Title
Market Design and Analysis for Uncertain, Flexible, and Decentralized Power Systems

Author
Mather, Jonathan William

Publication Date
2018

Peer reviewed|Thesis/dissertation

Market Design and Analysis for Uncertain, Flexible, and Decentralized Power Systems

by

Jonathan Mather

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Engineering - Mechanical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Kameshwar Poolla, Chair
Professor Pravin Varaiya
Professor Francesco Borrelli

Fall 2018

Market Design and Analysis for Uncertain, Flexible, and Decentralized Power Systems

Abstract

Market Design and Analysis for Uncertain, Flexible, and Decentralized Power Systems

by

Jonathan Mather

Doctor of Philosophy in Engineering - Mechanical Engineering

University of California, Berkeley

Professor Kameshwar Poolla, Chair

Contemporary power systems are characterized by increasing penetrations of renewable and distributed energy resources (DERs), which present both opportunities and challenges for reliable system operation. Uncertainty in renewable generation must be managed over both shorter and longer timescales, necessitating greater flexibility in the supply and demand of electricity. Fortunately, this rise in renewables has been accompanied by the growth of flexible energy resources such as energy storage and aggregated demand flexibility. However, these assets are generally not adequately integrated into existing electricity market structures leading to many benefits of their flexibility going unrealized. Additionally, distributed energy resources, such as rooftop solar and electric vehicles, present a challenge to the traditional top-down management of power systems, requiring new decentralized control and market architectures to ensure their safe operation and to maximize their value.

This dissertation tackles a range of questions relating to the integration of flexible and distributed resources into electricity markets in the presence of uncertainty, using techniques from control theory, game theory, multiparametric programming, and distributed optimization.

The main contributions of this dissertation encompass novel analysis and design of markets, both centralized and decentralized, that explicitly integrate distributed and flexible resources, numerical methods providing exact solutions to multi-leader follower equilibrium problems among strategic bidders, and insights into the role of virtual bidders in managing demand uncertainty in two-stage markets.

- Chapter 1 introduces the context for the dissertation, describing trends driving the energy industry and the challenges and opportunities they pose. Current practice and existing literature are described with regard to uncertainty, flexibility, and decentralized markets and control.

- Chapter 2 lays out the mathematical background for the dissertation, describing models for power flow, electricity markets, and market participants, in addition to the theory of multiparametric programming.

- Chapter 3 demonstrates the equivalence of virtual bidding and stochastic optimization in two-stage markets under certain conditions. It also characterizes the equilibrium of a *crowd* of virtual bidders with heterogeneous beliefs and explores simple learning strategies for virtual bidders to reach equilibrium.

- Chapter 4 studies robust Cournot-Bertrand equilibria among generation firms. Firms anticipate demand and renewable uncertainty and optimize their actions to ensure their profits are robust to this uncertainty. The chapter concludes by discussing the effects on market outcomes of this behavior.

- Chapter 5 addresses the more general topic of strategic bidding by generation firms in electricity markets. This question involves solving non-convex bilevel optimization problems, which form part of the wider class of multi-leader follower problems. Both the single firm problem, constituting an MPEC, and the multi-firm equilibrium problem, constituting an EPEC, are solved exactly using techniques from multiparametric programming. This latter contribution, in particular, represents a significant advance over current methods, generalizing the solution of supply function equilibrium problems to cases with asymmetric players and transmission constraints.

- Chapter 6 analyzes a number of regulatory models for the integration of energy storage into wholesale electricity markets, including *Open Access Storage*, and investigates the implications and incentives of a variety of storage ownership structures under each model.

- Chapter 7 demonstrates the potential of load flexibility for congestion relief, defining the notion of a *congestion-free dispatch*.

- Chapter 8 describes the design of a decentralized coordination mechanism for DERs in a microgrid using blockchain technology. The implementation uses ADMM with a smart contract on the blockchain serving as the ADMM coordinator.

- Chapter 9 illustrates the design of an open-gate forward market with specific flexibility products for DERs, allowing users to hedge better against uncertainty and correctly express and maximize the value of their flexibility. A number of algorithms are presented to address the combinatorial scheduling and pricing problem involved with flexible loads.

To anyone who is writing, or has ever written a PhD Dissertation.

You are not alone.

Contents

List of Figures

List of Tables

Acknowledgments

Finishing the PhD has been one of the hardest things I have ever had to do. I am only here because of the efforts and good nature of a whole host of individuals, to whom I am eternally grateful.

Firstly I am indebted to my committee chair and PhD supervisor, Prof. Kameshwar Poolla. Throughout my time at Berkeley, Kameshwar has been a fantastic academic patron, giving me immense freedom and support to explore a diverse array of subjects and interests, to which the length of this dissertation is a fitting testament. He has always been there for me when it counted, and I feel lucky to have been his student.

Prof. Pravin Varaiya has an incredible ability to ask precisely the right question, and I have valued his advice and support throughout the PhD. Prof. Francesco Borrelli is a fantastic teacher, and I have appreciated his classes and in particular, his excellent textbook which has informed much of this dissertation. Prof. Shmuel Oren was also very generous with his time and gave helpful and insightful comments on a range of my work.

Other professors who have illuminated and enriched my journey at Berkeley include Prof. Scott Moura, who welcomed me into his research group and has always been a brilliant sounding board on matters both academic and personal. Prof. Sascha Von Meier gave me my first real introduction to power systems, and Prof. Severin Borenstein pushed my exploration of the subject further, also giving me the great pleasure of beating all the MBA students in his electricity market game.

I have been blessed with an exceptional host of colleagues and friends at Berkeley. Roel Dobbe and Jaime Fisac began the journey with me, and I have always been blown away by their intelligence, wit, and humanity. I'm proud to call them friends and excited to see where their careers will take them. Caroline Le-Floch persisted with the startup when I couldn't, and we had some great adventures together. Eric Munsing was a superb academic collaborator and remains a great source of wisdom. I had brilliant conversations with Junjie Qin in the later stages of the PhD that shaped a number of ideas in this dissertation, and I am grateful for his contributions.

To everyone in the lab, Jared, Chris, Octavio, Emmanuel, Akhil, Hamid, Sen, Galaxy, and Kate, I will miss you all dearly. To the Godfather, The King, and Daddy, you are all clearly very secure in your masculinity.

To the Cal Jazz Choir, you gave me so much joy and the greatest bunch of singers and friends I could have hoped for. Being part of the group was a transformative experience and remains one of the most treasured memories of my time at Berkeley.

Back in the UK, Charles and Ellie have been a constant source of support and friendship and I owe them so much. JPC, James, Ivan, Owen, Laura, and Alison have always been there for me, even when I would disappear for months on end, and I am so happy to have them in my life.

To my parents, I cannot thank you enough. You have given me so much throughout my life, and I hope that I've made you proud. I've never lost my sense of home, even if I do

have a tendency to find far-flung parts of the globe to spend years of my life. I think you gave me the bug in the first place!

To my brothers, Nif, Ekul, and Xela. Noj says nyeee. Thanks for keeping me grounded and I'm so proud of you all.

To Rose. No one has had more patience, more grace, or put up with as much as you have. The PhD has been as hard on you as it has on me, and I'm so happy for both of us to finally be out the other side. You are a saint, I love you, and I can't wait to see what life has in store for us next!

Chapter 1

Introduction

The world's energy infrastructure is transforming, characterised by four major trends commonly referred to as the four D's of power systems: Decarbonization, Decentralization, Digitalization, and Democratization.[1] These trends are being driven by technological advancement and social necessity, and present distinct opportunities and challenges to the status quo.

Decarbonization

Mitigating the risks of dangerous climate change requires a significant reduction in the contribution of the energy sector to global greenhouse gas emissions. In a recent special report published by the IPCC on the impacts of global warming of 1.5°C they state that the world must reduce its carbon emissions by 45% by 2030 from 2010 levels to limit warming to no more than 1.5 degrees above pre-industrial levels [3]. As can be seen in Figure 1.1, electricity and heat production is the largest sectoral contributor to global GHG emissions, making up 25% of total emissions in 2010. Consequently, decarbonizing electricity and heat production is a crucial step towards reducing emissions and mitigating climate change. Making this more difficult is a predicted growth in global energy demand of 28% by 2040 [4], due to increasing global population and developing countries seeking to grow their economies and improve their quality of life. Additionally, the majority of proposals for the decarbonization of transport, which makes up 14% of global emissions, involve significant electrification, whether directly through electric vehicles or indirectly through the production of hydrogen, which will further increase the demand for electricity.

Initial decarbonization efforts were led by climate policies such as feed-in-tariffs, renewable portfolio standards, and carbon markets, causing the closure of large thermal power plants, and encouraging the development of low-carbon generation technologies such as solar and wind power. However, in recent years the shift is being driven more and more by pure economics as wind and solar now represent the cheapest form of new power anywhere in

[1] A fifth D might be deregulation, although this has mostly been superceded by decentralization and democratization.

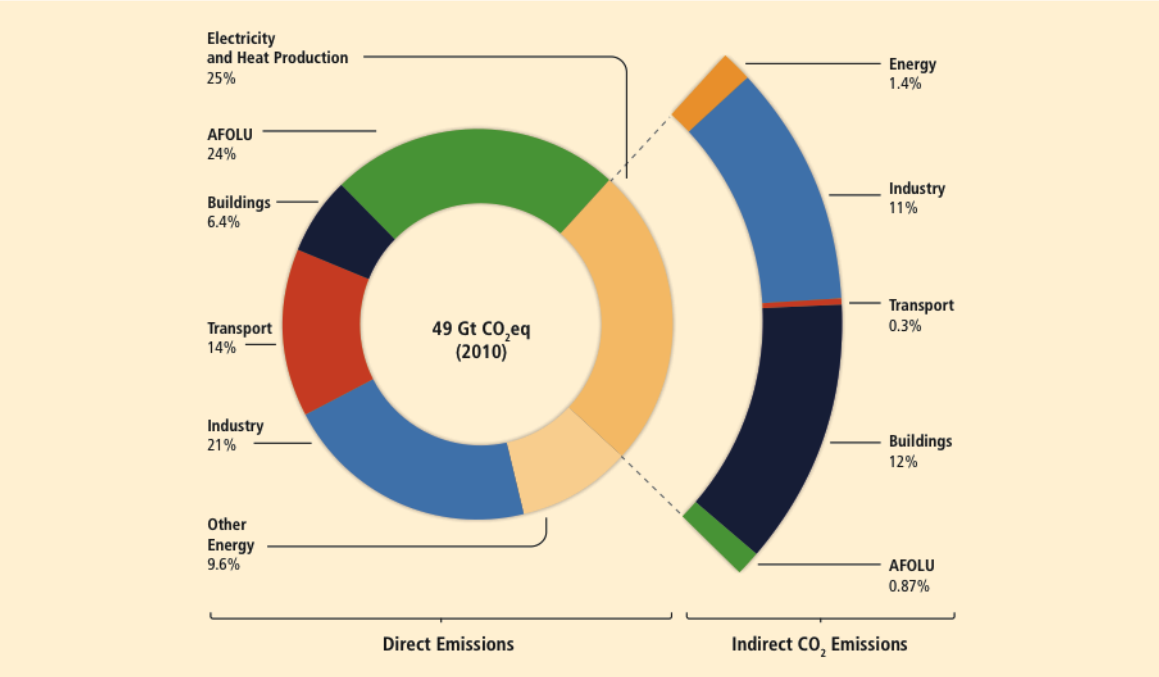


Figure 1.1: Global GHG emissions by sector in 2010. Figure from IPCC AR5 [1]

the world, beating out even natural gas turbines on a cost per watt basis. While they may provide cheap energy, renewables present serious grid integration issues as the power they generate is intermittent and uncertain. For grids with high penetrations of renewable generation, uncertain fluctuations must be smoothed, and intermittent capacity must be firmed to ensure the lights stay on.

Decentralization

Traditionally, power systems have been operated in a top-down fashion with centralized control of large, transmission-connected, thermal power plants, pushing power out to consumers on the grid edge. This paradigm is being called into question by both an increase in *embedded generation* connected to the distribution network, and the rise of distributed energy resources (DERs) such as residential solar PV, storage, and electric vehicles (EVs). Markets and control strategies that were designed for hundreds or thousands of resources now face having to scale to tens or hundreds of thousands of active, controllable resources. Distribution networks and protection systems that were designed for one-way power flow with passive consumption and control now face bidirectional power flows, active consumption, and the need for more dynamic control to ensure power quality.

Proposals to overcome these challenges generally involve hierarchies of decentralised markets and control strategies, managing balancing and constraints at multiple levels of the power system, in addition to developing new mechanisms to enable DER participation. This

set of ideas is commonly referred to as *Transactive Energy*, opening up existing markets to DERs where possible and creating new markets and control schemes more suited to the characteristics and capabilities of DERs where not [5]. This includes discussions around dynamic pricing, aggregation, and new business models for utilities, such as the role of a *Distribution system operator* (DSO). A DSO is similar in function to the ISO at the transmission level, administering new markets for local energy and services and managing distribution system operation.[2]

Digitalization

Designed in an age before the internet and smartphones, the communications and control infrastructure underpinning the power grid has not kept pace with current technology and is now being pushed to its limits. In today's world, information and computing are both ubiquitous and cheap. Vast quantities of data can now be processed, communicated, and stored, at speed and scale. The potential applications of this technology to the power system have been termed the *Smart Grid*, a physical power system combined with a cyber layer of communication, control, and automation, heralded to deliver improvements in operations, higher efficiency, and lower costs for consumers. Digital technology enables sensing and control at far more granular levels in terms of both location and time.

