planner to solve.
An alternative to this is to allow prices or other signals to encourage market participants to undertake efficient actions in their own interests. If such signals can be found using some mechanism without introducing the problems described above, conditions for which will be seen in the coming discussion, then the mechanism has advantages over a centralized solution. Any such mechanism could be referred to as a market mechanism.

Consider a partial Lagrangian of problem (5.1) in which the constraints $x_i \in X_i$ are retained but the constraint $\sum_{i=1}^{N} A_i x_i = b$ has been relaxed and the corresponding multiplier $\lambda \in \mathbb{R}^m$ introduced:

$$L(x_1, \ldots, x_N, \lambda) = \sum_{i=1}^{N} f_i(x_i) + \lambda' \left( \sum_{i=1}^{N} A_i x_i - b \right) \quad (5.2)$$

This function is separable between the participants:

$$L(x_1, \ldots, x_N, \lambda) = -\lambda' b + \sum_{i=1}^{N} \left[ f_i(x_i) + \lambda' A_i x_i \right] \quad (5.3)$$

Consider minimizing $L(x_1, \ldots, x_N, \lambda)$ over the vectors $x_i \in X_i$ for a given $\lambda$. From equation (5.3) it can be seen that this act coincides with the maximization of each participant $i$’s surplus

$$f_i(x_i) - \tilde{\lambda}_i x_i \quad (5.4)$$

where $\tilde{\lambda}_i := -A_i \lambda$ is the price seen by participant $i$ for commodities or states in vector $x_i$. For consumers, $x_i < 0$, and $f_i(x_i)$ represents the negated utility of consuming $x_i$ and $\tilde{\lambda}_i x_i > 0$ (assuming a positive price) is a payment made for the privilege. For producers, $x_i > 0$, $f_i(x_i)$ represents costs for producing $x_i$ and $\tilde{\lambda}_i x_i < 0$ represents income from selling it. In both cases, the surplus (5.4) gives the benefit accruing to participant $i$ from the market outcome. The market allows such surpluses to be created.

For any value of the multiplier $\lambda$, it can easily be shown that the minimum value of $L(x_1, \ldots, x_N, \lambda)$ over $x_i \in X_i$, $x_i^*(\lambda)$, is a lower bound for the optimal value of problem (5.1); see for example [BV09], §5.1.3.

In the case that each function $f_i$ is convex and each set $X_i$ is convex, and an additional constraint qualification such as Slater’s condition holds, strong duality holds for problem (5.1). An implication of this is that for a $\lambda^*$ that gives the maximum
possible lower bound, the optimal solution to \( (5.1) \) is among the minimizers of \( (5.2) \).

Practically, this means that under these conditions there exist surplus-maximizing actions for the participants that not only satisfy the market clearing constraint \( \sum_{i=1}^{N} A_i x_i - b \) but also represent an optimal solution to problem \( (5.1) \). The vector \( \lambda^\star \) is referred to as a vector of optimal dual variables, and the corresponding vectors seen by the participants, \( \tilde{\lambda}_i^\star := -A_i \lambda^\star \), are the optimal prices. Note that these interpretations may change slightly depending on what modelling role the matrices \( A_i \) play in a given application.

The above discussion paraphrases well-known convex optimization theory, with the relevant insight that when prices are in some sense “correct,” the selfish actions of the participants coincide with the global concern of the market operator, namely maximizing the welfare of the system as a whole and clearing the market.

### 5.1.5 Pricing in the absence of (strict) convexity

This dissertation is primarily concerned with markets that operate under standard convexity assumptions. If the underlying welfare optimization problem does not possess certain convexity properties, extra difficulties are introduced into the market mechanism design problem. This is because there is no longer any guarantee that a dual-based method will facilitate convergence to a global primal optimum, or even provide a feasible solution. The issues encountered for linear and non-convex problems are now briefly discussed.

#### Non-strictly-convex problems

Standard dual approaches [Lem01], when applied to market optimization problems, relies on a dual variable (or price) to indirectly satisfy the constraint that supply matches demand. The market clears when the prices expected by participants induce them to produce or consume exactly balancing quantities of the commodity. Clearly it is easier to clear the market to within a given tolerance if the actions of the participants \( x_i^\star(\lambda) \) are not too sensitive to the prices \( \lambda \) proposed. Roughly speaking, the smaller the curvature of the cost function \( f_i(x_i) \), the greater the sensitivity of participant \( i \) to changes in price, as illustrated in Fig. 5.1.
5. Distributed Solutions to Welfare Maximization Problems

Figure 5.1: Profit maximizing response \( x^*(\lambda) := \arg \min_x \{ f(x) - \lambda x \} \) to a price \( \lambda \) for three different participants with scalar outputs restricted to the range \( x_{\text{min}} \leq x \leq x_{\text{max}} \).

a) A strictly convex cost function \( f(x) \) leads to a continuous response to prices; b) and c) affine and concave cost functions lead to discontinuous responses. Behaviour in the multi-dimensional case depends more elaborately on the cost function and the constraints.

In the case where participant \( i \)'s cost function is linear in directions orthogonal to the coupling constraint, and it has polytopic local constraints \( \mathcal{X}_i \), \( x^*_i(\lambda) \) will be a discontinuous function of \( \lambda \). At the dual optimal solution \( \lambda^* \), a solution satisfying the coupling constraint can be found among the minimizers \( \{ x^*_i(\lambda) \}_{i=1}^{N_p} \), but the market participants cannot be expected to find or favour this solution in the case that the optimal solution set is an entire facet of the feasible set. For this reason, a dual-based method without some kind of regularization or “prox” term may not clear the market satisfactorily when some components of the costs are linear and the feasible set is faceted (e.g. polytopic).
Non-convex problems

Problems in which the feasible set or objective function are non-convex are even less amenable to solution using dual-based methods, since in contrast to the case described above a duality gap generally exists. In such cases the dual optimal solution may not be useful for extracting a good (or even feasible) primal solution.

Some special situations, for example where the cost function is linear and separable and some of the optimization variables are binary, have been studied, and results concerning the duality gap have been established. Bertsekas and Sandell proved that the relative duality gap, that is the duality gap divided by the problem size, diminishes as the problem size grows [BS82]. In other words, as long as there are enough market participants, price signals from a market may still be a good means of inducing market participants to collectively maximize social welfare. This agrees with the intuition that although many real markets feature many sources of non-convexity they often operate efficiently.

The result is relevant to this dissertation in that the description fits an electricity market where binary unit commitment decisions must also be made by generation companies, and has also motivated the use of dual-based methods to solve a centralized unit commitment problem [LSBP82 BLSJP83]. They may therefore be of relevance to the UC extension of the reserve policy problem described in Section 4.1.
5.2 Solution mechanisms

This section describes some existing methods for solving problems of the form (5.1) without the need for a central planner. Instead these methods use a market operator, or coordinator, to try and reach an efficient solution. The methods surveyed are those most relevant to the power systems applications in this thesis; many other mechanisms of course exists, some involving money and some not (see Nisan et al. [NRTV07] for an overview).

5.2.1 Double-sided auction

A participant in a double-sided auction submits bids, which consist of a quantity and price pair, and if the bid is accepted the participant is bound to produce (or consume) the quantity bid. The market operator constructs so-called ladders from these bids for the supply side (in decreasing price order) and demand side (in increasing price order), and these are overlaid as shown in Fig. 5.2. A clearing price and quantity is returned from the crossover point of the two ladders. Bids to the left of the crossover point are accepted, since these match consumers with producers who are willing to accept less than they have offered to pay, whereas bids to the right are rejected. All producers are paid the clearing price per unit, which is at least as much as they asked for, and all consumers pay the clearing price per unit, which is at most as much as they were willing to pay. The shaded areas on Fig. 5.2 represent surpluses made by the participants, i.e., the amount from which they profited from the difference between their bid prices and the clearing prices.

It is straightforward to show that if the bids approximate cost or utility functions representable using convex functions \(f_i(x_i)\), then the act of sorting the bids into supply and demand curves and clearing the market in this way approximately maximizes the welfare outcome of the market, i.e., returns a good solution to problem (5.1). The areas of the consumer and producer surplus regions on Fig. 5.2 represent approximately the sums of the quantities (5.4) seen by each participant, and the clearing price is approximately \(\lambda^*\).

However, when costs and constraints for market participants are coupled for several commodities, the most relevant example to this dissertation being electrical energy

---

1There are conventions for dealing with the case where this crossover is not a single point, see [PMNP06].
5.2. Solution mechanisms

traded over consecutive periods, a set of one-shot double auctions cannot clear optimal prices. This is because unless all participants possess a detailed model of all other participants’ behaviours they will not be able to submit suitable bids to trade off costs against expected income. This issue is discussed in detail below.

5.2.2 Iterative pricing mechanisms

The participants’ optimal bidding strategies in an auction are in general a function of the clearing prices, which are not known before the market clears. This is true of an electricity spot market in which prices and quantities of electricity are to be cleared for operating periods of, say, one hour each. The study of iterative pricing mechanisms for electricity will now briefly be motivated.

An energy storage operator’s strategy is to buy power at low prices and sell it back at high prices. However the price variations he wishes to profit from depend on the bidding actions of all participants, which are not known until the market has cleared and the bids can no longer be revised. Similarly, a power generation company operating difficult-to-ramp plants would prefer to know, as it chooses its bids, whether the sequence of hourly commitments it is making will be technically feasible and economically attractive, but must attempt to fulfil this objective without foreknowledge of the auction outcome. A mechanism attempting to find these
efficient clearing prices is called a price discovery mechanism.

In the earlier years of liberalized electricity markets, demand patterns were relatively predictable, only changing with the weather and seasons, and energy in-feeds from uncontrolled renewables were small. This led to a relatively predictable variation of both net load and prices throughout a given day, which in turn allowed generation companies to plan ramping and switching actions with relative certainty as to their incomes; and allowed energy storage companies such as pumped hydro operators to plan daily and seasonal patterns for buying and selling power, to take advantage of known price differences.

The transition to a large share of intermittent renewables in many countries is altering this situation. Wildly fluctuating in-feeds of “free” power break the regular daily patterns in net load that were relied on for scheduling purposes. As a result, market participants can no longer rely on their previous bidding strategies, which were based on assumptions about the price variation throughout the day. Instead, the prices that arise from the markets will be related to the renewable in-feeds and will not follow predictable patterns \[WNPK^{+12}\]. Since the mechanism used to clear the market does not systematically take into account the time-coupled nature of participants’ costs and constraints, the prices arrived at may be far from efficient. This issue motivates the further study of price discovery mechanisms for intraday markets, including the iterative mechanisms described below.

**Tâtonnement and Lagrangian relaxation**

It is an obvious intuition that if there is too much supply of a given commodity, prices should fall such that demand increases or supply decreases, so that an accumulation of excess goods is avoided. This intuition formed the basis for a pricing theory known as tâtonnement (\[Wal54\], see Section \[1.3.3\]). A market operator (either fictitious, representing the sentiments of a group of traders, or real) notes the net shortfall or surplus of a commodity, and adjusts the price appropriately until the supply matches the demand. In fact, some markets have been known to function with a real price coordinator in this way, an example being food commodities in Japan \[EW07\].

The coordinator announces prices and waits for participants to propose, or promise, their optimal profit- or utility-maximizing actions with respect to these prices. If there is a mismatch between supply and demand, the coordinator uses a rule, or
5.2. Solution mechanisms

perhaps just intuition, to adjust the price, which is then announced again to the market. After a number of iterations of this process, when the supply and demand match to within some given tolerance, the auction is terminated, the sellers are matched with buyers, and the promised transactions are made. Its 20th-century generalization is known as Lagrangian relaxation, and can be interpreted as a means of solving problems of the form (5.1) as follows:

1. Initialize multiplier $\lambda_k$ corresponding to the constraint $\sum_{i=1}^{N} A_i x_i = b$ to some starting guess.