An unfortunate consequence of greater reliance on digital technology is an increased vulnerability to cyber attacks that can compromise grid reliability and lead to blackouts and component damage. Attack vectors can include false data injections which lead control systems to take unnecessary and potentially damaging actions, all the way to direct override of power plant or system control. Such an attack in Ukraine in 2015 took out 30 substations leaving over 230,000 residents in the dark [6]. Additionally, the vast proliferation of data emerging from the smart grid associated with energy resources and energy usage has led to concerns over privacy and data breaches. For example, the occupancy of a home can easily be inferred from its energy consumption profile [7].

Democratization

Power systems have come a long way from natural monopolies and large vertically integrated utilities, with the energy industry now opening up to a wide variety of new players and entrants. Homes and businesses can now own DERs such as solar PV, storage and EVs, allowing them to export power as well as consume it. New business models such as aggregation and energy management[3] have emerged to enable consumers to make the most efficient use of

[2]The idea of a DSO as an alternative Utility business model arose predominantly in response to the so-called 'utility death spiral' where a Utility's ratebase shrinks as more customers self-generate and defect from the grid, but the Utility remains on the hook for existing sunk costs and new investments required to upgrade network infrastructure to deal with the impacts of DERs.

[3]This includes both residential and commercial energy management, generally using DERs to minimize energy bills.

their energy assets, in addition to coordinating the actions of many small resources to provide services to grid and distribution operators. Growing awareness of climate change and the ubiquity of DERs has also led to a rise in *Community energy*, encompassing everything from crowd-funded community solar projects, self-consumption communities [8], local microgrids, to community choice aggregators (CCAs). Community energy often focuses on consumer choice in the provenance of energy, with a general preference for local, low-carbon sources of energy. An extreme of both community energy and the aforementioned transactive energy, and perhaps the ultimate form of retail choice, is peer-to-peer energy (P2P). In theory, P2P enables all energy actors (homes, businesses, retailers, suppliers, generators) to buy and sell energy and services as necessary, rather than relying on traditional gatekeepers. It remains to be seen whether energy customers have the desire to engage on such an active level with their energy usage and become traders in their own right, but the principle certainly makes sense for larger participants and aggregations.

In this dissertation, I seek to address aspects of three technical challenges that must be solved to address the industry transformation described by these four D's. These are Uncertainty, Flexibility, and Decentralized Markets and Control.

1.1 Uncertainty

As we have seen, higher penetrations of renewable energy entail much greater uncertainty on our power systems. Uncertainty is indeed nothing new to the grid, and a variety of mechanisms already exist for both physical and financial risk management. However, it remains to be seen if these will hold up to scrutiny at ever higher levels of renewable penetration.

Today, physical risk management is required in power systems to handle load forecast error and deviations arising from unanticipated faults and contingencies on the grid. System operators have typically taken a robust approach to this uncertainty, ensuring the grid remains secure and stable in the face of any failure. The first line of defence is the *N-1 security criterion*, which broadly states that the power system should remain in a feasible and stable state following the loss of any single generator, transmission line, or transformer. In practice, operators ensure the system is robust to a set of *credible contingencies*, which leads to additional constraints being included in commitment and dispatch optimization problems. Moving to real-time operation, system operators procure both frequency response and reserve capacity to manage minor imbalances and more substantial contingencies respectively.[4] Frequency response regulates small imbalances between generation and demand and operates via automatic control systems which respond proportionally to frequency deviations. Reserve capacity is usually sized to at least the largest generator on the system or a given fraction of the peak load and is generally split into spinning and non-spinning

[4]Frequency response and reserves are generally split into Primary, Secondary, and Tertiary frequency response, differentiated by their response time, and capacity. In control theory parlance, primary response can be thought of as a proportional control, and secondary and tertiary response can be thought of as integral control, where secondary response is automatic and tertiary response is manual.

reserve. Spinning reserves are typically triggered by automatic generation control (AGC), and non-spinning reserves are typically called up manually by system operators.

Financial risk management mitigates uncertainty and variance in returns or costs to electricity market participants, both in investments and operations. The electricity spot price is volatile due to uncertainty in a range of factors, both short and longer term. In the short-term, this includes load forecasts, renewable generation, fuel costs, and transmission congestion. In the long term, this includes demand growth, impacts of new market entrants, development of new technologies, and regulatory risk. Since the majority of market participants are risk averse, this leads participants to hedge against this uncertainty, generally paying some risk premium for more predictable cash flows. For example, many electricity distributors and retailers sign long-term power purchase agreements (PPAs) with suppliers for some or all of their energy needs, formulated as contracts for difference settled against the market spot price. This insulates both buyers and sellers from the volatility of the spot price. Participants can also purchase financial transmission rights (FTRs), which provide a perfect hedge against transmission congestion and associated congestion rents.[5]

Closer to delivery, multi-stage markets enable participants to further hedge and limit their exposure to the spot price. In North America, this typically takes the form of a two-stage market, with Day-Ahead (DA) and Real-Time (RT) energy markets. In the DA market, participants submit bids and offers to the system operator for each hour of the coming day, and the market is cleared by the system operator, incorporating system and resource constraints in the optimization. The outcome of the DA market is a schedule of physical dispatch commitments and a set of DA clearing prices, producing one financial settlement. The RT market, unsurprisingly, is run closer to real time, generally at five- or fifteen-minute intervals. The RT Market balances the differences between DA commitments and actual RT demand and supply of electricity. The RT market produces a separate, second financial settlement, and establishes the RT locational marginal price (LMP) that is either paid or charged to participants in the DA market for demand or supply that deviates from their DA commitments [9]. The DA market can be thought of as a forward market which is settled against the RT price. An additional mechanism for participants to hedge against the spread between DA and RT, and for financial speculators to profit, is known as *virtual bidding* (VB). A detailed discussion of VB is given in Chapter 3, but briefly it allows participants to take virtual supply or demand positions in the DA market, which are cleared on equal footing with physical bids, and must be entirely liquidated in the RT market, allowing participants to arbitrage the price spread between DA and RT markets.

Following the DA market, the system operator will generally run a second optimization after the DA market has closed, known as reserve adequacy analysis (RAA), in which resources will be committed or decommitted based on the operator's forecast of system load, rather than what loads have bid in to the DA market, and to satisfy reliability constraints. In this way the DA and RAA also represent a form of physical risk management for the system operator, having an idea of system state and resource output ahead of time, rather

[5] A more detailed discussion can be found in Chapter 6

than just relying on whatever is available in real-time.

Uncertainty in power systems then presents three issues. First is the ability to feasibly mitigate this uncertainty through appropriate control and planning with sufficient resource availability and capability to deliver all reliability services that are required. The second is, assuming feasibility, to mitigate this uncertainty at the lowest cost. System operators often value reliability at the *value of lost load* (VOLL), representing the cost incurred by all consumers if they were to lose power. However, this number is incredibly difficult to calculate accurately, and current estimations are generally thought to be far too high. On this basis, system operators ensure that the lights stay on at almost any cost and have little incentive to economize as costs are usually socialised among all participants, or indeed only energy consumers. Third is the ability of market participants to hedge against this uncertainty appropriately.

The first and second issues can be addressed in part by flexibility, which we will discuss in the subsequent section. The second issue of mitigating uncertainty at the lowest cost can be addressed through changes in market design and the current philosophy of system operation. System operators currently think of uncertainty robustly, managing the system and procuring reserves to be resistant to the worst thing that could happen at any single time. However, this approach fails to recognize that the worst thing that could happen is generally improbable, and is thus extremely conservative. An approach to managing the degree of conservatism for a unit commitment problem is proposed in [10], and for multi-period economic dispatch in [11] using adaptive uncertainty sets. An alternative approach that takes into account the likelihood of any given event or contingency, explicitly accounting for the distribution of uncertainty, is referred to as *Stochastic*. In this way, the expected system cost is minimized, and the likelihood of a constraint violation or system failure can be explicitly parametrized through chance constraints. Such a method is used to solve a stochastic security constrained unit commitment problem in [12]. Anecdotally, such stochastic approaches, despite being well understood, have not been taken forward or implemented by ISOs, as they remain very opaque to actual system operators in ISO control rooms. It seems they prefer to retain full control of system reliability, rather than let an algorithm decide which reserves are unlikely to be needed on a given day.

The second and third issues can also be dealt with by introducing additional multi-stage markets or continual rolling (forward-gate) markets, in contrast to the two-stage markets of today. Uncertainty generally decreases as delivery approaches and introducing additional market stages allows participants to adjust their positions based on any new information that has emerged since the previous market stage, i.e. an updated forecast. System operators can also better plan resource commitment based on new information at each market stage to minimize the cost of providing reliability. While this brings benefits in the management of uncertainty, there is, however, always an inherent trade-off between the complexity and usability of a given market design. Introducing additional market stages or rolling markets increases the burden on traders and participants to process more information, make complex computations to inform their bidding strategy, and interact with the market more actively and regularly. The additional benefits of multi-stage markets are of no use if no one par-

ticipates! It could be argued that the success of the current two-stage market structure lies in its simplicity. With the advent of automation and machine learning, such rapid trader interactions may no longer be so far-fetched, and we address such a rolling market structure in Chapter 9 in the context of decentralized market design.

In the literature, one method, referred to as Risk-Limiting Dispatch, combines the two approaches of multi-stage markets and stochastic optimization to characterize the optimal decision for a system operator at each market stage [13]. Again, despite being well understood, perhaps for reasons of complexity or operator resistance, this method has not yet made its way into industry practice.

In this dissertation, we consider uncertainty from a range of angles and participant perspectives. In Chapter 3 we show that an existing mechanism, Virtual Bidding, can under certain conditions implicitly capture many of the same benefits of explicit stochastic optimization. In Chapter 4 we demonstrate the impacts to participants and equilibrium market outcomes of generation resources robustly accounting for demand uncertainty in their bidding. In Chapter 5 we take this a step further and directly address strategic bidding, considering the single-firm and multi-firm problems, presenting methods for handling uncertainty in the single-firm case. In Chapter 9 we propose a forward-gate market design for a distribution-level flexibility market, which allows participants to best hedge against uncertainty.

1.2 Flexibility

To manage ever-increasing penetrations of renewable generation, power system operators require a higher capacity of flexible energy resources and significant volumes of energy storage. Flexibility can be broadly characterised as the ability of an energy resource to increase or decrease its net power injection in response to a control or price signal. Ramping capability and speed of response are usually implicit in this definition, with resources that can ramp up or down more quickly being *more* flexible.

One striking real-life example of the need for additional flexibility is the *Duck Curve* observed in the CAISO net demand profile and shown in Figure 1.2. California leads the US in both residential and grid-scale solar installations, and capacity is only predicted to grow leading to the huge dip in the net demand curve seen in the middle of the day. The evening ramp along the *neck* of the duck requires a rapid increase in power generation of roughly 4 GW/hour in the CAISO 2020 forecast [14]. This is expected to exceed the ramping abilities of the existing generation fleet, resulting in significant overgeneration during the mid-day hours in order to ensure that evening peak load can be served. Assuming no additional flexibility the only solution to this problem would be to significantly curtail solar generation, resulting in additional carbon emissions and energy costs.

All types of resources, with perhaps the exception of nuclear power, can provide flexibility, but this is always at some cost, and in the case of generation at the expense of providing energy and often additional emissions. Thermal generation can provide flexibility both up

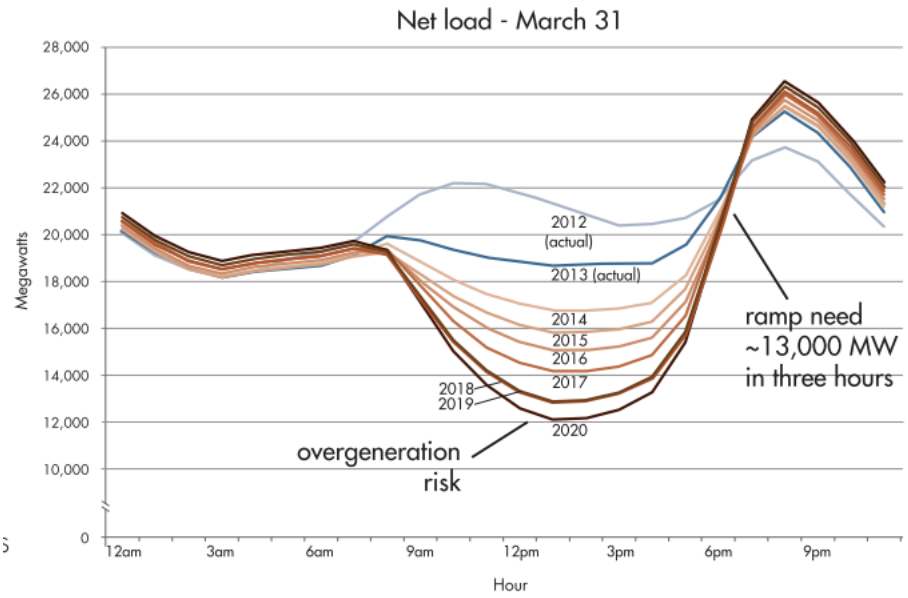


Figure 1.2: Forecasts for a 'duck curve' with high renewable penetrations [2].

and down by leaving some headroom between their setpoint and their capacity, allowing them to ramp up to their capacity, or ramp down to zero. Renewable generation, while uncontrollable, can provide a similar service by pre-curtailing their output, allowing them to ramp up to their maximum available output, or ramp down by curtailing further [15]. On the demand side, load flexibility comes in two general forms: temporal load flexibility[6] and demand response. Temporal load flexibility refers to the ability of a load to shift their consumption in time: controlling the start time of a deferrable load, controlling the power profile of a shapeable load, or reducing consumption in one period and increasing consumption in a later period.[7] Demand response is a loose term but habitually refers to loads reducing their consumption for some period, either for economic reasons or in a system emergency, with no explicit anticipation that consumption will be recovered in a later period.

Energy storage has been described as a swiss army knife for the future grid with the ability to provide instantaneous flexibility up and down for both short and long durations dependent on its state of charge, making it suitable for many applications, including arbitrage, congestion relief, voltage support, frequency regulation, firming renewable power, and deferring infrastructure investment costs. Developments in telecommunications, personal electronics, and most recently electric transportation, have dramatically improved the quality and reduced the cost of electric energy storage, with the cost of lithium-ion batteries

[6]Spatial load flexibility is also possible, for example deciding to charge electric vehicles at one location rather than another, or assigning computing jobs to one data center rather than another. However, this remains a far more limited application than temporal load flexibility, so we do not address it further in this dissertation.

[7]Full technical definitions of these terms are provided in Chapters 8 and 9.

falling from $490/kWh in 2014 to $225/kWh in 2018 [16]. Grid-scale storage, once restricted to large, geographically dependent pumped hydro installations, can now be provided by modular installations of lithium-ion batteries. In 2017 the US exceeded 1GWh of installed capacity of grid-scale energy storage [17]. Such deployments have for the most part been driven by state and federal mandates, recognizing the critical system benefits of energy storage. In particular, the California Storage Mandate requires the three large investor-owned utility companies in California to install 1.3GW of energy storage by 2020, in addition to procuring 500MW of behind-the-meter storage [17]. Recent examples of significant energy storage deployments include the 100MW/129MWh battery in South Australia, developed by Tesla, brought online at the end of 2017, and three projects totalling 70MW in Southern California to mitigate the Aliso Canyon gas leak, brought online in January 2017 by AES and Tesla.[8] Larger developments are planned across the world with a 200MW/800MWh Vanadium Flow battery under construction in Dalian, China, and two projects under development by PG&E near the Bay Area, 300MW/1200MWh and 182.5MW/730MWh by Vistra and Tesla respectively. [18]

While much development has been in front-of-the-meter, behind-the-meter and residential energy storage represent a growing fraction of new installations in the US, with 180MW expected to be installed in 2018, and over 1.5GW forecast to be installed by 2023 [16]. Commercial and industrial customers can often save on demand charges by installing batteries to smooth their loads, with project developers additionally using the battery to provide services to grid operators when feasible. With offerings such as the Tesla Powerwall, residential customers are increasingly being sold on energy storage as a source of backup power, in addition to maximizing the value of solar installations.

In addition to behind-the-meter storage, flexible and controllable loads also represent a growing class of flexible resources. Demand response programs have existed for many years, but have generally been limited to load shedding in an emergency, rather than anything more controlled or market-based. Aggregations of flexible loads, such as electric vehicles, smart thermostats, and pool pumps can provide both reductions and increases in load, either scheduled or in response to prices, in addition to providing ancillary services to grid operators. In this way, these assets can be coordinated to act as *virtual storage* mimicking the operation of a battery [19], or combined with behind-the-meter storage and distributed generation to create *virtual power plants* (VPPs) [20].

The need for flexibility is driven first by a *feasibility* argument, as discussed in the previous subsection and illustrated by the duck curve in Figure 1.2. Ever greater capacity of flexible resources will be required to manage renewable variability and intermittency and associated ramping requirements. It is also desirable that this flexibility should be low-carbon or at least carbon-neutral, not detract from the operation of existing firm power plants, and avoid the curtailment of renewable generation. In this way, grid-scale storage and distributed energy resources present the only viable future-proof solutions for energy storage. As we

[8]The author developed bidding and storage optimization algorithms for the Tesla Hornsdale Power Reserve battery in South Australia.

have seen, volumes of these assets are predicted to proliferate in the coming years. However, these assets currently face a lack of appropriate integration into existing energy markets, and in many cases, markets are missing altogether to capture and maximize the value of DERs, leading to flexible resources being underutilized.

At the transmission level, many system operators have minimum capacity requirements, typically on the order of 1MW, to access wholesale markets, limiting the ability of smaller assets to participate. Additionally, assets typically require a single meter, making it challenging to integrate aggregations of smaller resources. Furthermore, very few if any ISO dispatch optimizations include explicit time coupling between time periods, meaning that loads with flexible schedules cannot appropriately express their flexibility to the system operator. On a similar note, there is often no accounting for the specific characteristics of energy storage in these optimizations, with storage being forced to enter the market alternatively as a demand resource, then a generation resource. This means that a storage module must plan its charge and discharge schedule ahead of time, before knowing energy and service prices, to avoid making infeasible bids, rather than merely expressing its capability and availability. Finally, very few markets offer the opportunity to *revenue-stack*, a key piece of the business case for building flexible resources, in particular, energy storage. Flexible resources can often provide multiple services simultaneously in a 'non-rival' manner, although system operators frequently only permit a resource to provide a single service at a given time. In summary, flexible resources face a market integration issue, where their capabilities are not fully recognised by existing market structures, and their access to markets is limited, leaving system benefits and private value unrealized.

Some progress has been made in recent years on this integration issue. In terms of new products targeted at flexible resources, in 2016 CAISO introduced a flexible ramping product to ensure they have sufficient capacity to combat the duck curve. Also in 2016, the UK TSO National Grid introduced an additional frequency response product targeted at batteries called Enhanced Frequency Response (EFR) with 1-second response time. They are also currently consulting on opening access to their frequency products and the balancing mechanism to non-traditional providers. Regarding participation and access, in 2016 CAISO introduced a new class of market participant aimed at aggregations, a distributed energy resource provider (DER). This allows a DERP to aggregate individual DERs in order to meet the CAISO minimum size requirement, but they are restricted geographically due to concerns around congestion management. DERP aggregations within a single pricing node may be heterogeneous, i.e. a mix of DER types, and are not restricted in maximum capacity, whereas DERP aggregations spanning multiple pricing nodes must be homogenous, i.e. the same type of DER, and are restricted to 20 MW in capacity. This potentially limits the feasibility of spatially diverse VPPs. At the national level, in 2017 FERC issued order 841 which "directs RTOs/ISOs to establish a participation model consisting of market rules that, recognizing the physical and operational characteristics of electric storage resources (ESRs), facilitates their participation in the regional transmission organisation (RTO) and independent system operator (ISO) markets." [17] This has led to a discussion around the best method of integrating energy storage into energy markets, including how they should

be optimized, controlled, and remunerated.

We address the issue of storage integration in Chapter 6, where we analyse private incentives, system benefits, and the impacts of storage ownership, comparing a proposed mechanism, Open Access Storage (OAS), to the status quo of merchant storage operation. Regarding addressing markets that appropriately incorporate the characteristics of flexible resources, in Chapter 9 we present the design of an explicit flexibility market with products that enable resources to express their availability and capabilities fully. This market is designed for local and decentralised assets, the context for which will be addressed in the following section, however, a similar design could just as easily apply at the transmission level.

A final application of flexibility lies in constraint management and deferred transmission and distribution investment costs, so-called *non-wires alternatives* (NWA). When faced with upgrading a section of the distribution network due to constraints, utilities employ demand management and flexibility to resolve the constraint at a lower cost than installing new components. This idea has received much attention as part of the New York 'Reforming the Energy Vision' REV process, and California's Distribution Resource Plan proceeding. As of mid-2017, New York had 956MW of NWA projects in the pipeline and is home to the largest distributed NWA project to date, the Brooklyn Queens Demand Management project, which encompassed 46MW as of the fourth quarter of 2016 [21]. In Chapter 7 we describe how load flexibility can be used to manage and avoid constraints at both the transmission and distribution level.

1.3 Decentralized Markets and Control

The problem of decentralization can broadly be thought of as how to determine the *right price* for a given product or service at any given time and location.[9]. The right price, in this case, is the economically efficient one, which is, in general, the marginal cost of providing a given product or service at a given time and location. It should be noted that the right price takes into account the preferences of participants (their willingness to pay or sell), all operational (participant capacities and capabilities) and system constraints (network capacities, security constraints), and should technically include all externalities. One could imagine that if this right price were communicated to all participants by some oracle, and assuming all participants were rational, then all participants would behave efficiently and the greatest social welfare would be generated by those actions. Of course, despite the fact that the right price does exist, in reality, no such oracle exists, so a mechanism must be created for discovering this price and communicating it to participants. Usually, this role is left to markets, where participants communicate their willingness to pay or sell through bids and offers, which are cleared either through a central limit-order book as in stock markets, or through centralized clearing and welfare maximization as is performed by ISOs. There

[9]One could also add to this determining the correct futures or forward price at any given time and location for a given product or service

are a number of issues in both price-discovery and communication of price in power systems populated by vast numbers of DERs.

The first obstruction to the right price in power systems, not specific to DERs, is the decoupling of electricity markets and power system operation. Power systems must balance supply and demand instantaneously, however running a market on the same time scale would place an insurmountable informational and computational burden on traders and market operators. The solution is to discretize, with energy being sold in blocks (5 minutes, 15 minutes, 1 hour), and then contracting capacity to deliver ancillary services such as frequency response, with automatic control systems managing deviations within this capacity. It is interesting to note that a hypothetical instantaneous power price would obviate the need for both energy and capacity markets. Some recent work has demonstrated a way to break the economic and physical hierarchy using innovative control design limited to smart inverters [22]. However, this work is far from the mainstream in terms of current practice. For this dissertation, we will assume the hierarchy has not been broken and that we have both energy and ancillary service markets.

The second obstruction, again not specific to DERs, is strategic behavior in markets, including the exercise of market power and strategic bidding. In these scenarios, participants do not reveal their genuine willingness to pay and use monopolistic or oligopolistic positions to increase their profits at the expense of other participants and social welfare. This can be mitigated in the first place through antitrust legislation, and penalizing any attempts at market manipulation, as is already done by FERC. Second, it is an issue of mechanism design, creating markets that are incentive-compatible encouraging participants to reveal their actual preferences. There is always a trade-off in mechanism design, perhaps in welfare maximization vs obtaining truthfulness, but electricity market designs have remained relatively unchanged in the last twenty years, and the advent of smaller decentralized markets offers an opportunity for innovation in this area. We do not specifically address mechanism design in this dissertation, although it remains an interesting avenue of research.

The third is one of full participation and market integration. If a market does not include all participants, who have some willingness to pay or sell then it is missing information and cannot discover the right price. This issue was mentioned in the previous section in that DERs often do not have access to wholesale markets, and thus cannot add additional capability to these markets or maximize their value. Including all distributed assets in a single centralized clearing is unlikely to be the answer as this would present a significant computational, communication and control challenge to the system operator due to the sheer number of assets that would have to be considered. On a related note, the fourth obstruction is that markets often do not account for the specific characteristics, availability or capabilities of DERs. Again, as mentioned in the previous section a lack of time coupling in market optimization prevents resources with temporal load flexibility from revealing their true capability, willingness to pay, and maximizing their value. The fifth is that smaller DERs are frequently behind-the-meter and face flat retail rates that do not account for time of use or correlate in any way with wholesale prices. This means that DER owners have no incentive to optimize their consumption and generation in ways which would provide value

both to utilities, system operators, and the DER owner themselves.[10]

A number of solutions present themselves to these latter two issues, all falling under the umbrella of transactive energy. The first, and least extreme, is retail rate design. Time-of-Use (TOU) and dynamic pricing expose retail customers to the time-varying value of electricity, allowing them to optimize the operation of their DERs against meaningful price signals. The second is aggregation, coordinating the actions of a large number of smaller DERs so that they appear as a single asset, either a virtual battery or VPP, to the system operator at the wholesale level. The third is the creation of microgrids, small subsections of the distribution grid containing DERs which can island themselves from the main grid. Management and control of DERs are either handled centrally by a microgrid operator or through some market mechanism. The fourth involves decentralizing some or all of the dispatch and system operation functions of the ISO to distribution system operators (DSOs). DSOs would be responsible for active management of distribution systems and potentially facilitating decentralized local marketplaces for energy, balancing, and local services.

All of these solutions involve the coordination and control of DERs in some fashion, but the exact mechanism and exactly who has control can differ by application. In one mechanism, an aggregator or utility can offer a rebate or regular payment for voluntary customer action or direct control of DERs. OhmConnect is an example of an aggregator providing demand response services to the ISO through rebates to customers who voluntarily participate during *Ohm hours* by reducing their consumption [23]. Some utilities, and potentially future DSOs, want to retain tight control over distribution systems and may require direct control over DERs [24]. The DSO can then optimize and control DERs centrally, but it is unclear how this aligns with the private incentives of the DER owners. Rather than send control signals a DSO or utility may send price signals, similar to the concept of dynamic pricing. Under this mechanism, DER owners retain control of their devices and optimize against a price signal delivered by their utility or a DSO, minimizing the control and communication overhead. However, it is unclear exactly how a dynamic price should be derived and how to ensure that this is the 'right price'. This price-discovery question can be resolved by introducing decentralized local markets for DERs to buy and sell energy and services. How these decentralized markets should interact with one another, and intermediate or wholesale markets above them, remains to be seen. Some degree of integration or coordination must be required to ensure meaningful price signals being passed up and down the hierarchy.

Focusing on decentralized markets, these can function at a number of levels, with some propositions being more mature and others more speculative. One missing market that would be feasible to implement today would involve local balancing and constraint management, connecting embedded generation, microgrids, aggregators, retailers, suppliers, and DSOs. The market addresses a current need, and these entities would all be of sufficient size that managing trading operations would not be an issue. Also, they all, with perhaps the

[10] An issue related to retail rates is the ability or lack thereof of consumers to self-generate or being forced to sell their power to the utility at below-market rates. Policies such as net energy metering (NEM) which overvalue DERs by paying them a retail rate are good for encouraging the initial uptake of DERs but are not desirable in the long term as they distort the market.

exception of aggregators, would possess revenue grade meters that would enable settlement and billing.