2. At iteration $k$, minimize Lagrangian $L(x_1, \ldots, x_N, \lambda_k)$ with respect to primal variables $x_i \in X_i$.

3. Take the solution $x_i^*(\lambda_k)$ and update the multiplier according to

   $$
   \lambda_{k+1} \leftarrow \lambda_k + \alpha_k \frac{\sum_{i=1}^{N} A_i x_i^*(\lambda_k) - b}{|| \sum_{i=1}^{N} A_i x_i^*(\lambda_k) - b ||}
   $$

4. Return to step 2 unless some convergence criterion is met, such as $|| \sum_{i=1}^{N} A_i x_i^*(\lambda_k) - b || \leq \epsilon$, in which case terminate the algorithm.

The price update step size $\alpha_k$ is chosen according to a rule designed to guarantee convergence. A common choice is $\alpha_k = \alpha_0 / \sqrt{k}$, which produces step lengths satisfying the sufficient conditions given by e.g. Bertsekas [Ber99]:

$$
\sum_{k=1}^{\infty} \alpha_k = \infty, \quad \sum_{k=1}^{\infty} \alpha_k^2 < \infty, \quad \text{or} \quad \lim_{k \to \infty} \alpha_k = 0.
$$

The intuition behind these conditions is that the multipliers should be able to reach the optimal solution $\lambda^*$ from any starting point, since the steps are able to cover an arbitrarily large distance in the dual space, but should ultimately terminate in some small neighbourhood of $\lambda^*$ as the step size decreases towards zero. The same arguments extend to the case where the optimal multiplier is not unique.

Proofs of convergence of this algorithm to a dual-optimal value $\lambda^*$ are available both for the case where the dual update is as stated in step 3 above, and in the alternative case where the constraint violation is not normalized:

$$
\lambda_{k+1} \leftarrow \lambda_k + \alpha_k \left( \sum_{i=1}^{N} A_i x_i^*(\lambda_k) - b \right).
$$
These can be found under Theorems 2.2 and 2.3 by Shor in [SKR85], but were known even before the 1980s. Convergence of the primal variables to an optimal value satisfying the coupling constraint \( \sum_{i=1}^{N} A_i x^*_i(\lambda^*) = b \) relies on rather strict conditions, namely that for all \( i \) the feasible sets \( \mathcal{X}_i \) must be convex, and the cost functions must be strictly convex. Even when these conditions hold, convergence is only asymptotic, with a rate that is highly dependent on both the minimum curvature and the conditioning of the Hessian. In practice, then, the performance of Lagrangian relaxation is highly problem-specific and may be poor, needing thousands of iterations for a usable result. That said, numerous methods for improving the “vanilla” implementation above have been proposed (see Section 1.3.3), and its use in a few real markets suggests that the assumptions on which it relies are broadly reasonable.

### ADMM

The alternating direction method of multipliers (ADMM) is one of several decomposition approaches able to overcome some of the convergence limitations of standard Lagrangian relaxation by introducing a so-called regularization term to the problem. The method was originally reported in the 1970s [GM76] but has recently been re-popularized by Boyd [BPC10] as a means of exploiting modern computational capabilities to solve large-scale problems. Regularization approaches rely on the fact that modifications to problem (5.1), such as

\[
\min_{x_1, \ldots, x_N} \sum_{i=1}^{N} f_i(x_i) + \frac{\rho}{2} \| \sum_{i=1}^{N} A_i x_i - b \|_2^2
\]

subject to:

\[\sum_{i=1}^{N} A_i x_i = b, \quad x_i \in \mathcal{X}_i, \quad i = 1, \ldots, N,\]  

(5.5)

can be made such that the modified problem has the same solution as problem (5.1), since the extra norm term is equal to zero for all feasible solutions. In the case of a quadratic modification, the additional term adds curvature in directions orthogonal to the linear constraint, which improves the convergence behaviour of dual-based algorithms.

A typical way of using ADMM to break down large problems is to duplicate some of the variables \( x \) in the original problem, creating a new vector \( z \). The linear constraint is then augmented to enforce consistency between \( x \) and \( z \). For a
5.2. Solution mechanisms

generic problem of the form (taken from [BPC+10])

\[
\min_{x,z} \quad f(x) + g(z)
\]

subject to: \[Ax + Bz = c,\] \hspace{1cm} \text{(5.6)}

the algorithm steps are as follows, where

\[
L_\rho(x, z, \rho) := f(x) + g(z) + \lambda^\prime(\rho) + \frac{\rho}{2}||Ax + Bz - c||^2
\]

is the augmented Lagrangian of problem (5.6), and \(\rho > 0\):

0. Initialize \(k = 0\), set \(\lambda^0\) and \(z^0\) to some initial guess.
1. \(x^{k+1} \leftarrow \text{arg min}_x L_\rho(x, z^k, \lambda^k)\).
2. \(z^{k+1} \leftarrow \text{arg min}_z L_\rho(x^{k+1}, z, \lambda^k)\).
3. Check convergence using a criterion based on the residual \(Ax + Bz - c\), a dual residual and possibly a duality gap term. If not yet converged, perform \(\lambda^{k+1} \leftarrow \lambda^k + \rho(Ax^{k+1} + Bz^{k+1} - c)\), increment \(k\), and return to step 1.

A proof of convergence of the algorithm can be found in Appendix A of [BPC+10]. Note that step 2 uses the \(x\) result from step 1, so that the augmented Lagrangian is minimized sequentially over \(x\) and then over \(z\). Note also that \(f(x)\) and/or \(g(z)\) are defined over all space (i.e. the entirety of \(\mathbb{R}^n\) where \(n\) is the appropriate dimension) but may implicitly include a constraint set by using the extended-value extension [BV09 §3.1.2]. For example, to enforce \(x \in \mathcal{X}\) one can define \(f(x) = +\infty\) for all \(x \notin \mathcal{X}\).

An attractive interpretation is to assume the variable \(z\) is seen by the coordinator, who tries to enforce coupling constraints between participants, whereas the variable \(x\) may consist of private state information held by agents. For example, in a problem of the form (5.1), \(x\) can include all variables \(x_i \in \mathcal{X}_i\) and the generic constraint \(Ax + Bz = c\) can be used to represent the coupling constraint, written \(\sum_{i=1}^N A_i z_i = b\), along with a consistency constraint that \(x_i = z_i\) for \(i = 1, \ldots, N\). It can be shown that in this case the augmented Lagrangian \(L_\rho(x, z, \lambda)\) is separable in \(x\) for fixed \(z\), which makes each step of the algorithm rather straightforward. Step 1 becomes a separable (and parallelizable) optimization for each market participant \(i\); step 2 becomes an equality-constrained quadratic program, for which an analytic solution exists; step 3 just requires a gathering of the solutions and an evaluation of the residual.
5. Distributed Solutions to Welfare Maximization Problems

The conclusion from the discussion above is that ADMM relies on the same communication pattern as conventional Lagrangian relaxation, namely passing penalty information known by the coordinator (in this case $\lambda$ and an extra quadratic term) to agents and receiving from them an optimal $x^*$ for that iteration. However its performance is markedly better, in that it generally takes an order of magnitude fewer iterations to obtain a usable solution. Moreover its convergence behaviour is rather insensitive to the choice of $\rho$, the only parameter that the coordinator needs to specify. In contrast, the convergence of LR depends strongly on the step size rule and initialization, and heuristics are often needed to produce good results [WGMM12a].

ADMM’s drawback in a market context, and the reason it was not employed in [WGMM12a], is that each local minimization performed by the $N$ participants in step 1 of the algorithm now contains a quadratic term that penalizes deviations between $x_i$, the agent’s own profit-maximizing solution, and $z_i$, the coordinator’s version of the variable that satisfies the coupling constraint. However, it is important to note that this penalty will disappear (has value 0) when the algorithm has converged and $x = z$. Assuming the participants know this from prior experience, one can infer that they have no reason to take it into account during intermediate iterations. In other words, the participants could instead perform the same profit-maximizing optimization as in conventional LR, and the ADMM algorithm may fail to converge.

Bialek’s blame allocation

Other market clearing methods are specific to given types of market. Bialek proposed a method for tracing responsibility for flows in a network which can be applied to electricity grids [Bia96]. Given a particular power flow configuration, this consists of the following steps:

1. Determine the direction of power flows in all lines by solving a load flow equation.

2. For each node, identify the incoming and outgoing flows, and assume that the outgoing flows are each a perfect mixture of the incoming flows.

3. Use a matrix inversion technique to arrive at a vector of contributions of each source to each line flow.
This can be used as the basis for a price update rule that quickly resolves cases where transmission lines are overloaded by a proposed power flow, by using the contributions in step 3 above to determine a congestion price for each participant, depending on its location on the grid. Such an approach was presented in [WMJM10] as a means of solving problems of the form (5.1) featuring additional coupling inequality constraints arising from line congestion limits.

In addition to the congestion price, the method also measured the shortfall or excess of net power injected into the network, and updated a base price associated with this constraint as in the usual tâtonnement approach described above. Then, in the manner of conventional LMPs consisting of a base price plus a congestion price, the net power mismatch over the network and the line congestion blame allocations contribute to the price seen by each participant.

The mechanism appears to have very good convergence properties, in that prices that induce participants to satisfy the system constraints are found in only a few tens of iterations. However, the solution obtained is generally suboptimal (sub-optimality of the order 20% was observed anecdotally for the case studies tried). The mathematical properties of the allocation step are difficult to analyze due to the fact that the allocation is highly dependent on the sign of the power flow in each line. That said, a recent analysis of the properties of the allocation algorithm based on matrix power series was made in [ADBO10], and may be of further use for designing price update rules.

**Margin-based mechanisms**

Another approach a coordinator could apply when issuing prices is to ask market participants to make offers of power production or consumption which are valid over a given range described using a compact set \( \Theta_i \) containing the origin. At iteration \( k \) the participant bids \( x_i^{\ast \Theta_i}(\lambda_k) \) would be made in the knowledge that the coordinator could request \( x_i^{\ast \Theta_i}(\lambda_k) + \theta \) for any \( \theta \in \Theta_i \), accounting for the appropriate robustness margin in its decision. The coordinator could then choose a \( \theta \) that clears the market (projects the solution onto the constraint \( \sum_{i=1}^{N} A_i x_i = b \)) without having to ask for more bids, as long as the sets \( \Theta_i \) are sufficiently large to encompass a point that clears the market. A choice of \( \theta = 0 \) implies the coordinator’s acceptance of the nominal plan proposed by the participant.

The mechanism has the following attractive property. If each \( \Theta_i \) could be con-
tracted to the single element \( \{0\} \) in some way while keeping the market-clearing projection feasible, then the optimal welfare-maximizing actions of the participants would be discovered, as their actions \( x^\star_{\{0\}}(\lambda_k) \) would then result from a minimization of the Lagrangian (5.3) of the original problem. The result would be a market mechanism that can generate a feasible clearing at every iteration (unlike standard tâtonnement, which only becomes feasible asymptotically), and whose solution can get closer to optimal with every iteration. This would allow a fixed number of iterations to be used, thus avoiding the drawback associated with other dual ascent methods that the number of iterations needed for an \( \epsilon \)-feasible solution cannot be predicted.

A more detailed description and initial results concerning the convergence of this mechanism under standard convexity assumptions on the participants can be found in the recent thesis [Hau13].
Chapter 6

Multi-Period Optimal Power Flow

Section [6.1] represents the main contribution of this chapter, and describes the application of a dual subgradient method to a semidefinite relaxation of the multi-period AC OPF problem. The formulation allows for an arbitrary interconnection of market participants over an AC network, and the dual ascent method has the useful interpretation of a price negotiation that respects information privacy on the part of participants. Two heuristics are used to improve the convergence behaviour of the algorithm, and it is noted that the decomposition has additional scalability advantages over a centralized computation. Section [6.2] proposes a receding horizon market model in which prices are periodically re-negotiated in the light of new nominal forecasts of net load.