A more speculative application of decentralized markets is peer-to-peer (P2P) energy, where end customers are free to buy and sell both energy and services to utilities and one another. P2P receives the most press of all decentralized market applications but likely remains the furthest off in terms of technological feasibility and consumer acceptance. The most common design, which we will term 'active P2P', involves participants trading energy with one another in advance of delivery and places a significant burden on homes and DER owners to actively 'trade' their energy. It is indeed unclear whether every household wants to become an energy trader, but automation and trading bots may offer an answer to this problem. However, this does not resolve the need for significant communication and computational market architecture, in addition to how final settlement is handled. In most DR and aggregation programs, consumers are remunerated based on actions relative to a baseline, calculated from prior consumption patterns using a given methodology. It is unclear how such baselining could work in a market context, notwithstanding a host of issues around baselining methodologies in general. The best baseline is a baseline of zero with participants being charged for exactly what they have consumed or generated at a given time. This would, of course, mean that a participant's full energy usage would have to be transacted through such a P2P platform, although this would not preclude third parties from offering hedging services to those who want them, i.e. flat rate bills, much as utilities effectively do today for all consumers.

An alternative implementation, which we will term 'passive P2P', involves participants, beneath a single energy retailer or provider, consuming energy and operating their DERs as they see fit. After delivery, each participant has a net energy position, long or short, dependent on their usage. These positions are matched and netted off against one another based on prices and preferences submitted in advance by participants, and any outstanding energy balance is settled with the backstop energy provider. In this way, participants can prioritise locally generated energy or energy from renewable sources, without the administrative burden of active trading. The downside to this approach is that energy consumption behavior is not led in any direct way by price, rather there is a longer feedback loop between an energy bill, the prices and preferences the participant will set for the next pricing 'round' and then how they will behave. This approach represents a far simpler implementation than active P2P, since it is effectively an accounting problem, and could be feasibly implemented given existing smart meter and basic cloud technology.

A concept that often comes up in discussions around P2P energy is that of blockchain. A more detailed definition is given in Chapter 8, but in short, a blockchain is a distributed ledger that can securely execute and transactions between parties in an immutable and verifiable way without the need for a trusted third party. The most famous implementation of blockchain technology is the Bitcoin protocol, which as of November 2018 has a market capitalization of 78Bn USD. A more interesting evolution of Bitcoin is the Ethereum protocol which enables 'smart contracts', effectively computer code or logic, to be executed on the blockchain, turning the network into a vast distributed virtual machine. This has provoked

excitement for the potential of managing decentralized marketplaces and even the power system itself on blockchain, doing away with the need for utilities and system operators. This vision, while laudable, is fanciful for a number of reasons.

On a purely technical level, for what is gained in terms of cybersecurity and trustlessness using blockchain, a considerable cost is paid in terms of performance and speed relative to centralized databases and cloud systems. At the current maturity of the technology, it is a terrible idea to attempt to run any large-scale computation on blockchain. Additionally, the ability to store data on-chain is extremely limited, especially if the use case is large-scale time series datasets on energy production and consumption from smart meters. From an energy perspective, the power system has two inherent centralizations which make it unsuitable to be managed by a decentralized system. First, someone has to own, maintain, and manage the wires. Introducing blockchain does not remove the need for a system operator to monitor the grid and ensure stability and reliability. Second, each smart meter represents an oracle data source for the blockchain, meaning it is the sole source of truth for a participant's energy position. This presents two issues, how to verify the identity of a smart meter, and how to ensure the data coming from the smart meter is correct. The first issue can be dealt with via whitelisting; however, this requires an external regulator or verifier to approve smart-meters. Similarly, for the second issue, automatic data quality checks can be put in place, but ultimately a party outside of the blockchain has to guarantee the trustworthiness of a smart meter. The reliance on a system operator (at the least maintainer), and external validators defeat the premise of the removal of third parties. Just because a power system is decentralized, does not mean it has to be managed using a decentralized system such as blockchain.

All that said, blockchain is good at managing identity, atomic transactions of currency and data, and enforcing rules, which perhaps makes it suitable for managing the permissions of off-chain data stores and markets among untrusting parties. It is also a suitable use case for replacing the back-office and accounting functionality of utilities, automating billing and settlement processes. In general, for energy, blockchain should be thought of as a potential implementation layer for other applications with specific desirable and undesirable properties, in comparison to implementation layers such as Amazon Web Services. Blockchain is not a panacea in itself. In Chapter 8 we explore the potential of blockchain in serving as the coordinator for a distributed optimization to manage DERs in a microgrid. In reference to the earlier point, very little computation is handled on-chain.

As mentioned in the previous section, in Chapter 9 we address the design of an explicit flexibility market, aimed at distribution networks with the goal of increased integration of DERs. Flexible loads such as smart thermostats or EVs can specify their exact time-dependent availability, either as a shapeable load, or a deferrable load. The platform is also agnostic to the level in the hierarchy at which is implemented, coordinating transactions between microgrids as easily as transactions between homes.

A final note on decentralization is that the future seems to be heading in a far more market-driven direction for the provision of energy. The role of energy as a public good should not be lost in this discussion, and those who cannot afford to be flexible should not

be unduly left out or penalised as a result. Potential scenarios include residential customers with no flexibility becoming a higher risk segment for utilities and suppliers, leading to higher energy rates. Low-income customers who have benefited from subsidy schemes under regulated utilities may lose out in a more market-based system. On this point, any social program or redistribution, in deference to the 'right price', should ideally be handled outside of any market system rather than through the market system itself, as direct discounts or subsidies could cause market distortions.

Chapter 2

Mathematical Background

In this section we will review general notation, background theory, and relevant models which lay the groundwork for the rest of the dissertation.

2.1 General Notation

We denote by $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{R}_+$ the sets of real, complex, and nonnegative real numbers, respectively. We denote the transpose of a vector $x \in \mathbb{C}^n$ by $x^\top$, and its complex conjugate transpose by $x^\star$. Let $x_{-i} = (x_1, .., x_{i-1}, x_{i+1}, .., x_n) \in \mathbb{R}^{n-1}$ be the vector including all but the i^{th} element of x.

For any Euclidean vector space $\mathbb{R}^n$, we use $\mathbf{1} \in \mathbb{R}^n$ to denote the vector of ones, and $\mathbf{1}_k \in \mathbb{R}^n$ to denote the kth elementary vector, *i.e.* the vector with all zeros except for its kth element which is 1. We denote by $\mathbf{E} := \mathbf{1}\mathbf{1}^\top$ a square matrix of all ones.

For a matrix $x \in \mathbb{R}^{m\times n}$ with any given positive integers m and n, we use $x_{i,j}$ to denote its (i,j)th entry, $x_j := (x_{1,j}, \ldots, x_{m,j})^\top \in \mathbb{R}^{m\times 1}$ to denote its jth column, and $x_i^\top := (x_{i,1}, \ldots, x_{i,n}) \in \mathbb{R}^{1\times n}$ to denote its ith row. Denote by $\text{diag}(x_1, \ldots, x_n)$ the diagonal matrix with diagonal elements $\{x_i\}_{i=1}^n$.

For any real number z, we use $(z)^+ := \max(z, 0)$ to denote the positive part of z, and $(z)^- := -\min(z, 0)$ to denote the negative part of z such that $z = (z)^+ - (z)^-$.

Every effort has been made to maintain consistent notation of similar variables between chapters, however in some cases symbols have been reused and redefined. Any new definitions or usages are always defined in the text.

2.2 Power Flow Models

Modelling power flow is a crucial part of power systems analysis and optimization, forming the backbone of transmission expansion planning, contingency analysis, optimal power flow, and unit commitment, and will be required in all but one of the chapters of this dissertation. Power flow analysis can be used to determine the steady operating state of a power system,

where given certain known quantities, typically the amount of power generated and consumed at different locations, the remaining quantities of interest, such as the power flow on each line, can be calculated. Alternatively the power flow equations can be used to determine the *feasible injection region* of a given power system. System operators determine safe operating limits for the power network to avoid component damage and voltage collapse, usually in the form of capacity limits for transmission lines and operating ranges for nodal voltages. Additionally operators may further constrain the system using *contingency constraints* to ensure the network operating state is feasible in the event of a contingency such as a generator or component failure. Combining these limits with the power flow equations yields a feasible constraint set for the power system variables, which can then be used as the constraint set for an optimization problem. This latter application will be the focus in this dissertation.

Power flow is completely described by physical laws that are well understood, however there often exists no analytical solution due to the nonlinear and underdetermined nature of the full problem, and the non-convexity of its solution set. As such, a variety of approximations are frequently employed to make power flow more amenable to calculation and analysis. Each approximation uses a different set of assumptions that make them more or less appropriate to a given scenario.

AC Power Flow

The full AC power flow equations can be written as both a *branch-flow* model, involving only branch variables, and a *bus-injection* model, involving only nodal variables. Their equivalence has been shown in [25], and both have different properties that are useful in different contexts. Approximations of both models will be used throughout this dissertation, so both are included for completeness. In each case we consider a power network that is represented by a set of nodes connected by lines. The nodes are indexed by $i = 1, \ldots, n$, and the lines are indexed by $l = 1, \ldots, m$. The voltage of node 1 is assumed known and serves as a reference.[1] The complex nodal power injection is denoted as $s \in \mathbb{C}^n$. We assume that each line between buses i and j is a 3-phase AC line with balanced phases, meaning it can be modelled as a single line characterized by a complex impedance z_{ij}; we let its admittance be $y_{ij} := 1/z_{ij}$. The following is adapted from [25].

Branch-Flow Model

The power network is modeled by a connected *directed* graph with an arbitrary orientation, $\mathcal{G}(\mathcal{N}, \mathcal{E})$, where $\mathcal{N} := \{1, \ldots, n\}$, and $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$ such that if $(i, j) \in \mathcal{E}$ then $(j, i) \notin \mathcal{E}$, and (i, j) denotes a directed link from i to j. The network topology is described by the incidence

[1] The units in power systems are typically specified in the *per-unit system* where a base power and voltage are specified, and all quantities are specified relative to these base quantities. This means the reference voltage is typically equal to 1. The advantage of this system is that per-unit quantities do not change when referred from one side of a transformer to the other.

matrix $M \in \mathbb{R}^{n\times m}$ defined as

$$M_{il} = \begin{cases} 1, & \text{if link } l \in \mathcal{E} \text{ leaves node } i \\ -1, & \text{if link } l \in \mathcal{E} \text{ enters node } i \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

Let $\Delta_y := \text{diag}(y_l) \in \mathbb{R}^{m\times m}$, $l \in \mathcal{E}$. Let $V \in \mathbb{C}^n$ be the complex nodal voltage vector. For each $(i,j) \in \mathcal{E}$, let I_{ij} be the complex current and S_{ij} be the *sending-end* complex power from buses i to j. The branch flow model consists of the set of variables (S, I, V, s) and the following three sets of equations:

$$\text{Kirchoff's Law:} \quad I = \Delta_y M^\top V \tag{2.2a}$$

$$\text{Power Definition:} \quad S_{ij} = V_i I_{ij}^\star, \ (i,j) \in \mathcal{E} \tag{2.2b}$$

$$\text{Power Balance:} \quad s_i = \sum_{j:i\to j} S_{ij} - \sum_{k:k\to i} \left(S_{ki} - z_{ij}|I_{ij}|^2\right), \ i \in \mathcal{N} \tag{2.2c}$$

Eliminating phase angles from the complex voltages V and currents I, the branch-flow model can be more simply reformulated as

$$v_j = v_i - 2\text{Re}(z_{ij}^\star S_{ij}) + |z_{ij}|^2 l_{ij}, \qquad (i,j) \in \mathcal{E} \tag{2.3a}$$

$$l_{ij} = \frac{|S_{ij}|^2}{v_i}, \qquad (i,j) \in \mathcal{E} \tag{2.3b}$$

where $v_i := |V_i|^2$, and $l_{ij} := |I_{ij}|^2$.[2]

Bus-Injection Model

The power network is modeled by a connected *undirected* graph $\tilde{\mathcal{G}}(\mathcal{N}, \tilde{\mathcal{E}})$, where $\mathcal{N} := \{1, \ldots, n\}$, and and $\tilde{\mathcal{E}} \subseteq \mathcal{N}\times\mathcal{N}$, such that $(i,j) \in \tilde{\mathcal{E}}$ if and only if $(j,i) \in \tilde{\mathcal{E}}$. Let $M \in \mathbb{R}^{n\times m}$ be the network incidence matrix, with arbitrary orientation, as defined in (2.1). Let $Y \in \mathbb{C}^{n\times n}$ be the network admittance matrix, which can be represented as $Y = M\Delta_y M^\top$, where $\Delta_y := \text{diag}(y_l) \in \mathbb{R}^{m\times m}$, $l \in \mathcal{E}$. The network admittance matrix has the following structure

$$Y_{ij} = \begin{cases} \sum_{(i,k)\in\mathcal{E}} y_{ik}, & \text{if } i = j \\ -y_{ij}, & \text{if } i \neq j \text{ and } (i,j) \in \tilde{\mathcal{E}} \\ 0 & \text{otherwise} \end{cases} \tag{2.4}$$

Y is symmetric but not necessarily Hermitian, and represents the weighted graph Laplacian matrix of the power network. Let $\tilde{I} \in \mathbb{C}^n$, $\tilde{S} \in \mathbb{C}^n$ be the complex nodal current and power

[2]We abuse notation for l as the squared magnitude of the complex current here, to be consistent with the literature.

injection vectors respectively. The bus-injection model consists of the set of nodal variables $(\tilde{S}, \tilde{I}, V, s)$ and the following three sets of equations

$$\text{Kirchoff's Law:} \quad \tilde{I} = YV \tag{2.5a}$$

$$\text{Power Definition:} \quad \tilde{S}_i = V_i \tilde{I}_i^{\star}, \ i \in \mathcal{N} \tag{2.5b}$$

$$\text{Power Balance:} \quad s = \tilde{S} \tag{2.5c}$$

Substituting for $\tilde{S}$ and $\tilde{I}$, one can derive a more familiar form the AC power flow equations as

$$p_i = \sum_{j=1}^{n} |V_i||V_j| \left[g_{ij} \cos(\theta_i - \theta_j) + b_{ij} \sin(\theta_i - \theta_j) \right], \ i \in \mathcal{N} \tag{2.6}$$

$$q_i = \sum_{j=1}^{n} |V_i||V_j| \left[g_{ij} \sin(\theta_i - \theta_j) - b_{ij} \cos(\theta_i - \theta_j) \right], \ i \in \mathcal{N} \tag{2.7}$$

where $s_i = p_i + \mathbf{i} q_i$, and p_i, q_i denote the real and reactive components of the complex nodal power injection respectively, $y_{ij} = g_{ij} + \mathbf{i} b_{ij}$, $\forall (i,j) \in \tilde{\mathcal{E}}$, where g_{ij}, b_{ij} are the conductance and susceptance of the line respectively, and θ_i is the voltage phase angle at bus i, [26].

DC Power Flow

The DC power flow model is an approximation of the bus-injection model, most commonly used to model transmission networks. It makes the following simplifying assumptions:

- Transmission lines are lossless, $g_{ij} \simeq 0$, $(i,j) \in \tilde{\mathcal{E}}$.
- Voltage angle differences are small, $\sin(\theta_i - \theta_k) \simeq (\theta_i - \theta_k)$.
- Voltage magnitudes are close to 1, $|V_i| \simeq 1$, $i \in \mathcal{N}$.
- Real power flows are much larger than reactive power flows, $p_{ij} \gg q_{ij}$, $(i,j) \in \tilde{\mathcal{E}}$.

Applying these assumptions to equations 2.6 and 2.7, we obtain

$$p_i = \sum_{j=1}^{n} b_{ij}(\theta_i - \theta_j), \qquad i \in \mathcal{N} \tag{2.8}$$

$$q_i = 0, \qquad i \in \mathcal{N} \tag{2.9}$$

and it can also be shown that

$$p_{ij} = b_{ij}(\theta_i - \theta_j), \qquad (i,j) \in \mathcal{E} \tag{2.10}$$

$$q_{ij} = 0, \qquad (i,j) \in \mathcal{E} \tag{2.11}$$

We see that reactive power flow on the network is zero, and the real power flow on a line is proportional to the difference in voltage angle at the sending and receiving nodes. The DC

power flow assumptions are generally true of transmission networks, and this approximation is used widely in practice forming the basis of the power flow model for economic dispatch and unit commitment in Day-Ahead markets in the majority of North-American ISOs.

To specify the feasible injection region, we denote the real power flow capacity of lines in the network as $\hat{c} \in \mathbb{R}^m$. To express the feasible injection region compactly, it is also useful here to introduce the idea of the *shift-factor matrix* or *power transfer distribution factor matrix*, which relates line flows to nodal power injections, eliminating voltage angles. To form the shift factor matrix, we note that $\text{rank}(Y) = n-1$. Taking bus 1 to be the reference bus, we let $\overline{Y} \in \mathbb{R}^{(n-1)\times(n-1)}$ be the sub-matrix of Y which contains all entries of Y except its first row and column. Define a generalized inverse of Y to be $Y^\dagger := \begin{bmatrix} 0 & 0 \\ 0 & \overline{Y}^{-1} \end{bmatrix}$. We can then relate line flows $f \in \mathbb{R}^m$ to the nodal power injections $p \in \mathbb{R}^n$, using a linear map $\hat{H} \in \mathbb{R}^{m\times n}$

$$f = \hat{H}p, \quad \text{with } \hat{H} := \Delta_y M^\top Y^\dagger, \tag{2.12}$$

where $\hat{H}$ is the shift factor matrix. Let $H := \begin{bmatrix} \mathbf{I} \\ -\mathbf{I} \end{bmatrix} \hat{H} \in \mathbb{R}^{2m\times n}$ and $c := \begin{bmatrix} \hat{c} \\ \hat{c} \end{bmatrix} \in \mathbb{R}^{2m}$. We state the following properties of H

$$H\mathbf{1}_1 = 0, \quad \mathbf{1}^\top H = 0, \quad \text{rank}(H) = n-1 \tag{2.13}$$

and the columns of $H^\top$ are linearly independent of $\mathbf{1}$.

The feasible injection region, denoted $\mathcal{P}$, can be represented as

$$\mathcal{P} := \{p : \mathbf{1}^\top p = 0,\ Hp \le c\} \tag{2.14}$$

where the first equation enforces net power balance in the network, since power flows are assumed lossless, and the second equation limits the line flows induced by the net injection vector within transmission line capacities. The DC Power Flow equations will be used in Chapters 4, 5, 6, and 7.

Distflow

The Distflow equations represent a recursive reformulation of the branch-flow equations in (2.3) for radial networks, and were first proposed in [27, 28]. In a radial network each node in the network has a unique parent node, allowing for spatial recursion. We use notation from [29].

An additional node is introduced, indexed at 0, to be the feeder or root node of the tree, such that $\mathcal{N} = \{0, 1, \ldots, n\}$. Each line in $\mathcal{E}$ is denoted by the directed pair (i, j), where j is the node closer to the feeder, *i.e.* every line points towards node 0. We call j the parent of i, denoted by $\pi(i)$, and call i the child of j. Denote the child set of j as $\delta(j) := \{i : (i, j) \in \mathcal{E}\}$. Thus a link (i, j) can be denoted as $(i, \pi(i))$. Each line is now associated to its unique child node i, such that for each line $(i, \pi(i)) \in \mathcal{E}$, let $\mathrm{z}_i = \mathrm{r}_i + \mathbf{i}\mathrm{x}_i$ be the impedance of the line,

let I_i be the complex current flowing from node i to $\pi(i)$, and $S_i = P_i + \mathbf{i}Q_i$ be the complex power flowing from node i to $\pi(i)$. As defined previously, for each node $i \in \mathcal{N}$, let V_i be the complex voltage, and $s_i = p_i + \mathbf{i}q_i$ be the net complex power injection. The complex voltage V_0 at the root node is assumed given and fixed, and define $l_i := |I_i|^2$, $v_i := |V_i|^2$.

Reformulating (2.3) using this notation, the Distflow equations can then be written as

$$p_i = P_i - \sum_{k \in \delta(i)} (P_k - \mathrm{r}_i l_i), \qquad i = 0, \ldots, n \tag{2.15a}$$

$$q_i = Q_i - \sum_{k \in \delta(i)} (Q_k - \mathrm{x}_i l_i), \qquad i = 0, \ldots, n \tag{2.15b}$$

$$v_i = v_{\pi(i)} + 2(\mathrm{r}_i P_i + \mathrm{x}_i Q_i) - (\mathrm{r}_i^2 + \mathrm{x}_i^2) l_i, \qquad i = 1, \ldots, n \tag{2.15c}$$

$$l_i = \frac{P_i^2 + Q_i^2}{v_i}, \qquad i = 1, \ldots, n \tag{2.15d}$$

where $S_0 = 0 + \mathbf{i}0$ at the root node. Given (P, Q, l, v), phase angles can be uniquely determined for radial networks [30].

To specify the feasible injection region, we let $\overline{\mathrm{v}} \in \mathbb{R}^n$, $\underline{\mathrm{v}} \in \mathbb{R}^n$, be the upper and lower bounds on squared voltage magnitude, respectively, such that $\underline{\mathrm{v}} \leq v \leq \overline{\mathrm{v}}$.[3] We also assume real power flow constraints on the lines, $\hat{c} \in \mathbb{R}^m$, as in the DC power flow model, such that $-\hat{c} \leq P \leq \hat{c}$. This leaves us with the Distflow equations, however the final equation (2.15d) would give a non-convex set as it is a quadratic equality. This constraint is typically relaxed to an inequality, which yields a second-order cone constraint:

$$l_i \geq \frac{P_i^2 + Q_i^2}{v_i} \quad \Longleftrightarrow \quad \left\| \begin{matrix} 2P_i \\ 2Q_i \\ l_i - v_i \end{matrix} \right\|_2 \leq l_i + v_i \tag{2.16}$$

This gives the feasible injection region as

$$\mathcal{P} := \{P, Q, l, v : \underline{\mathrm{v}} \leq v \leq \overline{\mathrm{v}},\ -\hat{c} \leq P \leq \hat{c},\ (2.15\mathrm{c}),\ (2.16)\} \tag{2.17}$$

The Distflow equations are primarily applicable to distribution networks, which are typically radial in topology. For Optimal Power Flow (OPF) problems (the primary type of problem considered in this Dissertation), the SOCP relaxation in (2.16) is exact, *i.e.* the constraint holds with equality, if the upper voltage constraints are non-binding [31]. The Distflow equations will be used in Chapter 8.

Simplified/Linear Distflow

Real distribution network lines usually have very small r, x, *i.e.* $r, x \ll 1$, while $v \simeq 1$. Thus real and reactive power losses are typically much smaller than power flows P_i, Q_i. Following

[3] Distribution network power quality standards usually require the voltage at each node to be within $\pm 5\%$ of the nominal voltage.

this insight, higher order real and reactive power terms are neglected by setting $l_i = 0$, $\forall i$, to give this linear approximation to the Distflow equations, known as *Simplified Distflow*, again introduced in [27, 28].

$$p_i = P_i - \sum_{k \in \delta(i)} P_k, \qquad i = 0, \dots, n \tag{2.18a}$$

$$q_i = Q_i - \sum_{k \in \delta(i)} Q_k, \qquad i = 0, \dots, n \tag{2.18b}$$

$$v_i = v_{\pi(i)} + 2(\mathrm{r}_i P_i + \mathrm{x}_i Q_i), \qquad i = 1, \dots, n \tag{2.18c}$$

From (2.18) it can be derived that the voltages and net power injections satisfy the following equation

$$v_k = \sum_{i=1}^{n} R_{ki} p_i + \sum_{i=1}^{n} X_{ki} q_i + v_0, \ k = 1, \dots, n \tag{2.19}$$

where $R_{ki} := 2 \sum_{h \in \mathcal{W}_k \bigcap \mathcal{W}_i} r_h$, $X_{ki} := 2 \sum_{h \in \mathcal{W}_k \bigcap \mathcal{W}_i} x_h$, and $\mathcal{W}_i$ is the unique path from node 0 to node i, *i.e.* $\mathcal{W}_i := \{k : \ k \text{ is on the path from node 0 to node } i\}$.

To specify the feasible injection region of Simplified Distflow, we assume voltage and line capacity constraints as described and notated previously. It is also common to assume that reactive power injections, q, are fixed and known, defining $\hat{v}_k := \sum_{i=1}^{n} X_{ki} q_i + v_0$, such that

$$v_k = \sum_{i=1}^{n} R_{ki} p_i + \hat{v}_k, \ k = 1, \dots, n \tag{2.20}$$

The feasible injection region is then compactly expressed as

$$\mathcal{P} := \{p : \mathbf{1}^\top p = 0, \ \underline{\mathrm{v}} \leq Rp + \hat{v} \leq \overline{\mathrm{v}}, \ Hp \leq c\} \tag{2.21}$$

where $R \in \mathbb{R}^{n \times n}$ is defined appropriately from (2.20), and $H \in \mathbb{R}^{(n-1) \times n}$ is the shift factor matrix defined previously. Clearly the Simplified Distflow equations are similar to the DC Power Flow equations, enabling the inclusion of voltage constraints, but limiting their application to radial networks. The Simplified Distflow equations will be used in Chapter 9.

2.3 Optimal Power Flow, Economic Dispatch, and Electricity Market Models

The *Optimal Power Flow* (OPF) problem finds the optimal solution to an objective function subject to power flow and operational constraints, and is the basis for much of power systems operation and optimization [32]. It typically refers to the market clearing and dispatch problems solved by ISOs in day-ahead (DA) and real-time (RT) markets, but tranmission expansion planning and general network investment problems are also examples of OPF

problems. The most general formulation of OPF is as follows

$$\min_{p,z} f(p,z) \tag{2.22}$$
$$\text{s.t. } p \in \mathcal{P} \tag{2.23}$$
$$z \in \mathcal{Z} \tag{2.24}$$
$$g(p,z) \leq 0 \tag{2.25}$$

where p is the vector of power flow variables (real and reactive power injections or flows, currents, voltages), z is the vector of operational variables (generation dispatch, storage controls, system component settings), $\mathcal{P}$ is the feasible injection region of the power flow equations, as decribed in the previous section, $\mathcal{Z}$ is the feasible set of the operational variables, $f(p,z)$ is a general objective function (dispatch or social cost, power flow losses, investment cost), and $g(p,z)$ is a general constraint function that links power flow and operational variables.

There are many variants of OPF, generally named for the power flow approximation which they use, including ACOPF, DCOPF, Decoupled OPF, among others. In this dissertation we will be focused on the dispatch problem faced by ISOs which is sometimes referred to as *Security-Constrained Economic Dispatch* (SCED), an OPF variant which consists of minimizing the social cost of resource dispatch (maximizing social welfare) subject to power flow and operational constraints. Depending on the application, SCED can employ either AC or DC power flow models, and include varying levels of constraints and approximations thereof. From here on we will use the general term *Economic Dispatch* (ED) to refer to this family of SCED problems.[4] Before describing more specific formulations of the economic dispatch problem and the exact market structures they relate to, we detail the classes of market participants and how we model their preferences and operational constraints.