6.1 A dual ascent mechanism for solving multi-period OPF

This section describes a market mechanism that minimizes operation costs for power systems on intra-day time-scales, employing a dual ascent method modified using heuristics to improve convergence behaviour. Two modelling assumptions for the transmission network are common in OPF problems:

1. The full AC power flow equations, where lines are lossy, voltage magnitudes can vary between nodes, and reactive as well as real power flows are modelled.

2. A linearization known as a DC approximation, in which the complex-valued
voltage phasors at each bus are assumed to have equal, constant magnitude, and phase angles between pairs of buses are assumed to be small. This corresponds to a lossless AC network where the phase difference between two ends of a line are small enough that the sine approximation \( \sin(\theta_i - \theta_j) \approx \theta_i - \theta_j \) can be used.

This section will focus on multi-period OPF under the first set of assumptions, making use of a recent semidefinite relaxation of the problem \([BWFW08, LL12]\). Section 6.2 will report on a receding-horizon implementation under the linearized model, which is attractive due to its lower computational cost.

### 6.1.1 Problem statement

The scope of the problem considered is as follows:

- The transmission network consists of \(N\) buses connected by transmission lines. A distribution system of some kind exists at buses where real and/or reactive power is injected or extracted, but is not modelled.

- Each line has complex admittance \(y_{lm} \in \mathbb{C}\) with non-negative real and imaginary parts, where \(l\) and \(m\) are the buses between which the line runs. Ground susceptance is neglected.

- The market’s participants include both producers and consumers (both will be referred to generically as “participants”). Participants can combine price-elastic and price-inelastic real and reactive components in their power production or consumption.

- Participant actions are to be optimized over a series of \(T\) time intervals or steps, indexed by \(k = 1, \ldots, T\). The optimization is assumed to take place during step 0, so that any constraints or costs relating to participant actions may be a function of their states at step 0.

- Cost functions are defined for all price-elastic real and reactive production or consumption, as are constraints governing their actions over the time horizon (for example ramp rates, or storage dynamics).

The following sets are defined:

- \(\mathcal{C}\) is the set of all participants connected to the network.
6.1. Dual ascent for multi-period OPF

- $\mathcal{N}$ is the set of buses comprising the network. All participants are connected to a bus.

- $\mathcal{C}_n$ is the set of all participants connected to bus $n \in \mathcal{N}$. $\bigcup_{n \in \mathcal{N}} \mathcal{C}_n = \mathcal{C}$ and $\mathcal{C}_n \cap \mathcal{C}_m = \emptyset, \forall n \neq m$.

- $\mathcal{L}$ is the set of pairs of buses $(l, m)$ between which a transmission line is connected.

An optimal solution to a multi-period OPF problem can also be interpreted as an optimal market outcome for the trade of electricity over the time period $[WMM11]$.

For an AC network, OPF over multiple time steps requires solution of a non-convex optimization of the form:

$$\min_{\{p_i, q_i\}, \{v_k\}} \sum_{i \in \mathcal{C}} J_i(p_i, q_i)$$  \hspace{1cm} (6.1a)

s.t. $(p_i, q_i) \in \mathcal{PQ}_i$ \hspace{1cm} $\forall i \in \mathcal{C}$, \hspace{1cm} (6.1b)

$$\sum_{i \in \mathcal{C}_n} [p_i + \hat{p}_i]_k = \text{Re}\{v_k^* Y_n v_k\} \hspace{1cm} \forall n \in \mathcal{N}, \forall k,$$ \hspace{1cm} (6.1c)

$$\sum_{i \in \mathcal{C}_n} [q_i + \hat{q}_i]_k = \text{Re}\{v_k^* Y_n v_k\} \hspace{1cm} \forall n \in \mathcal{N}, \forall k,$$ \hspace{1cm} (6.1d)

$$\underline{v}_n \leq |[v_k]_n| \leq \overline{v}_n \hspace{1cm} \forall n \in \mathcal{N}, \forall k,$$ \hspace{1cm} (6.1e)

$$|v_k^* Y_{lm} v_k| \leq \overline{S}_{lm} \hspace{1cm} \forall (l, m) \in \mathcal{L}, \forall k,$$ \hspace{1cm} (6.1f)

where $J_i(p_i, q_i) : \mathbb{R}^T \times \mathbb{R}^T \rightarrow \mathbb{R}$ is the cost to participant $i \in \mathcal{C}$ of injecting positive real power $p_i \in \mathbb{R}^T$ and reactive power $q_i \in \mathbb{R}^T$ into the network over $T$ steps. The set $\mathcal{PQ}_i \subset \mathbb{R}^T \times \mathbb{R}^T$ defines the feasible real and reactive power schedules available to participant $i$. For example, for a thermal generator, $J_i(p_i, q_i)$ represents fuel and other variable costs and $\mathcal{PQ}_i$ would be determined by its static operating chart and ramping constraints. Models for $J_i(p_i, q_i)$ and $\mathcal{PQ}_i$ for different types of participant are described in Table 6.1. Fixed (i.e. price-inelastic) real and reactive power injections due to participant $i$ are denoted $\hat{p}_i \in \mathbb{R}^T$ and $\hat{q}_i \in \mathbb{R}^T$. Some participants may be modelled as a summation of fixed and variable powers.

The $k$th element of a vector $p$ is denoted $[p]_k$, so that $\sum_{i \in \mathcal{C}_n} [p_i + \hat{p}_i]_k$ denotes the net real power injection at bus $n$ at time $k$. $v_k \in \mathbb{C}^N$ is the vector of complex-valued bus voltages at time $k$, and $[v_k]_n$ represents the voltage at the $n$th bus at time $k$. $v_k^*$ denotes the conjugate transpose of $v_k$. $\underline{v}_n$ and $\overline{v}_n$ are the minimum and maximum allowed magnitudes of the voltage $[v_k]_n \in \mathbb{C}$ at bus $n$ for all times $k$. The apparent power flow along line $(l, m)$ at time $k$ is denoted $|v_k^* Y_{lm} v_k|$, and may not exceed limit $\overline{S}_{lm}$. 


6. Multi-Period Optimal Power Flow

Note that this notational difference between \( p_i \) and \( q_i \), which are \( T \)-dimensional for each participant \( i \), and \( v_k \), which is \( N \)-dimensional for each time step \( k \), is necessary since constraints in participant actions are time-coupled but network-independent, whereas voltage constraints (and any current constraints implied by them) are network-coupled but instantaneous.

Constants \( Y_n, \overline{Y}_n, Y_{lm} \in \mathbb{C}^{T \times T} \) are admittance matrices derived from the AC power flow laws holding at each bus \( n \) and for each line \((l, m)\). They are defined as in [LL12].

The solution of problem (6.1) is approached by reformulating it in terms of the following real-valued matrix variables \( W_k \), for each step \( k = 1, \ldots, T \) (\( v' \) denotes the transpose of \( v \)):

\[
W_k := \begin{bmatrix} \text{Re}\{v_k\} \\ \text{Im}\{v_k\} \end{bmatrix} \quad \begin{bmatrix} \text{Re}\{v_k\}' \\ \text{Im}\{v_k\}' \end{bmatrix} \in \mathbb{R}^{2N \times 2N} \quad (6.2)
\]

The following optimization is therefore equivalent to (6.1), where \( \langle A, B \rangle \) denotes the trace of the product \( A'B \):

\[
\min_{\{p_i, q_i, W_k\}} \sum_{i \in C} J_i(p_i, q_i) \quad (6.3a)
\]

s. t. \((p_i, q_i) \in \mathcal{P}_i, \forall i \in C\) \quad (6.3b)

\(\langle Y_n, W_k \rangle - \sum_{i \in C_n} [p_i + \hat{p}_i]_k = 0 \quad \forall n \in \mathcal{N}, \forall k,\) \quad (6.3c)

\(\langle \overline{Y}_n, W_k \rangle - \sum_{i \in C_n} [q_i + \hat{q}_i]_k = 0 \quad \forall n \in \mathcal{N}, \forall k,\) \quad (6.3d)

\(\frac{v_n^2}{2} \leq \langle M_n, W_k \rangle \leq v_n^2 \quad \forall n \in \mathcal{N}, \forall k,\) \quad (6.3e)

\(\langle Y_{lm}, W_k \rangle^2 + \langle \overline{Y}_{lm}, W_k \rangle^2 \leq S_{lm}^2 \quad \forall (l, m) \in \mathcal{L}, \forall k,\) \quad (6.3f)

\(W_k \succeq 0\quad \forall k,\) \quad (6.3g)

\(\text{rank}(W_k) = 1\quad \forall k,\) \quad (6.3h)

\(Y_n, \overline{Y}_n, M_n, Y_{lm}, \) and \( \overline{Y}_{lm} \) are real-valued \( 2N \times 2N \) matrices defined in terms the real and imaginary parts of \( Y_n, \overline{Y}_n, \) and \( Y_{lm} \), as in [LL12].

However, the solution \( W_k^{\text{opt}} \) must be such that a nonzero voltage vector \( v_k^{\text{opt}} \) satisfying the definition (6.2) can be obtained. It can easily be shown that this is possible if and only if \( W_k^{\text{opt}} \) is a positive semidefinite, symmetric matrix of rank 1. This requirement means that the reformulated problem still remains very difficult to solve.
Fortunately, as reported in [LL12], such a $W_k^{\text{opt}}$ very often exists when rank constraint (6.3h) is ignored. The approach described in this section will rely on this observation formally, as stated in Assumption 4 below.

### 6.1.2 Market-based solution

The approach considered here is to solve problem (6.3) using a distributed optimization method consistent with market-based price determination (a *tâtonnement* process).

#### Assumptions

In order to construct an efficient market-based solution method for the problem (3), the following standard convexity assumptions are made:

**Assumption 1.** The set $\mathcal{P}_i \subset \mathbb{R}^T \times \mathbb{R}^T$ is compact and convex for each participant $i \in \mathcal{C}$.

**Assumption 2.** The cost function $J_i(p_i, q_j)$ is strictly convex for each participant $i \in \mathcal{C}$.

**Assumption 3.** Problem (6.3) admits a feasible solution, and a Slater point exists.

**Assumption 4.** Solving problem (6.3) with constraint (6.3h) dropped gives a solution for which $\text{rank}(W_k) = 1$ or $2$ for all $k$. Furthermore, where a matrix $W_k$ has rank 2, a $W_k$ of rank 1 can be generated without changing the objective value of the solution or making it infeasible. This assumption is based on the findings in [LL12], which appear to hold for physically realistic networks, and is in practice usually satisfied.

Under Assumptions 3 and 4 the problem can be approached as a convex one while retaining optimality of the solution, strong duality holds, and decomposition strategies relevant to market clearing can be employed.