Market Participants

Here we address market participants that own physical assets, *i.e.* that intend to actually take or make delivery of energy in the market. Purely financial participants, such as virtual bidders, also exist but are only relevant to Chapter 3 in this dissertation, so we leave further discussion of this class of participants until then. We consider the costs and constraints of each participant in a general multi-period setting (not to be confused with a multi-stage setting) over T periods, with the set of periods denoted by $\mathcal{T} := \{1, \ldots, T\}$, and indexed by $t \in \mathcal{T}$.

Producers

Producers generate energy and inject it into the grid. This can be produced by thermal generation such as coal plants or natural gas turbines, or renewable generation such as solar

[4] In the literature, economic dispatch historically referred to the minimization of generation cost to meet a given demand, generally neglecting power flow constraints. In recent years however it has come to be a shorthand for the SCED problem

PV or wind turbines. We consider a generation vector $g_t \in \mathbb{R}^n$, with nodal generation $g_{i,t}$, $i = 1, \ldots, n$, representing the aggregate of all generation resources at that node in the tth period. The notion of generation firms and strategic behavior is only relevant to Chapters 4 and 5, so we leave further discussion of these ideas until then. For now, the generation at each node can be assumed to be independent and to behave competitively.

Generation is assumed to have variable and fixed costs of production, but faces no startup, shutdown or no-load costs, nor ramping constraints. The aggregate cost of generation in each period is measured by a non-decreasing convex function $C_t(g_t)$, $t \in \mathcal{T}$ taking the general form

$$C_t(g_t) = \frac{1}{2} g_t^\top Q_t g_t + a_t^\top g_t, \quad t \in \mathcal{T} \tag{2.26}$$

where $a_t \in \mathbb{R}^n$, $Q_t \in \mathbb{R}^{n \times n}$, $a_t \geq 0$, $Q_t \succeq 0$, $\forall t \in \mathcal{T}$. Q_t, $t \in \mathcal{T}$ is generally assumed to be diagonal, i.e. there is no *cross-cost* of generation. In the case of purely linear generation costs, Q_t is omitted. It should be noted that this cost function is the integral of the producer's inverse supply function, the latter being commonly used in microeconomic theory.

The feasible generation set is

$$\mathcal{G}_t = \{g_t \, : \, 0 \leq g_t \leq \overline{g}_t\}, \quad t \in \mathcal{T} \tag{2.27}$$

where $\overline{g}_t \in \mathbb{R}^n$, $t \in \mathcal{T}$ is the vector of upper generation limits.[5]

Load Serving Entities

Load Serving Entities (LSEs) represent end-customer loads which withdraw or consume energy from the grid. These can typically be thought of as utility companies or load aggregators who purchase energy on the wholesale market and then retail this energy to their customers. We consider a demand[6] vector $d_t \in \mathbb{R}^n$, $t \in \mathcal{T}$, with nodal demand $d_{i,t}$, $i = 1, \ldots, n$, representing the aggregate of all demand at that node in the tth period.

Demand can be modeled in two ways, either as elastic or inelastic, where elasticity refers to the sensitivity of demand to changes in price. Elastic demand is assumed to have an aggregate concave benefit function in each period, denoted $B_t(d_t)$, $t \in \mathcal{T}$ taking the general form

$$B_t(d_t) = -\frac{1}{2} d_t^\top M_t d_t + b_t^\top d_t + r_t, \quad t \in \mathcal{T} \tag{2.28}$$

where $M \in \mathbb{R}^{n \times n}$, $M \succeq 0$, $b_t \in \mathbb{R}^n$, and $r_t \in \mathbb{R}$. Similarly to above, M_t, $t \in \mathcal{T}$ is also generally assumed to be diagonal. It should be noted that this benefit function is the integral of the load's inverse demand function, the latter being commonly used in microeconomic theory. Inelastic demand is modeled as having some fixed demand profile, $\overline{d}_t \in \mathbb{R}^n$, $t \in \mathcal{T}$, and as

[5] One could also include lower generation limits $\underline{g}_t$, but we omit them here for simplicity and assume that $\underline{g}_t = 0$, for all generators.

[6] Load and demand are used interchangeably throughout the dissertation

willing to pay any price to have this demand satisfied. This preference is represented as a finite and constant benefit function $B_t(d_t) = r_t$, $r_t \leq \infty$, $t \in \mathcal{T}$.[7]

We model demand d_t, $t \in \mathcal{T}$ as a variable in the ISO's problem. For elastic load the feasible load set is

$$\mathcal{D}_t = \{d_t \ : \ d_t \geq 0\}, \quad t \in \mathcal{T} \tag{2.29}$$

which formalizes the general concept that demand cannot be negative.[8] For inelastic load the feasible load set is

$$\mathcal{D}_t = \{d_t \ : \ d_t = \overline{d}_t\} \quad t \in \mathcal{T} \tag{2.30}$$

Clearly rather than a variable, inelastic demand is a parameter of the ISO's problem, *i.e.* a constant. For convenience here, we maintain the formalism as if it were a variable, however in all future chapters involving inelastic demand it will be treated as a parameter. It should also be noted that inelastic demand can be treated as *net demand*, which is the nominal demand less any zero marginal cost renewable generation. Assuming renewables are must-take and uncontrollable, this is a convenient way to model these assets.

Energy Storage and Load Flexibility

We use *load flexibility* to refer to *temporal load flexibility*, in other words the ability of a load to shift its consumption in time[9]. We make the distinction from *demand response*, which usually refers to loads reducing their consumption for some period of time, either for economic reasons or in a system emergency.[10] Demand response can be seen as a subset of load flexibility, although here we generally assume that loads want to recover any foregone consumption. Flexible loads can also *increase* their consumption in a given period, which is a novel concept in the current paradigm of demand response.

We use *energy storage* to refer to assets that can withdraw energy from the grid at one time, store it, and then inject this energy back into the grid at a later time. This encompasses assets such as batteries and pumped hydro storage. Additionally, aggregations of flexible loads can be modeled and controlled as *virtual batteries* or *virtual power plants* [19, 20]. This enables a consistent mathematical modeling framework for both batteries and aggregations of flexible loads, with the difference that virtual batteries generally require some kind of

[7] Technically the benefit of inelastic loads is infinite, however in practice this is not true due to finite measures such as the value of lost load.

[8] If loads make use of storage then it is possible that some loads could be negative but we preclude this for now.

[9] Spatial load flexibility is also possible, for example deciding to charge electric vehicles at one location rather than another, or assigning computing jobs to one data center rather than another. However, this remains a far more limited application than temporal load flexibility so we do not address it further in this dissertation.

[10] Economic demand response is actually an example of elastic load, as described previously.

baseline, whereas real batteries do not. Load flexibility in the absence of baselines will be discussed in Chapters 8 and 9.

With regard to asset ownership, we assume that load flexibility can only be owned by LSEs. Energy storage can be owned either by producers, LSEs, or as a standalone asset which we refer to as *merchant storage*.

To model energy storage, we present the ideal storage model used in Chapter 6. A simpler model of pure load flexibility is used in Chapter 7, and a more realistic storage model involving charge/discharge efficiencies and charge leakage is used in Chapter 8. The properties and models of distinct classes of flexible load is presented in Chapter 9.

We adopt a similar storage model to that of [33] and [34]. For each bus i, the storage's state of charge (SOC) $x_i(t)$ evolves as

$$x_{i,t+1} = x_{i,t} - u_{i,t}, \quad t = 1, \dots, T-1 \tag{2.31}$$

where $u_{i,t}$ is the amount of energy discharged (if $u_{i,t} > 0$) or charged (if $u_{i,t} < 0$) in time period t, and the initial SOC is assumed to be $x_{i,t} = 0$. Each storage device has an energy capacity $s_i \geq 0$, such that

$$0 \leq x_{i,t} \leq s_i, \quad t \in \mathcal{T} \tag{2.32}$$

where $s_i = 0$ if there is no storage connected to bus i. Equations (2.31) and (2.32) can be compactly expressed in the following vector form

$$0 \leq Lu_i \leq s_i \mathbf{1}, \quad i = 1, \dots, n \tag{2.33}$$

where $u_i \in \mathbb{R}^T$ is the vector of storage controls over T periods at bus i, and $L \in \mathbb{R}^{T \times T}$ is a lower triangular matrix with entries $L_{ij} = -1$ for $i \geq j$. We denote the general set of feasible storage controls given a storage capacity s as

$$\mathcal{U}(s) := \{u \in \mathbb{R}^{n \times T} : 0 \leq Lu_i \leq s_i \mathbf{1}, \ \forall i\} \tag{2.34}$$

Market Mechanism and Formulation

Pricing, Social Welfare, and Participant Surplus

Before solving the ED problem, the ISO collects bids and offers from market participants to supply or consume a given amount of energy at a certain price. The exact bidding mechanism differs by ISO, but this can generally thought of as each participant specifying the parameters of its cost or benefit function in (2.26) and (2.28) respectively. Assuming that participants' bids reflect their true costs and benefits, *i.e.* they bid competitively, then if the ISO maximizes social welfare, subject to sytem and operational constraints, and adopts *Locational Marginal Pricing* (LMP), then the resulting dispatch and prices represent an *efficient market equilibrium*. This formulation, sometimes known as *bid-cost minimization* (BCM) with a *market-clearing price* (MCP) mechanism, is an extension of the basic concept of the equilibrium price and quantity obtained by intersecting supply and demand curves

from classical microeconomics, and was first introduced in [35]. Social welfare is defined as the sum of producer and consumer surplus, and it can be shown that maximizing this objective is equivalent to maximizing the consumer benefit function less the generation cost function as will be seen in (2.35a). In a sufficiently competitive market, the locational marginal pricing mechanism is designed to incentivize participants to bid truthfully, however participants may not reveal their true cost and benefit functions and behave strategically, marking up (or down) their bids or witholding capacity from the market in an effort to increase their profits or surplus. In this case the market outcome will deviate from the social optimum.

Assumptions

Single, two, and multi-stage markets are considered throughout the dissertation. Assumptions that apply to the whole dissertation can be categorized as:

- Energy only markets - no ancillary services or co-optimization
- Do not consider reserve provision or any kind of recourse
- No unit commitment, units assumed either to be pre-committed by the ISO, or generators internalize these commitment decisions and costs.

Multi-Period Economic Dispatch Formulation

The *economic dispatch problem* is solved by the ISO to determine the most efficient generation dispatch to meet an apparent demand, $d_t \in \mathbb{R}^n, t \in \mathcal{T}$, subject to network constraints. When there are storage devices connected to the network, energy can be moved across time periods, coupling T-single period economic dispatch problems. This results in the *multi-period economic dispatch problem* (*MPED*):

$$\min_{p,g,d,u} \sum_{t=1}^{T} C_t(g_t) - B_t(d_t) \tag{2.35a}$$

$$\text{s.t. } \gamma_t : \mathbf{1}^\top p = 0, \qquad t \in \mathcal{T} \tag{2.35b}$$

$$\beta_t : Hp \leq c, \qquad t \in \mathcal{T} \tag{2.35c}$$

$$\lambda_t : p_t = g_t + u_t - d_t, \qquad t \in \mathcal{T} \tag{2.35d}$$

$$\mu_i : Lu_i \leq s_i \mathbf{1}, \qquad i \in \mathcal{N} \tag{2.35e}$$

$$\nu_i : Lu_i \geq 0, \qquad i \in \mathcal{N} \tag{2.35f}$$

$$g_t \in \mathcal{G}_t, \qquad t \in \mathcal{T} \tag{2.35g}$$

$$d_t \in \mathcal{D}_t, \qquad t \in \mathcal{T} \tag{2.35h}$$

where, $g_t \in \mathbb{R}^n$ is the generation vector in each time period $t \in \mathcal{T}$, $J^\star(s,d)$ is the total optimal system cost, and $C_t(g_t)$, $B_t(d_t)$ are as defined above.

In general we will consider *MPED* as being parametrized by the storage vector, s, and the apparent demand, $d = \{d_t\}_{t\in\mathcal{T}}$, hence the notation for the optimal cost, $J^\star(s,d)$. We

also distinguish between notation for the storage capacity optimized by the ISO, s, and the apparent demand, d, and the storage capacity owned by market participants, and the nominal demand, denoted $\overline{s}$, $\overline{d}$, respectively.

Key Market Concepts

We now review the following useful market concepts related to economic dispatch:

Locational Marginal Prices The dual variables, $\lambda_t \in \mathbb{R}^n, t \in \mathcal{T}$ associated with the net injection constraint in (2.35d) are known as *Locational Marginal Prices* (LMPs). If bus i is a net demander, $\lambda_{t,i}$ equals the marginal benefit to consumers, and if bus i is a net supplier, $\lambda_{t,i}$ equals the marginal cost of generation. Since generation costs are assumed to be increasing functions, it must follow that $\lambda_{t,i} > 0$. The LMP can be decomposed into three components: a pure energy term, whose value is the same at each node, a loss term, and a congestion term [36]. Since the DC formulation is lossless, the loss term of the LMP does not appear in our analysis. It can be shown that the *Locational Marginal Prices* (LMPs) are equal to

$$\lambda_t^\star(s,d) = \gamma_t^\star(s,d)\mathbf{1} - H^\top \beta_t^\star(s,d), \quad t \in \mathcal{T} \tag{2.36}$$

Merchandising Surplus The *merchandising surplus* (MS) is defined as

$$MS = \sum_{t=1}^{T} \lambda_t^{\star\top}(s,d)(d_t - g_t^\star(s,d)) \tag{2.37}$$

This is the money that is left over after the system operator has paid all generators and collected all revenue from loads. It can be shown that this is always nonnegative for an economic dispatch [37]. The merchandising surplus will be discussed further in Chapter 6, where it will be shown that the MS is only non-zero in the presence of transmission or storage congestion.

2.4 Multiparametric Programming

In Chapters 4, 5, and 6, multiparametric programming will be used to examine equilibria among market participants and analytically characterize market outcomes under different regulatory frameworks for energy storage. The following theory is condensed from [38].

General Formulation of Multiparametric Program (MP)

We consider an optimization problem of the form

$$J^\star(x) = \inf_z \; J(z,x) \tag{2.38}$$

$$\text{s.t. } g(z,x) \leq 0 \; : \; u \tag{2.39}$$

where $z \in \mathcal{Z} \subseteq \mathbb{R}^s$ is the optimization vector, $x \in \mathcal{K} \subseteq \mathbb{R}^n$ is the parameter vector, $J : \mathbb{R}^s \times \mathbb{R}^n \to \mathbb{R}$ is the value function, $g : \mathbb{R}^s \times \mathbb{R}^n \to \mathbb{R}^m$ are the constraints, and $u \in \mathcal{U} \subseteq \mathbb{R}^m$ are the dual variables. We wish to study the behavior of the value function $J^\star(x)$, the optimizer $z^\star(x)$, and the dual variables $u^\star(x)$ as we vary x.

Results on general MPPs of this type require the concepts and theory of point-to-set mappings. We restrict our attention to two important classes of MP with linear constraints, mpLP and mpQP.

Multiparametric Programs with Linear Constraints

We consider the multiparametric program with linear constraints

$$J^\star(x) = \min_z \; J(z,x) \tag{2.40}$$

$$\text{s.t. } Gz \leq w + Sx \; : \; u \tag{2.41}$$

where $G \in \mathbb{R}^{m \times s}$, $w \in \mathbb{R}^m$, and $S \in \mathbb{R}^{m \times n}$. Given a closed and bounded polyhedral set $\mathcal{K} \subseteq \mathbb{R}^n$ of parameters, denote by $\mathcal{K}^\star \subseteq \mathcal{K}$ the region of parameters $x \in \mathcal{K}$ such that (2.41) is feasible

$$\mathcal{K}^\star = \{x \in \mathcal{K} \; : \; \exists z \text{ s.t. } Gz \leq w + Sx\} \tag{2.42}$$

If the domain of $J(z,x)$ is $\mathbb{R}^{s+n}$ then $\mathcal{K}^\star$ is a polytope, which can be proved by projection. i.e. $\mathcal{K}^\star$ is the intersection of the feasible parameter set $\mathcal{K}$ with the projection of the constraint set $\{x, z \; : \; Gz \leq w + Sx\}$ onto the parameter space. We assume that $\mathcal{K}$ is fully dimensional, and S has full column rank, otherwise a smaller set of parameters can be used.

We now define a critical region of the MPP. Let $\mathcal{J} = \{1, ..., m\}$ be the set of constraint indices. For any $A \subseteq \mathcal{J}$, G_A and S_A are the submatrices of G and S, with the rows indexed by A. G_j, S_j and w_j are the j-th row of G, S and w, respectively.

Definition 1 (Critical Region)**.** *We define CR_A as the set of parameters x for which the set A of constraints is active at the optimum.*

At x define:

$$A(x) = \{j \in \mathcal{J} \; : \; G_j z^\star(x) - S_j x - w_j = 0\} \tag{2.43}$$

$$NA(x) = \{j \in \mathcal{J} \; : \; G_j z^\star(x) - S_j x - w_j < 0\} \tag{2.44}$$

$(A(x), NA(x))$ are disjoint and their union is $\mathcal{J}$.[11] For a given $\overline{x} \in \mathcal{K}^\star$ let $(A, NA) = (A(\overline{x}), NA(\overline{x}))$, and let

$$CR_A = \{x \in \mathcal{K}^\star \; : \; A(x) = A\} \tag{2.45}$$

To find these critical regions, the general procedure is to begin with a single parameter vector $\overline{x} \in \mathcal{K}^\star$, solve the optimization problem and find the set of constraints $A(\overline{x})$ active

[11] This formulation assumes a unique optimizer $z^\star(x)$ for a given x. To extend to the case of multiple optimizers we have that $z^\star(x) \in \mathcal{Z}^\star(x)$.

at the optimum. This allows the calculation of CR_A from the KKT conditions of the optimization problem. The parameter space is then iteratively partitioned into critical regions. We note that the set of critical regions forms a *polyhedral partition* of the feasible parameter set, as defined in [38]. Each critical region is full-dimensional, and an open set as seen in (2.44). The boundaries of full-dimensional critical regions are in fact lower-dimensional of zero measure.

We now examine the specific cases of mpLP and mpQP.

mpLP

We consider the mpLP

$$J^\star(x) = \min_z \; c^\top z \tag{2.46}$$

$$\text{s.t. } Gz \leq w + Sx \; : \; u \tag{2.47}$$

We wish to compute the set of critical regions $\{CR_A\}_{A=1}^{n_{CR}}$, and closed form expressions for the value function, optimizers, and dual variables. We first form the Lagrangian

$$\mathcal{L}(z, x, u) = c^\top z + u^\top (Gz - Sx - w) \tag{2.48}$$

From the Lagrangian we derive the KKT conditions

$$c + G^\top u = 0 \tag{2.49}$$

$$u^\top (Gz - Sx - w) = 0 \tag{2.50}$$

$$Gz - Sx - w \leq 0 \tag{2.51}$$

$$u \geq 0 \tag{2.52}$$

We consider a point $x_0 \in \mathcal{K}^\star$. A possible choice for x_0 is the Chebyshev center of the polytope $\mathcal{K}$, which is the center of the largest Chebyshev ball contained within $\mathcal{K}$. Let $z_0^\star$ and $u_0^\star$ be the solutions of the Primal Problem (PP) and the dual variables, respectively, at x_0, which can be found using a standard LP solver. Determine $(A, NA) = (A(x_0), NA(x_0))$ for the PP at the optimum. From the primal feasibility conditions

$$G_A z_0^\star(x) = w_A + S_A x \tag{2.53}$$

$$G_{NA} z_0^\star(x) < w_{NA} + S_{NA} x \tag{2.54}$$

for all $x \in CR_A$.

We assume that G_A is square and invertible.[12] From (2.53), the optimizer $z_0^\star(x)$ is

$$z_0^\star(x) = G_A^{-1} S_A x + G_A^{-1} w_A = F_0 x + g_0 \tag{2.55}$$

[12]The general case where this assumption does not hold is addressed using QR decomposition in [38].

We note that $z_0^\star(x)$ is an affine function of x. From (2.54) and (2.55), the critical region CR_A is

$$CR_A = \{x \in \mathcal{K}^\star \,:\, (G_{NA}F_0 - S_{NA})x < w_{NA} - G_{NA}g_0\} \tag{2.56}$$

We note that for an LP, $u_0^\star(x) = u_0^\star$, $\forall x \in CR_A$. That is the vector of dual variables is a constant within each critical region. Then, using equations (2.49), (2.50), we can compute the value function as

$$\begin{aligned} J_0^\star(x) &= c^\top z_0^\star(x) && (2.57)\\ &= -u_0^{\star\top} G z_0^\star(x) && (2.58)\\ &= -u_0^{\star\top}(S_A x + w_A) && (2.59) \end{aligned}$$

To summarize these results, the critical region is an open polyhedral set, defined by strict inequalities. Within this critical region the value function and optimizers are affine functions of the parameter x, and the dual variables are constant.

Once we have found the critical region $\mathcal{CR}_0$, containing x_0, we wish to explore the parameter space and discover all other critical regions. Two prevailing methods exist for approaching this problem in a principled manner.

1. **Reversing Inequalities** This method involves partitioning the set $\mathcal{K}^\star \setminus \mathcal{CR}_0$, based on the constituent inequalities describing $\mathcal{CR}_0$. Each inequality of $\mathcal{CR}_0$ is sequentially reversed such that a new set $\mathcal{R}_i$ is generated for each inequality.[13] The reader is referred to Theorem 5.2 and Figure 5.10 of [38] for further details. This produces a strict polyhedral partition of $\mathcal{K}^\star \setminus \mathcal{CR}_0$, such that $\mathcal{K}^\star \setminus \mathcal{CR}_0 = \bigcup_i \mathcal{R}_i$. For each new set $\mathcal{R}_i$, the above procedure is repeated recursively. Generate a new point x_i using the Chebyshev center of $\mathcal{R}_i$, check if the problem is feasible in this set, and if so, discover a new critical region and partition the remainder of the set. The full algorithm is given in *Algorithm 7.1* of [38] This algorithm guarantees that the set $\mathcal{K}^\star$ is explored in a finite number of iterations, and that all critical regions are discovered. A problem with this algorithm is that critical regions can be artificially divided between multiple $\mathcal{R}_i$, implying that they can be 'discovered' multiple times. A simple solution to this is to keep track of critical regions already discovered.

2. **Crossing the Facets** For each facet of $\mathcal{CR}_0$ a point outside the region but close to the facet is selected and the above procedure is repeated. This guarantees that critical regions are computed in one piece, with no artificial splitting. However this requires heuristics to decide how far to step over the facet, and there is no guarantee that the whole of $\mathcal{K}^\star$ is covered. In practice however this method typically outperforms the reversing inequalities strategy.

Both of these algorithms provide a principled way to discover all critical regions of an mpLP problem.

[13] It should be noted that these $\mathcal{R}_i$ are not critical regions.

It is also relevant to discuss the propagation of the set of active constraints between critical regions. *i.e.* What happens to the set of active constraints when we cross from one critical region to another? Under no primal or dual degeneracy, the following hold:

i Full critical regions are described by a set of active constraints of dimension n.

ii Two neighbouring full-dimensional critical regions CR_{A_i} and CR_{A_j} have A_i and A_j differing only in one constraint. *i.e.* As one constraint becomes inactive, another new constraint becomes active.

iii CR_{A_i} and CR_{A_j} will share a facet which is a primal degenerate critical region CR_{A_p} of dimension $n-1$ with $A_p = A_i \cup A_j$.

iv Full dimensional critical regions are described by a number of constraints s.[14] In the case of primal degeneracy, critical regions may be described by more than s constraints, and vice versa for dual degeneracy.

Now that we have described how to discover all critical regions, we can describe and summarize the global properties of the mpLP solution. The proofs are detailed in [38].

Theorem 1. *Global Properties of the mpLP Solution*

i The feasible set $\mathcal{K}^\star$ is a polyhedron.

ii If the optimal solution $z^\star$ is unique $\forall x \in \mathcal{K}^\star$, the optimizer function $z^\star(x) : \mathcal{K}^\star \to \mathbb{R}^s$ is:

- *continuous*
- *polyhedral piecewise affine (PPWA) over $\mathcal{K}^\star$. In particular it is affine in each critical region $\mathcal{CR}_i$, and every $\mathcal{CR}_i$ is a polyhedron.*

Otherwise it is always possible to choose such a continuous and PPWA optimizer function $z^\star(x)$.

iii The value function $J^\star(x) : \mathcal{K}^\star \to \mathbb{R}$ is:

- *continuous*
- *convex*
- *PPWA over $\mathcal{K}^\star$, in particular affine in each $\mathcal{CR}_i$.*

iv The dual variable function $u^\star(x) : \mathcal{K}^\star \to \mathbb{R}^m$ is:

- *discontinuous*
- *PPWA over $\mathcal{K}^\star$, in particular constant in each $\mathcal{CR}_i$.*

[14] The polyhedral representation of critical regions will typically be described by more than s constraints, since they are intersected with the feasible parameter set $\mathcal{K}^\star$.

mpQP

We consider the mpQP

$$J^\star(x) = \min_z \ \frac{1}{2} z^\top H z \tag{2.60}$$

$$\text{s.t. } Gz \leq w + Sx \ : \ u \tag{2.61}$$

We assume that $H \succ 0$, to avoid primal degeneracy. It should be noted that quadratic cost functions of the form $\frac{1}{2} z^\top H z + x^\top F z$, can be transformed to the above form by the change of variables $\tilde{z} = z + H^{-1} F^\top x$.

We wish to compute the set of critical regions $\{CR_A\}_{A=1}^{n_{CR}}$, and closed form expressions for the value function, optimizers, and dual variables. The algorithm is conceptually identical to that for mpLP. We first form the Lagrangian

$$\mathcal{L}(z, x, u) = \frac{1}{2} z^\top H z + u^\top (Gz - Sx - w) \tag{2.62}$$

From the Lagrangian we derive the KKT conditions

$$Hz + G^\top u = 0 \tag{2.63}$$

$$u^\top (Gz - Sx - w) = 0 \tag{2.64}$$

$$Gz - Sx - w \leq 0 \tag{2.65}$$

$$u \geq 0 \tag{2.66}$$

We consider a point $x_0 \in \mathcal{K}^\star$. A possible choice for x_0 is the Chebyshev center of the polytope $\mathcal{K}$, which is the center of the largest Chebyshev ball contained within $\mathcal{K}$. Let $z_0^\star$ and $u_0^\star$ be the solutions of the Primal Problem (PP) and the dual variables, respectively, at x_0, which can be found using a standard QP solver. Determine $(A, NA) = (A(x_0), NA(x_0))$ for the PP at the optimum. From the primal feasibility conditions

$$G_A z_0^\star(x) = w_A + S_A x \tag{2.67}$$

$$G_{NA} z_0^\star(x) < w_{NA} + S_{NA} x \tag{2.68}$$

for all $x \in CR_A$.