#### Structure of the partial Lagrangian

Let the following sets comprise all voltage matrices satisfying (6.3e), (6.3f), (6.3g):

$$\mathcal{W}_k := \{W_k : (6.3e), (6.3f), (6.3g) \text{ hold for that } k\}$$

$$\mathcal{W} := \{W = \text{diag}(W_1, \ldots, W_T) : W_k \in \mathcal{W}_k, \forall k\}$$
Note each $\mathcal{W}_k$, and therefore $\mathcal{W}$, is an intersection of convex cone and affine inequality constraints, and is therefore itself convex. The following also holds:

**Lemma 1.** $\mathcal{W}_k$ is compact for all $k$, as is $\mathcal{W}$.

**Proof.** Constraint (6.3e) is equivalent to $\sum_{n \in N} \mathbb{S}^2 \leq |W_k|_{nn} + |W_k|_{(N+n)(N+n)} \leq \mathbb{S}^2$. Therefore $\sum_{n \in N} \mathbb{S}^2 \leq \text{tr}\{W_k\} \leq \sum_{n \in N} \mathbb{S}^2$. Since $W_k \succeq 0$ this implies that the largest eigenvalue of $W_k$ is bounded by some $\theta$, implying $W_k \preceq \theta I$. $W_k$ cannot have an unbounded direction $D \in \mathbb{R}^{2N} \times \mathbb{R}^{2N}$, since for any $W_k \in \mathcal{W}_k$, $0 \preceq W_k + \beta D \preceq \theta I$ is violated for sufficiently large $\beta \in \mathbb{R}_+ \setminus 0$.

Compactness of the block-diagonal concatenation $\mathcal{W}$ follows from a similar argument. \hfill \square

Using these definitions, the following partial Lagrangian of optimization problem (6.3) (with rank constraint (6.3h) dropped) can be obtained by introducing penalties for violating constraints (6.3c) and (6.3d):

$$L(p, q, W, \lambda, \mu) := \sum_{i \in C} J_i(p_i, q_i) \quad \text{(6.4a)}$$

$$+ \sum_{k=1}^T \sum_{n \in N} \lambda_{nk} \left( \langle Y_n, W_k \rangle - \sum_{i \in C_n} [p_i + \hat{p}_i]_k \right) \quad \text{(6.4b)}$$

$$+ \sum_{k=1}^T \sum_{n \in N} \mu_{nk} \left( \langle Y_n, W_k \rangle - \sum_{i \in C_n} [q_i + \hat{q}_i]_k \right) \quad \text{(6.4c)}$$

Multiplier $\lambda_{nk}$ is associated with each nodal real power constraint (6.3c), and $\mu_{nk}$ is associated with each nodal reactive power constraint (6.3d), with $\lambda$ and $\mu$ denoting the concatenation of these variables. Note that $\lambda$ and $\mu$ are unsigned because they correspond to equality constraints. The vectors $p$ and $q$ are similarly defined as shorthand for the concatenation of the vectors $p_i$ and $q_i$ for all $i \in C$, with $(p, q) \in \mathcal{P}_Q$ representing the corresponding constraint set.

Terms (6.4b) and (6.4c) are separable in $W_k$, and since all participants occupy a single node, each of the products $\lambda_{nk}[p_i]_k$ and $\mu_{nk}[q_i]_k$ can be associated with the relevant participant. Therefore the Lagrangian can be rearranged (defining $\lambda_n := [\lambda_{n1} \cdots \lambda_{nT}]'$ and $\mu_n := [\mu_{n1} \cdots \mu_{nT}]'$) to give

$$L(p, q, W, \lambda, \mu) := \sum_{n \in N} \sum_{i \in C_n} (J_i(p_i, q_i) - \lambda_n' p_i - \mu_n' q_i)$$

110
6.1. Dual ascent for multi-period OPF

\[ + \sum_{k=1}^{T} \left\langle \sum_{n \in \mathcal{N}} (\lambda_{nk} Y_n + \mu_{nk} \bar{Y}_n), W_k \right\rangle \]  \hspace{1cm} (6.5)

A corresponding dual of problem (6.3) is then

\[ d(\lambda, \mu) := \min_{(p,q) \in \mathcal{PQ}, W \in \mathcal{W}} L(p, q, W, \lambda, \mu). \]  \hspace{1cm} (6.6)

Note that since both \(\mathcal{PQ}\) and \(\mathcal{W}\) are compact and the Lagrangian (6.5) is continuous, the usual infimum used in the definition of the dual function will always be attained, and can be written in (6.6) simply as a minimum.

**Solution by dual decomposition**

The dual function (6.6) can be maximized using a routine that repeatedly polls participants for their optimal power production or consumption policies given prices \(\lambda\) and \(\mu\), and combines this with a \(W\) that minimizes the Lagrangian. Evaluating the resulting constraint violations gives an element of the dual subgradient (actually the supergradient since a concave function is being maximized, but the same notation as for subgradients will be used here, so that \(g \in \partial d(\lambda, \mu) \iff d(\tilde{\lambda}, \tilde{\mu}) \leq d(\lambda, \mu) + g^T (\tilde{\lambda} - \lambda, \tilde{\mu} - \mu), \forall \tilde{\lambda}, \tilde{\mu} \in \mathbb{R}_T^T\)). Choosing a step length \(\alpha_j\), where \(j\) is the iteration number, new prices \(\lambda\) and \(\mu\) are then chosen and the process repeated. This is the well-known subgradient method [Ber99, Nes04].

Inspection of the Lagrangian (6.5) shows that for fixed dual variables \(\lambda\) and \(\mu\) it is separable between all pairs \((p_i, q_i)\) and linear in each \(W_k\). This leads to the following observation:

**Lemma 2.** If price schedules \(\lambda_n\) for real power and \(\mu_n\) for reactive power are available to participant \(i \in C_n\), then the Lagrangian (6.5) is minimized over \((p_i, q_i) \in \mathcal{P}_i\) when participant \(i\) maximizes its net income or utility considering only those prices given.

**Proof.** The maximization of surplus for participant \(i \in C_n\) is equivalent to finding an optimal pair:

\[ (p_i^*(\lambda_n, \mu_n), q_i^*(\lambda_n, \mu_n)) := \operatorname{argmin}_{(p_i, q_i) \in \mathcal{P}_i} \{ J_i(p_i, q_i) - \lambda'_n p_i - \mu'_n q_i \}. \]  \hspace{1cm} (6.7)
6. Multi-Period Optimal Power Flow

The term \( J_i(p_i, q_i) \) represents the cost of executing the schedules of positive real and reactive power given by vectors \( p_i \) and \( q_i \), while \( -\lambda_n p_i \) and \( -\mu_n q_i \) represent incomes for the net real and reactive power produced. Once the pair \((p_i^*(\lambda_n, \mu_n), q_i^*(\lambda_n, \mu_n))\) is found, the remainder of the Lagrangian \((6.5)\) is independent of \( p_i \) and \( q_i \). For consumers with utility functions, an analogous argument can be used.

Once a minimizer of \((6.7)\) has been computed for each participant \( i \in C \), the remainder of \((6.5)\) can be minimized independently over \( W \in \mathcal{W} \), a minimization that can be split into \( T \) independent (and therefore parallelizable) subproblems to find each \( W_k \):

\[
W_k^*(\lambda, \mu) := \text{argmin}_{W_k \in \mathcal{W}_k} \left\{ \sum_{n \in \mathcal{N}} (\lambda_{nk} Y_n + \mu_{nk} Y_n) , W_k \right\} . \tag{6.8}
\]

Examining the KKT conditions of optimization \((6.8)\) shows that Assumption 4 should typically hold when the prices \( \lambda \) and \( \mu \) are positive. Denoting the multiplier for the constraint \( k\)th constraint of \((6.3g)\) \( Z_k \), one finds that \( Z_k \) should usually have a rank of at least \( 2N - 2 \), and from the complementarity condition \( \langle W_k, Z_k \rangle = 0 \) it follows that \( \text{rank}(W_k) \leq 2 \) \cite{LL12}.

For a given \( \lambda \) and \( \mu \), let \( p^*(\lambda, \mu) \), \( q^*(\lambda, \mu) \), and \( W^*(\lambda, \mu) \) denote values of \( p \), \( q \), and \( W \) which minimize Lagrangian \((6.5)\) subject to \((p, q) \in \mathcal{PQ} \) and \( W \in \mathcal{W} \). Then the following lemma holds:

**Lemma 3.** \( d(\cdot, \cdot) \) is a concave function, and an element of its supergradient \( \partial d(\lambda, \mu) \) is given by \( g := [g_{\lambda_1} \cdots g_{\lambda_T} g_{\mu_1} \cdots g_{\mu_T}]' \), where each \( g_{\lambda_k}, g_{\mu_k} \in \mathbb{R}^N \) with

\[
[g_{\lambda_k}]_n := \langle Y_n, W_k^*(\lambda, \mu) \rangle - \sum_{i \in C_n} [p_i^*(\lambda, \mu) + \hat{p}_i]_k ,
\]

and

\[
[g_{\mu_k}]_n := \langle Y_n, W_k^*(\lambda, \mu) \rangle - \sum_{i \in C_n} [q_i^*(\lambda, \mu) + \hat{q}_i]_k .
\]

**Proof.** \( d(\lambda, \mu) \) is always concave in both \( \lambda \) and \( \mu \) since it is the point-wise infimum of a set of concave functions. Therefore \( g \in \partial d(\lambda, \mu) \Leftrightarrow d(\lambda, \mu) + g' \frac{\xi - \lambda}{\mu - \mu} \), \( \forall \lambda, \mu \in \mathbb{R}^T \). Now

\[
d(\lambda, \mu) \leq L(p^*(\lambda, \mu), q^*(\lambda, \mu), W^*(\lambda, \mu), \tilde{\lambda}, \tilde{\mu}) \text{ by def. (6.6)} \]
6.1. Dual ascent for multi-period OPF

\[
L(p^\star(\lambda, \mu), q^\star(\lambda, \mu), W^\star(\lambda, \mu), \tilde{\lambda}, \tilde{\mu}) - [\lambda' \mu']g + [\lambda' \mu']g
\]

by expansion of (6.5).

The dual function is not smooth everywhere. Because the primal objective (6.3a) is not strictly convex in \(W\), the Lagrangian (6.5) is not in general minimized by a unique \(W\). Consequently the sufficient condition for dual continuous differentiability (Prop. 6.1.1 of [Ber99]) is not met. Quoting Danskin’s theorem (e.g. from [Ber99], Prop. B.25, and substituting \((p,q,W)\) and \(\mathcal{R} \times \mathcal{W}\) for \(z\) and \(Z\), \((\lambda, \mu)\) for \(x\), and \(-L(p,q,W,\lambda,\mu)\) for \(\phi(x,z)\)), one finds that although the the minimizer of term (6.7) of the Lagrangian is always unique for a given \((\lambda, \mu)\), the minimizer of (6.8) is generally not. In those cases, it is always possible to find a direction \(y \in \mathbb{R}^{2NT}\) in which the dual function is not differentiable.

Unfortunately, the lack of smoothness makes maximizing the dual difficult, because only an arbitrary element of \(\partial d(\lambda, \mu)\) can be computed from optimizations (6.7) and (6.8) at each iteration. Cutting plane methods, where the dual function is approximated using hyperplanes constructed from the value and gradient at each iteration, are not usable because the dual’s value is not accessible to the price issuer. Methods employing, for example, an augmented Lagrangian [Ber99], adding penalty terms to the primal objective function in order to smooth the dual, are not suitable either, on the grounds that they either introduce non-monetary quadratic changes to the participants’ optimizations, or prevent separation of the Lagrangian.

The approach considered here is therefore restricted to the subgradient method, but improvements leading to faster convergence are proposed in Section 6.1.3.

An implementation of the subgradient method is formalized in Algorithm 1 and shown in Fig. 6.1. By choosing step length \(\alpha_j\) such that \(\sum_{j=1}^{\infty} \alpha_j = \infty\), \(\lim_{j \to \infty} \alpha_j = 0\) (see [Ber99], Chapter 6), it can be shown that \(\lambda\) and \(\mu\) converge to a dual optimal pair \((\lambda^{\text{opt}}, \mu^{\text{opt}})\). A common choice is \(\alpha_j = K/\sqrt{j}\).

Theorem 1. Executing Algorithm 1 with

\[
\begin{bmatrix}
  f_\lambda^j(g_\lambda, g_\mu) \\
  f_\mu^j(g_\lambda, g_\mu)
\end{bmatrix} := \frac{K}{\sqrt{j}} \begin{bmatrix}
  g_\lambda \\
  g_\mu
\end{bmatrix} / \left\| g_\lambda \right\| \left\| g_\mu \right\|
\]

\tag{6.9}

with \(K > 0\) solves problem (6.3) to within a specified maximum constraint violation \(\epsilon\) in a finite number of iterations, in that, given a terminating pair of vectors...
6. Multi-Period Optimal Power Flow

\[ \lambda^{\text{opt}} \text{ and } \mu^{\text{opt}}, \text{ minimizing } L(p, q, W, \lambda^{\text{opt}}, \mu^{\text{opt}}) \text{ over } \mathcal{R} \times \mathcal{W} \text{ gives an optimal solution } (p^{\text{opt}}, q^{\text{opt}}, W^{\text{opt}}) \text{ to the original problem (6.3)}. \]

Proof. Under Assumptions 3 and 4 strong duality holds, so any primal minimizers of \( L(p, q, W, \lambda^{\text{opt}}, \mu^{\text{opt}}) \) over \( \mathcal{R} \times \mathcal{W} \) give an optimal solution to (6.3) including constraint (6.3h), once matrices \( W_k \) of rank 1 have been generated from the output. Proof of convergence of the subgradient algorithm given in Algorithm 1 with step size rule (6.9) follows the line presented in \S 3.2 of [Nes04]. \qed

In fact, in numerical tests it was observed that the matrices \( W_k \) satisfy the rank condition even at intermediate iterations. This implies physically-meaningful voltage vectors can be extracted before convergence has been reached, however these will still be inconsistent with the nodal power injections proposed by the participants connected to the network.

6.1.3 Modified algorithms for improved performance

Although Algorithm 1 converges regardless of starting point and initial step size, there is no guarantee that convergence to an optimal solution of problem (6.3) will be achieved in an acceptable number of iterations. Numerical experiments show that a naive implementation of the basic subgradient method in Algorithm 1 indeed offers unacceptably slow convergence. Two modifications are presented here and summarized in Algorithms 2 and 3 by virtue of which significant improvement was observed in the numerical example to be described in Section 6.1.4.
Algorithm 1 Negotiated solution to the AC OPF problem.

1: \( \lambda \leftarrow \lambda^0 \)
2: \( \mu \leftarrow \mu^0 \)
3: \( X \leftarrow \text{FALSE} \) \{Set convergence flag to FALSE.\}
4: \( j \leftarrow 1 \) \{Initialize iteration index.\}
5: \( \text{while } X == \text{FALSE} \) do
6: \( \text{for each } n \in \mathcal{N} \) do
7: \( \text{for each } i \in C_n \) do
8: \( (p_i, q_i) \leftarrow (p_i^\ast(\lambda_n, \mu_n), q_i^\ast(\lambda_n, \mu_n)) \) \{Solve (6.7).\}
9: \( \text{end for} \)
10: \( \text{end for} \)
11: \( \text{for } k := 1 \text{ to } T \) do
12: \( W_k \leftarrow W_k^\ast(\lambda, \mu) \) \{Solve (6.8).\}
13: \( \text{for each } n \in \mathcal{N} \) do
14: \( [g_{\lambda_k}]_n \leftarrow \langle Y_n, W_k \rangle - \sum_{i \in C_n} [p_i + \hat{p}_i]_k \)
15: \( [g_{\mu_k}]_n \leftarrow \langle Y_n, W_k \rangle - \sum_{i \in C_n} [q_i + \hat{q}_i]_k \)
16: \( \text{end for} \)
17: \( \text{end for} \)
18: \( \text{if } ||g_{\lambda_k}||_\infty \leq \epsilon \text{ and } ||g_{\mu_k}||_\infty \leq \epsilon, \forall k, \text{ then} \)
19: \( \{ \text{Constraints satisfied.} \} \)
20: \( X \leftarrow \text{TRUE} \) \{Terminate and apply prices } \lambda \text{ and } \mu.\}
21: \( \text{else} \)
22: \( \{ \text{Constraints not yet satisfied.} \} \)
23: \( \lambda \leftarrow \lambda + f_\lambda(g_{\lambda}, g_{\mu}) \)
24: \( \mu \leftarrow \mu + f_\mu(g_{\lambda}, g_{\mu}) \)
25: \( j \leftarrow j + 1 \)
26: \( \text{end if} \)
27: \( \text{end while} \)

Constraint aggregation

The subgradient method enforces power conservation separately at every node, but in addition it is obvious that the same constraint must apply to the whole network itself. Summing constraints (6.3c) and (6.3d) over nodes \( n \) gives the following:

\[
\langle \sum_{n \in \mathcal{N}} Y_n, W_k \rangle - \sum_{i \in C} [p_i + \hat{p}_i]_k = 0, \forall k \quad (6.10a)
\]
Algorithm 2 Improved-initialization OPF mechanism.

1: Execute Algorithm 1 with $\epsilon = \epsilon^{agk}$ and update rule (6.11).
2: Reinitialize Algorithm 1 with $\lambda_k \leftarrow 1_n \Lambda_k$ and $\mu_k \leftarrow 1_n M_k$.
3: Execute Algorithm 1 again with the new initialization point and the original update rule given by equation (6.9), and run until an acceptable tolerance is reached in line 18.

\[
\left\langle \sum_{n \in N} Y_n W_k \right\rangle - \sum_{i \in C} [q_i + \hat{q}_i]_k = 0, \forall k \tag{6.10b}
\]

It can be inferred that when $W_k$ is feasible, $\langle \sum_{n \in N} Y_n W_k \rangle$ and $\langle \sum_{n \in N} Y_n W_k \rangle$ represent respectively the total real and reactive power losses in the lines during step $k$.

If in problem (6.3) constraints (6.3c) and (6.3d) are replaced with (6.10a) and (6.10b), the solution no longer has to be consistent with nodal real and reactive power conservation. Consequently, the minimum-cost solution with the aggregated constraints will typically set all voltages to equal values and no real or reactive power will be absorbed in lines.

Let $\Lambda_k, M_k \in \mathbb{R}$ be the multipliers associated with (6.10a) and (6.10b) respectively for each step $k$. Inspection of the Lagrangian of the new problem shows that nodal real and reactive power prices will become equal to $\Lambda_k$ and $M_k$ for all buses. Numerical tests showed that by running Algorithm 1 but ignoring constraints (6.3c) and (6.3d) and penalizing violation of (6.10a) and (6.10b) using the price update rule

\[
\left[ \begin{array}{c} f^A_j(g_A, g_M) \\ f^M_j(g_A, g_M) \end{array} \right] := \frac{K}{\sqrt{j}} \left[ \begin{array}{c} g_A \\ g_M \end{array} \right] / \left\| \begin{array}{c} g_A \\ g_M \end{array} \right\|, \tag{6.11}
\]

where $g_A$ and $g_M$ are defined analogously to $g_\lambda$ and $g_\mu$, a set of prices near to the optimal global prices can be obtained very quickly. Algorithm 1 can then be reinitialized with $\lambda_k \leftarrow 1_n \Lambda_k$ and $\mu_k \leftarrow 1_n M_k$, so that physically consistent bus voltages can then be found; see Algorithm 2.

Theorem 2. Algorithm 2 also terminates at a dual-optimal solution.
6.1. Dual ascent for multi-period OPF

Proof. By the same argument as in Theorem 1, the improved method finds an optimal solution to problem (6.3) with constraints (6.3c) and (6.3d) replaced by (6.10a) and (6.10b) in a finite number of steps for a given tolerance $\epsilon_{agg}$. Since this is used as the starting point for Algorithm 1, for which any starting point results in convergence to the dual-optimal solution, such convergence is guaranteed.

The aggregation method is in effect a dual ascent method applied to a different dual problem, with lower dimension and with lower optimal value than problem (6.3). The observed effect of reducing the dual dimension is to reduce the chance that the subgradient ascent direction is near-perpendicular to the desired change in prices. This phenomenon, discussed in detail in [SKR85], is responsible for most of the poor performance seen in subgradient methods.

Choosing feasible voltages

The method described above relies on the standard subgradient method to eventually give a globally optimal solution to problem (6.3). However, optimal values $W_k^*(\lambda, \mu)$ from (6.8) were observed to be very sensitive to the values of $\lambda$ and $\mu$, due to that part of the Lagrangian’s linear objective function (see the discussion of this effect in Section 5.1.5). Because of this, satisfying constraints (6.3c) and (6.3d) using the standard subgradient method requires many iterations (with many fine changes to $\lambda$ and $\mu$) to obtain usable voltages, in the course of which the participants’ operating points $(p_i, q_i)$ and prices $(\lambda, \mu)$ barely change.

In reality, network voltages are a consequence of the real and reactive power injections at each bus of the network, and not chosen according to an optimization. Thus, as long as bus voltages can be found such that consistency constraints (6.3c) and (6.3d) hold, the solution is implementable as long as the voltages found reflect a network state that can be maintained by the automatic controllers already present. This is likely to be the case since OPF solutions are almost always stable ones featuring low network losses, and this heuristic only makes small adjustments from a near-optimal starting point.

In other words, as long as matrices $W_k$ can be found satisfying (6.3c) and (6.3d) for each step $k$, with rank at most 2, the proposed power injections are feasible. If this is the case, for each time step $k$ the following convex optimization, with $p_i^*(\lambda_n, \mu_n)$ and $q_i^*(\lambda_n, \mu_n)$ as fixed input data, has optimal value $\delta = 0$ and the output matrices have rank at most 2:
Algorithm 3 Feasibility-prioritized OPF solution.

1: Execute Algorithm 1 with $\epsilon = \epsilon^{agg}$ and update rule 6.11.
2: Reinitialize Algorithm 1 with $\lambda_k \leftarrow 1_n \Lambda_k$ and $\mu_k \leftarrow 1_n M_k$.
3: Execute with $\epsilon = \epsilon^{phys}$, and with line 12 replaced with $W_k \leftarrow W_k^{phys}(\lambda, \mu)$.
   Augment termination condition in line 18 with rank($W_k^{phys}(\lambda, \mu)$) $\leq 2$, $\forall k$.

$$
\min_{W, \delta} W_k, \delta \\
\text{s.t. } |\langle Y_n, W_k \rangle - \sum_{i \in C_n} [p_i^*(\lambda_n, \mu_n) + \hat{p}_i]_k| \leq \delta \forall n ,
$$

(6.12a)

$$
|\langle Y_n, W_k \rangle - \sum_{i \in C_n} [q_i^*(\lambda_n, \mu_n) + \hat{q}_i]_k| \leq \delta \forall n ,
$$

(6.12b)

$$
W_k \in W_k .
$$

(6.12d)

The second modification, described by Algorithm 3, aims to find a consistent set of bus voltages as soon as the prices determined using the aggregated constraints (6.10) have been reached. This is done by using the output matrices from problem (6.12), denoted $W_k^{phys}(\lambda, \mu)$, in place of the true Lagrangian minimizer $W^*_k(\lambda, \mu)$, on line 12 of Algorithm 1. The algorithm terminates when a tolerance of $\epsilon^{phys}$ is achieved on line 18.

Although optimization problem (6.12) is no longer a minimization of the Lagrangian (6.5) over $W$, and therefore cannot be guaranteed to lead to an appropriate dual subgradient or a dual-optimal solution, it consistently produced matrices of rank 2 in various tests, which could then be used to derive feasible voltage vectors $v_k$.

### 6.1.4 Results

The pricing mechanism will now be illustrated using the adapted 39-bus network from [Pai89] shown in Fig. 6.2. Parameters for the participants and network are given in Table 6.1.

The network has 46 transmission lines, 7 generators, 2 storage units, 3 wind farms, and 19 inelastic loads. The multi-period optimization was carried out over a time horizon of $T = 10$ steps. Prices were initialized to $\lambda_0 = 250$ and $\mu_0 = 0$ for all
time steps and nodes. Loads and wind farm outputs $\hat{p}$ and $\hat{q}$ for the 10 steps were generated by multiplying $\hat{p}_{\text{nom}}$ and $\hat{q}_{\text{nom}}$ from Table 6.1 with the trajectories given.

A nominal voltage of $v_{\text{nom}} = 230$ kV was used, with $v_n = 0.95v_{\text{nom}}$, $v_n = 1.05v_{\text{nom}}$, $\forall n$. The per-unit line impedance values from the specification in [Pai89] were used with a base of 100 $\Omega$, leading to typical impedances of order $0.2 + 6i$ $\Omega$. To illustrate the incorporation of congestion constraints, an apparent power limit of 200 MVA was applied to line 15-16, i.e. $S_{15,16} = S_{16,15} = 200$; this constraint was active in the 15-16 direction over all times $k$ in the solutions obtained. Generator (i) was constrained to have a low maximum ramp rate (i.e. change in power output between steps); its output is denoted by the dark blue lines in Fig. 6.6.

The three algorithms detailed above were compared, and the results are summarized in Table 6.2. In each case step length parameter $K = 8$ was used, with tolerance $\epsilon^{\text{agg}} = 5$ in Algorithms 2 and 3. With a naive implementation of Algorithm 1 convergence took so long (thousands of iterations, equivalent to tens of hours of computation) that it was not possible to terminate the method at an acceptable solution. In fact Algorithm 2 was used to measure the optimal cost in this case. This approach is valid because the oscillating signs of the dual subgradient elements $g_\lambda$ and $g_\mu$ suggest that the dual optimal solution had been reached to
Table 6.1: Parameters of the participants shown in Fig. 6.2

<table>
<thead>
<tr>
<th>Type</th>
<th>Label</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generators</td>
<td>i</td>
<td>$b = 25$, $c = 0.800$, $\Delta p = 10$, $\bar{p} = 500$</td>
</tr>
<tr>
<td></td>
<td>ii</td>
<td>$b = 100$, $c = 0.200$, $\Delta p = 250$, $\bar{p} = 2000$</td>
</tr>
<tr>
<td></td>
<td>iii</td>
<td>$b = 65$, $c = 0.154$, $\Delta p = 250$, $\bar{p} = 1300$</td>
</tr>
<tr>
<td></td>
<td>iv</td>
<td>$b = 56$, $c = 0.179$, $\Delta p = 250$, $\bar{p} = 1120$</td>
</tr>
<tr>
<td></td>
<td>v</td>
<td>$b = 54$, $c = 0.185$, $\Delta p = 250$, $\bar{p} = 1080$</td>
</tr>
<tr>
<td></td>
<td>vi</td>
<td>$b = 83$, $c = 0.121$, $\Delta p = 250$, $\bar{p} = 1660$</td>
</tr>
<tr>
<td></td>
<td>vii</td>
<td>$b = 100$, $c = 0.100$, $\Delta p = 250$, $\bar{p} = 2000$</td>
</tr>
</tbody>
</table>

Table 6.2: Convergence time and optimality results for Algorithms 1, 2, and 3

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Number of iterations</th>
<th>Max constraint violation</th>
<th>Cost</th>
<th>Supoptimality</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Thousands</td>
<td>$\nabla$ 0</td>
<td>$7.0292 \times 10^6$</td>
<td>—</td>
</tr>
<tr>
<td>2</td>
<td>600</td>
<td>5</td>
<td>$7.0292 \times 10^6$</td>
<td>+0 %</td>
</tr>
<tr>
<td>3</td>
<td>144</td>
<td>$10^{-3}$</td>
<td>$7.1200 \times 10^6$</td>
<td>+1.29 %</td>
</tr>
</tbody>
</table>

within 4 significant figures after around 600 iterations. A constraint violation of 5 MW was still present; the sensitivity of the minimization step to the multipliers prevented the network constraints from being satisfied to within a finer tolerance.

Algorithm 3 was run with $e^{\text{phys}} = 10^{-3}$, convergence took 144 iterations (94 with aggregated constraints plus 50 to find feasible line flows). The modified method resulted in a cost increase of 1.29% over the global optimum in this case.
6.1. Dual ascent for multi-period OPF

The convergence behaviour of Algorithm 3 is illustrated in Figs. 6.3 and 6.4, and the final price and power schedules are shown in Figs. 6.5 and 6.6 respectively. Complex bus voltages $v_k$ extracted from the solution at termination are shown for $k = 1$ in Fig. 6.7.

Line 12 of Algorithm 1 and minimization (6.12), involve finding an optimal $2N \times 2N$ matrix $W_k^*(\lambda, \mu) \in \mathcal{W}_k$ for each step $k$. For symmetric $W_k$, this means $2N(N+1)$ primal variables must be found. For $N = 39$ buses this required around 1 minute of CPU time. However, the dual of (6.8) is in fact dramatically faster to solve (of the order 1 second), perhaps due to the dimension of the dual problem. This was achieved using the dualize command in YALMIP [Löf04], which uses the user-defined constraints to form the dual of (6.8). The multipliers

![Figure 6.3: Convergence of real prices $\lambda_k$ and reactive prices $\mu_k$ to their final values over 144 iterations. Only the plots for $k = 1$ are shown. The first part of Algorithm 3 terminates at iteration 94. Although prices for $k = 1$ converge quickly, others take longer, hence the apparent inactivity between iterations 20 and 94.](image)
Figure 6.4: Residuals (summed across nodes) for equality constraints (6.3c) (blue) and (6.3d) (green) under Algorithm 3. Until iteration 94, the residuals of the aggregated constraints (6.10a) and (6.10b) are decreasing, while those shown remain largely constant until disaggregation is triggered.
6.1. Dual ascent for multi-period OPF

Figure 6.5: Nodal real power prices over the 10 steps.

Figure 6.6: Final real and reactive power schedules for generators (solid lines) and storage units (dotted lines). Note that storage units were assumed not to produce or consume VARs and so do not appear on the second graph.

from the solution to this dual, output by the solver (SDPT3 v4.0 [TTT03]), include the required $W_k^\star(\lambda, \mu)$ or $W_k^{\text{phys}}(\lambda, \mu)$.
6. Multi-Period Optimal Power Flow

![Figure 6.7: Bus voltages $v_k$ for $k = 1$ under Algorithm 3. The voltages were extracted from $W_{1}^{\text{opt}}$ (this involved using the procedure from [LL12] to produce a matrix of rank 1). The dotted line denotes the upper voltage magnitude constraint, which is active at one of the nodes.](image)

6.1.5 Further improvements

The results show that when the dual subgradient method is applied to this problem, modifications to the naive method are required in order to obtain an acceptable number of iterations. In addition, some tuning of the parameters such as tolerances and step sizes is beneficial.

Additional methods such as the the $r$-method described in the survey (Section 1.3.3) were also implemented, but these did not result in consistent improvements to convergence rate.

The results of the simulated example exhibited prices that were relatively close to each other, with small differences caused by network losses and line congestion. It is possible that the heuristics used to improve convergence of the algorithm would produce rather more inefficient outcomes if the optimal solution featured larger differences in nodal prices across the network. This could happen if transmission line congestion constraints were to bind strongly. In this case, Algorithm 3 could be modified so that the constraints are disaggregated in stages. This would have the effect of producing a coarse nodal price distribution before the search for a feasible voltage vector is prioritized in the final step.
A further option would be to reduce the dimension of the dual problem (6.6) by moving those constraints (6.3c) and (6.3d) corresponding to nodes with no participating devices attached into the definition of set $\mathcal{V}$, so that power matching constraints are automatically satisfied at those nodes. This may improve convergence of the subgradient method, but at the cost of the algorithm not returning prices for those nodes.

### 6.1.6 Conclusions

This section demonstrated a mechanism for the solution of a general AC optimal power flow problem with mixed participant types, in a market-based environment. The mechanism generates optimal nodal prices for both real and reactive power over the time horizon considered. The structure of the reformulated OPF problem is shown to be well-suited to such a mechanism in that its Lagrangian is separable between the operational (voltage) side, for which the TSO is responsible, and the participant side, where private costs informing operational decisions can be kept private. In order to improve convergence behaviour, constraints were aggregated and then physically-realizable voltages were chosen in order to accelerate the convergence of the algorithm greatly. In this case study a sub-optimality of 1.29% was induced by these modifications.

From a computational point of view, the dual subgradient method presented has the advantage over a centralized AC dispatch that the original large non-linear cone program (6.3) need not be solved in one piece. Indeed, this proved impossible for the case study described, because even when the parser was able to build the centralized semidefinite problem, the solver was unable to solve it, possibly because the resulting problem had been built with an inappropriate sparsity structure. With or without the market interpretation, then, dual decomposition appears to be a particularly attractive approach to the multi-period AC OPF problem, given the mixture of temporal and spatial coupling present. Further decomposition along the graph-clique lines presented in [And11] and [Mol13] may be attractive.

The method is sensitive to the choice of step size and other tunable parameters. Although in the example shown relatively few iterations were needed after the application of the two heuristics described (at least by the usual standards of dual ascent methods), it is likely that such heuristics would have to be tailored further to a given scenario before a low-iteration-count solution could be obtained reliably.
The high computational cost of optimizing over the semidefinite constraint set $\mathcal{W}$ makes this process difficult until further speed increases can be identified.
6.2 Receding horizon (rolling window) electricity market

Section 6.1 described a negotiation mechanism that determines efficient prices and an accompanying OPF solution for participants whose costs and/or constraints are coupled across a time horizon. The idea can be used as a basis for the efficient operation of power systems under fluctuations from wind and other uncertainty sources, by performing the negotiation at regular trading intervals, in order to modify previously-agreed dispatch plans, as shown in Fig. 6.8. The method was proposed in [WMM11], the key results from which are reported here.

At each trading period, typically lasting 15 minutes to 1 hour, the actions of all participants over the following $T$ periods are negotiated given the information available at that time. However, only the first step of the cleared dispatch is implemented, after which the procedure is repeated. The subsequent steps of the original horizon that are not implemented nevertheless serve as useful price indications, which are vital to the efficient operation of devices with time-coupled costs and constraints. These affect the actions carried out at the first step of each planning horizon, coloured green in the figure.

![Figure 6.8: Receding horizon market. The passage of time is drawn vertically, and the time periods negotiated at a particular time are drawn horizontally.](image)

The receding-horizon market idea can be seen as a logical extension of recent developments in electricity markets. In European countries, an increasing number of intra-day markets have been introduced to cope with the fact that renewable in-feeds predicted for the purpose of clearing the day-ahead market turn out to be inaccurate by the time of delivery [Web10, ENT12]. In some US markets the ISO selectively utilizes bids submitted the day before to adjust generator actions in real
6. Multi-Period Optimal Power Flow

Figure 6.9: Above: Schematic illustration of PJM-like mechanism. Below: Proposed modification, in which re-dispatch can be performed on a receding horizon basis [WMM11].

time in response to system measurements [Ott10], but these adjustments are only just starting to account explicitly for time coupling across multiple time periods [XT12]. This coupling is especially important for managing energy storage units and generator ramping actions, both of which impact significantly on the ability of the grid to accommodate large intermittent renewable in-feeds. Fig. 6.9 illustrates a modification of PJM-like clearing mechanisms to account for time coupling.

6.2.1 Comparison with Economic MPC

The proposal shown in Fig. 6.8 is reminiscent of Model Predictive Control [Mac02], in that it relies on model-based short-term plans that are modified in the face of new information whenever such information is measured. Several proposals for so-called Economic MPC have been published in recent years [RA09]. The key characteristic of these is to replace the traditional quadratic state and input penalty function used in conventional MPC with an economic cost function, i.e., an approximation of the real costs of operating the system over a given time horizon [DAR11]. This may be useful in cases where the steady-state reference from which
deviations are traditionally penalized may not be available, or where the costs of the transient leading to a known steady state may be poorly modelled by a traditional quadratic penalty. The concept could be described as a hybrid between MPC and the generic notion of planning actions for a limited time horizon; it has found application in areas such as building climate control [HMPJJ12] and the management of refrigeration under uncertainty [HLJ11].

The application of most relevance to this thesis is predictive dispatch in energy networks, of which [XI09] is a representative example. In this approach, a model of system costs is used to perform a centralized minimization of short-term running costs in the presence of uncertainty. The limited-horizon plan is updated at every step. In contrast, the implementation proposed in [WMM11] is a distributed one in which participants carry responsibility for maximizing profit in response to prices proposed by the market operator, in the manner of a tâtonnement auction. A generic representation of the costs and constraints of each participant was used to guarantee convergence of a price negotiation algorithm, based on the same dual ascent principles used in Section 6.1.

### 6.2.2 Negotiated predictive dispatch

An illustrative example using the DC flow approximation was used in [WMM11] to show that cost savings could be made by negotiating the actions of generators and storage units (and determining efficient prices) over a multi-step time horizon rather than by determining actions one time step at a time. Under the DC approximation, the AC OPF problem (6.1) reduces to the following, in which nodal voltages $v_k$ and reactive powers $q_i$ have been eliminated, and the remaining notation is analogously defined:

\[
\begin{align*}
\min_{\{p_i\}} \sum_{i \in C} J_i(p_i) & \quad (6.13a) \\
\text{s. t.} \quad p_i & \in P_i \quad \forall i \in C, \quad (6.13b) \\
\sum_{i \in C} [p_i + \hat{p}_i]_k & = 0 \quad \forall k; \quad (6.13c) \\
\sum_{i \in C} \Gamma_i[p_i + \hat{p}_i]_k & \leq \bar{P} \quad \forall k. \quad (6.13d)
\end{align*}
\]
The matrices $\Gamma_i$ describe the contribution of power outputs of participants to line flows, and are derived in the Appendix. The vector $\mathbf{T} \in \mathbb{R}^{2LT}$ contains the line flow limits, in both directions and at each time. The multipliers $\mu \in \mathbb{R}^T$ and $\nu \in \mathbb{R}^{2LT}$ can be associated with constraints (6.13c) and (6.13d) respectively. In this case, following the argument of Section 5.1.4, the optimal nodal prices for problem (6.13) can be written in the form $\mu + \Gamma_i \nu$, or a base price plus congestion component, for the node to which participant $i$ is attached. This contrasts with the full AC case, in which network losses prevent a base price from being defined in the same way.

A negotiated solution to problem (6.13) along the lines of Algorithm 1 in the previous section was derived in \cite{WMM11}, and applied to a case study in which the market was settled in a receding horizon manner. The case study was a variation of the 39 bus example described in Section 6.1, with the addition of ramping costs on generators. The results are summarized in Table 6.3. The four methods compared are as follows:

1) **Global, fully prescient optimization**: Solves problem (6.13) centrally for all participants, with perfect knowledge of the coming wind profile, over the entire simulation period. Gives a lower bound on the cost of the dispatch achievable under any other candidate scheme.

2) **Global, receding horizon**: Performed on a receding horizon basis using the nominal wind prediction that would be available at the start of each period. Gives a lower bound on the cost obtainable by method 3).

3) **Negotiated predictive dispatch**: Prices are negotiated in a manner similar to Algorithm 1. In this particular scenario, a few hundred iterations at each step are required. These costs are slightly higher than those of method 2) because it is assumed that market participants would plan their actions with respect to prices extrapolated beyond the horizon end. Consequently a slightly different dispatch is

<table>
<thead>
<tr>
<th>Method</th>
<th>Cost of dispatch</th>
<th>Cost vs. method 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Global prescient</td>
<td>$1.0632 \times 10^8$</td>
<td>-</td>
</tr>
<tr>
<td>2) Global receding horizon</td>
<td>$1.0822 \times 10^8$</td>
<td>+1.7%</td>
</tr>
<tr>
<td>3) Negotiated dispatch</td>
<td>$1.0853 \times 10^8$</td>
<td>+2.1%</td>
</tr>
<tr>
<td>4) Isolated periods</td>
<td>$1.1185 \times 10^8$</td>
<td>+5.2%</td>
</tr>
</tbody>
</table>
obtained compared to the centralized solution.

4) Global, isolated time steps: Performed globally neglecting interstage generator ramping costs, which are added to the result retrospectively. This method can be interpreted as an approximation to the outcome of an idealized version of the PJM real-time control mechanism shown in Fig. 6.9, in which bids are utilized every period by the market operator based on their social welfare contribution, in response to a fluctuating net load. If bids reflect costs, and demand is assumed to be totally inelastic, then standard theory shows that the transmission-constrained bid selection problem also solves problem (6.13) applied to an isolated time step, since the sum of supply and demand surpluses will be maximized.

The results suggest that in cases where ramping costs are significant, receding-horizon pricing may improve efficiency by supplying advance pricing signals to market participants. Method 2) gives the best result that any receding horizon process could achieve, but a method appropriate to a market environment such as 3), which requires no private information and allows some freedom to participants, may be desirable in practice.

6.2.3 Conclusion

As reported in [XT12], the industry has now started to move in the direction of explicitly considering time-coupled costs and constraints in real-time dispatch planning, at least in markets where this function is carried out via a centralized ISO optimization. This has efficiency attractions but some market operators may find that it places excessive requirements on participants, who are forced to reveal more operational information, and possibly commit to more restrictive behavior than would otherwise be necessary. This is one of several arguments against bringing the market back to what increasingly appears to be a centrally-planned solution (see Section 5.1.4 and [WV99], which pointed out similar dangers for electricity markets). An iterative, or negotiated, model may offer a suitable compromise between restrictive central plans and the potential inefficiencies arising from current mechanisms.
Part IV

Research Outlook
Chapter 7

Short- and Medium-Term Objectives

This dissertation presented a number of contributions in the area of robust and distributed optimization for power systems, upon which several further studies could be built. This chapter suggests possible short- and medium-term future developments of the ideas presented in Parts II and III, and Chapter VIII will identify two broader areas of investigation relevant to this dissertation that may gain importance on a longer time-scale.

Section 7.1 suggests several ways that the affine reserve policy idea from Part II could be improved, beyond the more immediate extensions proposed there. Section 7.2 suggests ways that ISO decision-making could be improved using the results of this dissertation. Section 7.4 suggests ways the results of Part III may be used as a starting point for a plausible real-time pricing mechanism. Section 7.5 describes areas in which the semidefinite relaxation of the AC OPF problem could be investigated further.

7.1 Affine reserve policies

Numerical examples suggest that the robust optimization methods outlined in this dissertation represent a promising approach for operating reserves less conservatively, particularly in grids with high-penetration renewables.
A case study on a national or continental level using aggregated models of generation and energy storage capabilities could be used to quantify the effects of robust reserve policies on reserve costs more realistically than for the test systems in this dissertation. As reported in Section 4.1, the effects of affine reserve policies appear to be particularly attractive when large amounts of storage or time-coupled generation costs are present, so the results of a study based on a grid with these characteristics may prove interesting.

Efforts could be made to solve the coupled reserve policy and unit commitment problem for large-scale systems. These could draw on the wide range of heuristic methods that have been used to solve related problems for the last decade or two. Solving such large problems would be a necessary for a practical implementation of the more advanced reserve scheduling methods proposed in Part II. Another requirement for practical use is the incorporation of contingencies such as generator or line failure in the problem; these have not yet been considered, and ways of scaling the problem to deal with large numbers of extra constraints without introducing an unreasonable computational burden are needed. Solving for a reliable and efficient dynamically-changing unit commitment in the face of large-scale uncertainty is one of the most difficult and pressing problems in the field.

A key assumption behind the reserve policy optimization is that the uncertainty can be described in a low-order fashion through a vector of size $N_\delta$. There is great flexibility in the way this could be constructed, a framework for which was sketched in Fig. 3.4. At times of low network congestion, the only uncertain parameter of relevance might be the sum of all power in-feeds, i.e., the system’s net load. At others, it may be necessary or attractive to divide up the uncertainty representation on spatial grounds, as illustrated in the case study. The translation of uncertainty samples from a distribution with potentially large support into a compact uncertainty set also requires attention, and the treatment of this uncertainty depends on what kind of probabilistic system security guarantee the ISO wishes to provide.

A further extension is the possibility of piecewise affine responses to uncertainty, as a generalization of the double-sided response sketched in Section 4.3 and employing ideas from [CZ09] and [GWK10]. For example, a generator could commit to providing upward regulation only up to a given system-wide power imbalance, above which it may flatten out its response. Choosing such piecewise responses would allow risks to be shared with more granularity across the system, and make use of further statistical information concerning the uncertainty. To solve this prob-
lem, the description of the uncertainty set needs to be lifted to a higher dimension (including a good approximation of the convex hull of possible realizations) in order to formulate the robust constraints. The computational cost of this is potentially very high.

It would be computationally difficult, but still potentially very attractive, to be able to determine optimal multi-stage recourse rules for the timing of generator switching, i.e., an adjustable unit commitment. This would allow reduced conservatism compared to the approach described in this dissertation, which couples an open-loop UC with adjustable operating point decisions. The recent two-stage min-max model of Bertsimas et al. [BLS+13] is a starting point for this, and the idea of affine disturbance feedback policies for the control of hybrid systems has been demonstrated for the very simple example of a buck converter [VSWM13]. However despite the emergence of new work in this direction [BG13] there are still no scalable options for multi-stage recourse on binary decisions.

Lastly, almost nothing is known about adjustable robust solutions to AC optimal power flow. Mathematically it is far easier to build robustness into a DC-approximated solution, as described in this dissertation, than to consider the full AC equations. However it also apparent that real power systems exhibit significant robustness to uncertainty, thanks to the presence of automatic controllers that adjust generator outputs in real time. The question therefore arises whether the limits of this mechanism can be quantified, and whether it can be accounted for in an optimization framework.

### 7.2 Online decision-making tools

There is scope for improving the way large-scale optimization is employed in online decision-making tools for the benefit of system operators.

In contrast to local minima found by current non-linear interior point solvers, solutions of even complex, time-coupled AC OPF problems based on a semidefinite relaxation can now be certified as globally optimal. It would be possible to use certifiable solutions to AC power flow problems as an alternative basis for tools for operating transmission and distribution networks. A tool built on sufficiently generic optimization principles would be able to answer a range of complex operational questions posed by TSOs and DSOs. For example, distribution operators
may wish to coordinate the charging of electric vehicles while respecting local voltage constraints, while transmission operators may wish to operate HVDC lines efficiently alongside the conventional transmission network. An additional use would be the investigation of price volatility by market operators and regulators.

Another adaptation of the robustness approach presented in Part II could be to produce tools that either provide preventive and corrective security within a SC-OPF solution \[\text{SA12}\] less conservatively, or measure the security of a given grid operating point more effectively. The incorporation of recourse rules may lead to more elegant ways of factoring in a required corrective security level than by the iterative evaluation of individual scenarios, since planned responses to unforeseen events are inherent to the approach explored in Part II. In order to do this, it would be necessary to combine the effects of a large number of contingency scenarios into a carefully-chosen equivalent robustness set.

### 7.3 ADMM for large-scale combinatorial optimization problems relevant to power systems

It would particularly attractive to obtain usable bounds on the sub-optimality and effects of rounding in large-scale problems involving binary variables, of which many arise in power systems. Initial evidence (see Section \[\text{4.2.1}\]) suggests that such problems can be solved rather cheaply in computational terms, although sub-optimally, by an iterative algorithm involving an appropriate duplication and rounding of variables. Although older results exist on the duality gap in mixed-integer programming problems \[\text{BS82}\], it would be attractive to adapt such bounds for results gained via ADMM, which shows promise as a powerful tool for large-scale mixed-integer problems \[\text{BPC+10}\].

### 7.4 Pricing mechanisms

The discovery of efficient electricity prices becomes increasingly difficult in the presence of large fluctuations from renewable energy sources. Distributed optimization methods such as Lagrangian Relaxation and ADMM have useful interpretations as market clearing mechanisms, however they also have unfortunate requirements
such as an unreasonable numbers of iterations, or the need for non-linear “price”
functions that have counter-intuitive implications. This makes their use in real-time
markets such as the receding horizon approach described in Section 6.2 somewhat
implausible. Preliminary work described in Section 5.2.2 proposed the idea of a
fixed-iteration margin-based mechanism. This could be developed as a means of
clearing power and reserve markets simultaneously on intra-day time-scales without
an unreasonable exchange of information.

One possible extension of the ideas presented in Section 6.2 is the design of a
receding-horizon market clearing mechanism in which prices and bids are updated
continually, and possibly asynchronously, as a means of ensuring a constant down-
ward pressure on system costs. This could draw on results concerning asynchronous
distributed (dual) subgradient methods as analyzed by Nedic et al. [NBB01] or on
some system of bilateral trades [WV99]. It would also be beneficial for appropriate
reserve margins to be chosen as part of the same mechanism, since reserves are
typically provided at the expense of scheduled power.

7.5 AC optimal power flow

Continuing along the direction of results presented in Section 6.1, it would be at-
tractive to demonstrate that dual-based methods are capable of solving very large
time-coupled AC OPF problems. Although the ADMM-based decomposition down
to the level of individual lines reported in [KCLB12] was successfully implemented
for a large system, Andersen et al. [And11] suggest that unless an appropriate
chordal decomposition of the underlying constraint graph is used, the decomposi-
tion may not lead to an optimal solution of the original problem. There is still a
lack of clarity in how to decompose the semidefinite relaxation most efficiently.

More generally, there is a strong justification for studying the properties of the AC
OPF problem further. Despite the recent results demonstrating often-tight conic
relaxations of this difficult problem, many questions such as the following do not
appear to have been addressed sufficiently:

- When can a conic relaxation of the AC problem yield an exact solution?
  Anecdotal examples of cases where it fails are available [LMBD11], but no
  concise description of the limitation appears to be available. This would be
  useful for certifying the solutions of OPF problems in software applications.
7. Short- and Medium-Term Objectives

- Can conic relaxations be used for the analysis of transients and feedback control on AC networks?

- Can a dispatch planned on the assumption of a lossless or DC-approximated network model always be translated over to the AC model to account for losses and voltage constraints? What margins should be provided to generators in order to account for the change of network model?

- Can usable models of HVDC lines and other controllable components such as FACTS devices be incorporated efficiently into the relaxation?

These questions may be easier to answer using convex analysis after applying a conic relaxation to the OPF problem than would have been possible from inspection of the original non-linear constraints.
Chapter 8

Long-Term Challenges

This chapter raises two longer-term challenges prompted by the intermittent renewables growth described in Chapter [1] and it is possible that the robust and distributed optimization tools used in this dissertation may be of some use when it comes to tackling them.

8.1 Business as usual

It is commonly argued that the current methods for operating power systems and power markets will become unsustainable when large amounts of renewable energy are introduced [New10]. However this claim is not backed up by evidence, and the distinction between existing methods actually failing and existing methods simply becoming vastly more expensive has not been made. This has in part been made difficult by the variety of existing market structures in place (see the review in Chapter [1]).

It would nevertheless be interesting to use a more rigorous robustness approach to attempt to forecast levels of renewable penetration at which standard mechanisms, such as the European day-ahead-plus-balancing-payments approach, or the online optimization and pricing approach more wide-spread in the US, would fail. This could occur because, for example, the ramping of reserve generators cannot track wind fluctuations fast enough, or because system operators cannot dispatch such reserves fast or accurately enough. These insights would allow system operators and policy-makers to plan new market and operational rules before blackouts occur.
Chapter 8. Long-Term Challenges

It is also unclear whether the price spikes dogging today’s markets [WNPK+12] are an artefact of inefficiencies in the way markets operate (either without due regard to time coupling or without sufficient robustness [XT12]), or whether they in fact do provide appropriate signals for longer-term investment in fast-acting reserves. It would be useful to evaluate the effect of short-time-scale market failures on longer-term investments.

8.2 Sharing costs and profits at high renewable penetration

It is assumed that in the coming decades many grids will include a very high penetration of renewables, perhaps using them to satisfy all electricity requirements if combined with a large enough energy storage capability [RAG12]. Apart from the greenhouse gas reductions this would bring, an attraction of this transformation would be that power could be provided at almost zero marginal cost once the necessary infrastructure investments have been made. There is in fact some evidence that even at current penetration levels the cheapness of wind power is reducing spot electricity prices in some countries [SdMdRGV08, MRMS10].

However, this transformation would also bring with it the problem of how and whether an electricity market could still function. Clearing prices become difficult to determine when all production on offer has near-zero marginal costs and consumers still have a high willingness to pay for power, since the supply and demand curves fail to cross. Prices too near to the cost of generation (i.e., zero) will fail to reward investments made in generation, and high prices will result in undue profit to generation companies at the expense of consumers. Even if “fair” electricity prices solving this problem can be found, there still remains the question of how best to distribute the responsibility for generation amongst companies that will often have excess combined capacity.

A related question is whether energy storage units can still be operated using traditional “buy low, sell high” arbitrage strategies when prices are essentially flat. It is possible that some other system of incentives would be required in order to induce them to perform their usual smoothing function, since an energy storage industry that succeeds in totally flattening out variations in market prices will receive no net income at all under current assumptions. If this fails, the system
operator could adopt a command-and-control approach to managing stored energy.

These problems imply that future power systems with very high renewable penetrations will require drastically different market rules from those in use today. One could also argue that the transition period between the current energy mix and this future fully-renewable system will be even more difficult. This is because it is likely to be characterized by prices that are usually near zero, but with the added difficulty of spikes at the rare times when renewable infeeds simultaneously drop and expensive, seldom-used backup generation is needed to meet demand. The question of how to reward infrequently-used generators adequately was relevant even before the fluctuations of renewable in-feeds grew to significant levels [CS05] and is likely to become more pressing.
Appendix

DC load flow approximation: Computing the matrices $\Gamma_i$

Chapter 2 of this dissertation makes use of a reduced model of an AC transmission network known as a DC approximation in order to derive a tractable optimization problem. This is based on the following assumptions:

1. Transmission lines are lossless. More precisely, each line connecting node $l$ to node $m$ with complex reactance $X_{lm} = R_{lm} + jG_{lm}$ is modelled only by its susceptance, i.e., the real part of its admittance,

$$\sigma_{lm} = -\text{Im}\{X_{lm}^{-1}\}.$$  

Note that this is not equal to $1/G_{lm}$, though the latter may be a good approximation when $G_{lm} \gg R_{lm}$.

2. All nodal voltages are of equal magnitude $|V|$; this is a consequence of assumption 1. Only phase differences between nodal voltage phasors exist.

3. Phase differences between nodal voltages are small, so that power flow seen by node $l$ along the line towards node $m$ is given by $|V|^2\sigma_{lm}\sin(\theta_l - \theta_m) \approx |V|^2\sigma_{lm}(\theta_l - \theta_m)$.

The modelling approach in equation (2.10) of Chapter 2 relies on matrices $\Gamma_i \in \mathbb{R}^{2LT \times T}$, which describe the contributions of each participant $i$’s power output schedule $r_i + G_i\delta + C_i\mathbf{x}^i$ to line flows.

Since the network is memoryless on the minutes-to-hours timescales considered here, each matrix $\Gamma_i$ can be written in block-diagonal form

$$\Gamma_i = I_T \otimes \tilde{\Gamma}_i,$$
where $\tilde{\Gamma}_i \in \mathbb{R}^{2L \times 1}$ maps the contributions of power injected at the node to which participant $i$ is connected. In order to determine each $\tilde{\Gamma}_i$, the voltage phase angles must be written in terms of the bus power injections, and then eliminated ([CWW00] §III-B). For ease of communication the derivation is adapted here in the notation of Chapter 2.

Under Assumption 3 above, the line flows $P \in \mathbb{R}^{2L}$ (accounting for both directions of flow on each line) can be written as a linear combination of phase angles, represented using a vector $\theta \in \mathbb{R}^N$, so that, assuming per-unit voltages,

$$P = \Sigma \theta$$

for some $\Sigma \in \mathbb{R}^{2L \times N}$, whose rows contain zeros except for entries with $+\sigma_{lm}$ and $-\sigma_{lm}$ in the appropriate places accounting for each connected pair $(l, m)$.

In order to find a linear relationship between $P$ and the bus power injections, which will be denoted $p \in \mathbb{R}^N$ so that $[p]_l$ denotes the net power injected at bus $l$, a relationship

$$\theta = X p$$

is needed. From the conservation of power at each bus,

$$[p]_n = \sum_{m \in \mathcal{N}(l)} \sigma_{lm} ([\theta]_l - [\theta]_m),$$

where $\mathcal{N}(i)$ is the set of nodes neighbouring node $l$ (c.f. equations (6.1c) and (6.1d) for the full AC model in Chapter 5). From this relation a square matrix $B_T$ can be found such that $p = B_T \theta$. However this matrix is singular and cannot be inverted, and the physical interpretation of this fact is that an equal offset to all phase angles produces the same flows. This problem can be resolved by choosing a reference bus, say, $[\theta]_{N_n} = 0$, which leads to the relation

$$[p]_{1:N_n-1} = [B_T]_{1:N_n-1,1:N_n-1} [\theta]_{1:N_n-1}.$$  

The submatrix $\tilde{B}_T := [B_T]_{1:N_n-1,1:N_n-1}$ is invertible, and the relation

$$\theta = \begin{bmatrix} \tilde{B}_T^{-1} & 0 \\ 0 & 0 \end{bmatrix} p := T p$$

holds. Then the flows are given by

$$P = \Sigma T p.$$  

Each matrix $\tilde{\Gamma}_i$ is given by column $[\Sigma T]_{:,n(i)}$, where $n(i)$ is the node to which participant $i$ is attached.
Bibliography


Bibliography


[PMNP06] S. Parsons, M. Marcinkiewicz, J. Niu, and S. Phelps. Everything you wanted to know about double auctions, but were afraid to (bid or) ask. Technical report, Brooklyn College, City University of New York, NY, USA, 2006.


Publication List


multi-period optimal power flow exactly on AC networks with mixed participants, in American Control Conference, Montreal, 2012.


Curriculum Vitae

Sept. 2009 - Dec. 2013  Scientific Assistant  
Automatic Control Lab, ETH Zurich, Switzerland

July 2008 - July 2009  Control Design Engineer  
Wind Technologies Ltd., Cambridge, United Kingdom

Summer 2007  Summer Undergraduate Research Fellow  
Caltech, Pasadena CA, United States

Summer 2006  Intern, CPU Design Group  
ARM Ltd., Cambridge, United Kingdom

Summer 2005,  Testing & Development  
Easter 2006  Sherwood Scientific Ltd., Cambridge, United Kingdom

Oct. 2004 - June 2008  BA and MEng in Mechanical Engineering  
University of Cambridge, United Kingdom

Sept. 1997 - July 2004  Tiffin School  
Kingston Upon Thames, United Kingdom