From (2.63) we have that

$$z_0^\star = -H^{-1} G^\top u_0^\star \tag{2.69}$$

Substituting $z_0^\star$ into the complementary slackness condition (2.64)

$$u_0^{\star\top} (-G H^{-1} G^\top u_0^\star - w - S x_0) = 0 \tag{2.70}$$

Let $u_A^\star$ and $u_{NA}^\star$ be the dual variables corresponding to inactive and active constraints. For inactive constraints $u_{NA}^\star = 0$. For active constraints

$$(-G H^{-1} G^\top) u_0^\star - w - S x_0 = 0 \tag{2.71}$$

If the set of active constraints A is empty, then $u_0^\star = u_{NA}^\star = 0$, and therefore $z^\star = 0$, which implies that the critical region CR_A is

$$CR_A = \{x \ : \ Sx + w > 0\} \tag{2.72}$$

Otherwise we have the following. We assume that LICQ holds, *i.e.* the rows of G_A are linearly independent, implying that $(G_A H G_A^\top)$ is square, full rank, and thus invertible.[15] Then we have

$$u_A^\star = -(G_A H G_A^\top)^{-1}(w_A + S_A x) \tag{2.73}$$

We note that $u_A^\star$ is an affine function of x. Substituting $u_A\star$ into the equation for the optimizer we have that

$$z_0^\star = H^{-1} G_A^\top (G_A H G_A^\top)^{-1}(w_A + S_A x) \tag{2.74}$$

We note that $z_0^\star$ is also an affine function of x. $J^\star(x) = \frac{1}{2} z^\star(x)^\top H z^\star(x)$ and is therefore a quadratic function of x. The critical region CR_A is computed as the intersection of $\mathcal{P}_p$ and $\mathcal{P}_d$.

$$CR_A = \{x \ : \ x \in \mathcal{P}_p, x \in \mathcal{P}_d\} = \mathcal{P}_p \cap \mathcal{P}_d \tag{2.75}$$

where $\mathcal{P}_p$ is found by by substituting $z_0^\star$ into (2.65)

$$\mathcal{P}_p = \{x \ : \ G H^{-1} G_A^\top (G_A H G_A^\top)^{-1}(w_A + S_A x) < w + Sx\} \tag{2.76}$$

and $\mathcal{P}_d$ is found by substituting the dual variables into (2.66)

$$\mathcal{P}_d = \{x \ : \ -(G_A H G_A^\top)^{-1}(w_A + S_A x) \geq 0\} \tag{2.77}$$

To summarize these results, within the critical region the optimizers and dual variables are affine functions, and the value function is a quadratic function, of the parameter x. The process of computing the other critical regions after the first has been discovered is identical to that for mpLP, and an algorithm is given in *Algorithm 7.2* of [38].

In terms of the propagation of the set of active constraints between critical regions, for any two neighboring full-dimensional critical regions CR_{A_i} and CR_{A_j} we have $A_i \subset A_j$ and $|A_i| = |A_j| - 1$ or $A_j \subset A_i$ and $|A_i| = |A_j| + 1$. *i.e.* As one moves from one full-dimensional region to a neighboring full-dimensional region, one constraint is either added to the list of active constraints or removed from it.

We now describe and summarize the global properties of the mpQP solution, assuming $H \succ 0$.

Theorem 2. *Global Properties of the mpQP Solution*

i The feasible set $\mathcal{K}^\star$ is a polyhedron.

[15] Again, [38] has a full treatment for the case when LICQ does not hold involving QR decomposition.

ii The optimizer function $z^\star(x) : \mathcal{K}^\star \to \mathbb{R}^s$ *is:*

- *continuous*
- *polyhedral piecewise affine (PPWA) over* $\mathcal{K}^\star$*. In particular it is affine in each critical region* $\mathcal{CR}_i$*, and every* $\mathcal{CR}_i$ *is a polyhedron.*

iii The value function $J^\star(x) : \mathcal{K}^\star \to \mathbb{R}$ *is:*

- *continuous*
- *convex*
- *polyhedral piecewise quadratic (PPWQ) over* $\mathcal{K}^\star$*, in particular quadratic in each* $\mathcal{CR}_i$*.*

iv The dual variable function $u^\star(x) : \mathcal{K}^\star \to \mathbb{R}^m$ *is:*

- *continuous*
- *PPWA over* $\mathcal{K}^\star$*, in particular affine in each* $\mathcal{CR}_i$*.*

Chapter 3

Virtual Bidding - Equilibrium, Learning, and the Wisdom of Crowds

We begin by considering the role and impact of uncertainty in the context of multi-stage electricity markets, and ways in which this uncertainty can be mitigated or addressed. Specifically we examine the mechanism of *Virtual Bidding* (VB), and how a population of virtual bidders with heterogeneous beliefs at equilibrium approach the stochastically optimal market outcome. This chapter is the result of joint work with Prof. Eilyan Bitar and Prof. Kameshwar Poolla, and the majority of the results were first published in [39] by the author.

3.1 Introduction

In electricity markets, VB allows market participants to buy and sell electricity without the obligation to physically produce or consume it. This opens up market participation to financial entities or third parties without generation or load assets, allowing them to take advantage of arbitrage opportunities and promote market liquidity. VB is similar in nature to futures trading in more traditional commodity markets, where contracts are settled financially and no physical delivery takes place.

Deregulated electricity markets are typically characterized by centralized multi-settlement markets administered by an independent system operator (ISO). More specifically these markets have both a day-ahead (DA) and real-time (RT) market. In the DA market, the ISO collects demand bids and supply offers from participants and, based on the expected transmission network conditions, determines an economic unit commitment and dispatch with associated locational marginal prices (LMPs) for each hour of the next day. A similar economic dispatch procedure is conducted in the RT market, but in response to real-time system conditions, typically at five to fifteen minute intervals. The important distinction between the two markets is that cleared DA schedules are just financial contracts that can be settled at real-time prices, whereas the RT market represents physical delivery of energy *i.e.* no power flows in the DA market. It is this fact that enables the inclusion of VB that is not

backed by physical assets in electricity markets.[1] A more complete discussion of these issues can be found in [40].

A virtual bid in such a market structure is comprised of a buy (sell) bid in the DA market, matched by a sell (buy) offer in the RT market, such that any position taken up in the DA is completely liquidated in the RT market, with no obligation to physically produce or consume electricity. This allows virtual bidders to arbitrage the price difference between the DA and RT markets. This should in general cause the DA and RT prices to converge in expectation, as any price gap can be exploited by a risk neutral speculator. This is why VB is sometimes referred to as convergence bidding. It is also important here to highlight the difference between explicit and implicit VB. In the absence of an explicit VB mechanism, participants backed by physical assets can still make implicit virtual bids, for example bidding more capacity than they have available into the DA market and then purchasing the shortfall on the spot market in real time. Implicit VB can cause market power issues, and compromise the integrity of load and generation forecasts. Allowing a mechanism for explicit VB, as described above, goes some way to mitigating these issues. More broadly, whenever we discuss VB in this chapter, we are referring to explicit VB.

Prior Literature

The benefits of virtual bidding are discussed at length in [40–42], and are generally characterized as: improved liquidity, mitigation of market power, improved market efficiency and price formation, reduced price volatility, and providing market participants with the ability to hedge price risk. A potential downside of virtual bidding highlighted in the above works is the incentive for a virtual bidder in possession of a bilateral or external position to influence the profitability of this position through virtual trades. This is of particular relevance to those traders in possession of financial transmission rights (FTRs), as described in [43], to the effect that both ISO-NE and PJM enforce revenue capping when a participant makes a virtual bid which affects its own FTR revenue stream. [44] also suggest that virtual traders can exploit approximations in market designs to make profits without improving system operation, for example real-time ramping requirements that are not considered in the DA market. Some attempts have been made to quantify the efficiency effects of virtual bidding through empirical studies testing for the existence of profitable bidding strategies [45–48].

In this chapter we focus on the ability of virtual bidding to improve outcomes in electricity markets with uncertainty. [40] emphasizes this as one of the most valuable aspects of VB, yet also highlights the lack of rigorous work or analysis in this area, mainly due to the complexity involved. Modern electricity markets face increasing uncertainty in both supply and demand with a growing penetration of renewable and distributed generation. ISOs typically take a conservative approach to uncertainty, scheduling supply myopically in the DA market to meet expected demand, and neglecting the subsequent cost of recourse required to correct imbalances in the real-time (RT) market. They also hold significant reserve margins to

[1] VB is implemented in the majority of North-American ISOs, including PJM, NYISO, ISO-NE, MISO, and CAISO.

manage large deviations or deal with contingencies. This deterministic approach to power markets provides reliable and secure system operation, but it can be costly. Recent advances in stochastic and robust optimization have shown that significant cost reductions can be achieved by more explicitly incorporating uncertainty into market clearing algorithms [10, 12, 49]. Such approaches are tractable for real, large-scale, power systems; however, they face resistance from ISOs and system operators due to their perceived complexity, opacity, and reduction in system reliability.

Novel Contributions

We propose the novel thesis that, under certain assumptions, deterministic system operation with virtual bidding approximates the results of stochastic system operation, obviating the need for implementing new market algorithms. We demonstrate this result on a stylized model of a single bus, two-settlement electricity market. While a simple model, the results are instructive and point the way to models that more closely approximate the true operation of real power systems in future work. The model presented here is similar in nature to that proposed by [50], although the equilibrium analysis, welfare analysis, and learning dynamics presented here are novel. All of these analyses are shown to depend on the accuracy of the aggregate beliefs of the population of virtual bidders. In short, the wisdom of the crowd. The contributions are as follows:

- Characterization of the unique, pure strategy Nash equilibrium of a population of profit-maximizing virtual bidders with heterogeneous beliefs about the market in which they participate.
- We demonstrate that as the number of virtual bidders increases, the DA ISO schedule approaches the socially optimal schedule, and prices converge in expectation between the DA and RT markets.
- We investigate simple learning strategies for individual speculators and characterize conditions under which they converge to the unique Nash equilibrium.

Organization

The remainder of the chapter is organized as follows. In Section 3.2, we formulate a model of the two-settlement market and the virtual bidding mechanism. In Section 3.3 we characterize the pure Nash equilibrium among virtual bidders, and discuss its effect on social welfare. In Section 3.4 we propose simple learning dynamics under which virtual bidders reach the Nash equilibrium, and Section 3.5 concludes.

3.2 Market Model

We consider a simplified model of a two-settlement electricity market administered by an independent system operator (ISO) for a copper plate power system.[2] The electricity market is cleared in two stages: day-ahead (DA) and real-time (RT). In the DA market, the ISO must determine an initial dispatch of supply subject to uncertainty in the eventual realization of demand, which we assume to be perfectly inelastic and denote by $D \in \mathbb{R}_+$. We describe uncertainty in the ISO's prior belief about demand by modeling D as a random variable with mean $\mu := \mathbb{E}[D]$ and variance $\sigma^2 := \mathrm{Var}(D)$.

The ability to schedule supply in the DA market is essential, as certain generation resources (e.g., coal and nuclear) have limited ramping capability, and must therefore be scheduled well in advance of the required delivery time. We define the production cost in the DA market according to a convex quadratic function of the form

$$C_{\mathrm{DA}}(x) := \frac{1}{2}\alpha x^2 \tag{3.1}$$

for all production levels $x \geq 0$, and where $\alpha > 0$ is assumed to be fixed and known by the ISO.

In the RT market, demand is realized, and any mismatch between supply scheduled in the DA market, say x, and the realized demand D must be compensated through an adjustment of supply in the amount of $D - x$. The subsequent balancing cost incurred in the RT market is assumed to be a convex quadratic function of the form

$$C_{\mathrm{RT}}(D - x) := \frac{1}{2}\beta(D - x)^2 + \gamma(D - x), \tag{3.2}$$

where $\beta > 0, \gamma$, are assumed fixed and known by the ISO.[3] The inclusion of the affine term in the RT cost is an approximation of the fact that in reality the DA and RT cost functions will be coupled. Fast-ramping generators that have not been dispatched in the DA market may bid their spare capacity into the RT market.γ may be interpreted as the minimum marginal cost of fast-ramping generators in the RT market, such that $\gamma = \alpha\overline{x}$, where $\overline{x}$ is the total capacity available in the DA market.

We define the total expected cost of supply incurred under a DA schedule $x \geq 0$ as

$$J(x) := C_{\mathrm{DA}}(x) + \mathbb{E}[C_{\mathrm{RT}}(D - x)]. \tag{3.3}$$

Finally, the price at which energy is traded in each of the DA and RT markets is set by the ISO according to the marginal cost of supply in each market. Accordingly, given a DA dispatch of supply in the amount of $x \geq 0$, the DA and RT prices of energy are determined according to

$$P_{\mathrm{DA}}(x) := \alpha x \quad \text{and} \quad P_{\mathrm{RT}}(D - x) := \beta(D - x) + \gamma, \tag{3.4}$$

[2] We use the term *copper plate* here to imply a lossless, unconstrained transmission system.

[3] No assumption is made on the relative values of α and β, although generally in practice $\beta > \alpha$, reflecting the fact that it is more expensive to procure power in real-time than schedule it forward.

respectively, and the RT-DA price spread is defined as

$$\Delta(x) := P_{\mathrm{RT}}(D - x) - P_{\mathrm{DA}}(x). \tag{3.5}$$

Naturally, a priori uncertainty in demand will manifest itself as uncertainty in the RT price.

Remark 1. *Implicit in our assumption of quadratic cost functions, in both the DA and RT markets, is the assumption that the underlying aggregate supply function in each market is linear and affine respectively. This is a common assumption in the power system economics literature, see for example [51]. Throughout this chapter, we interpret these supply functions as representing the true marginal cost of generation in each market. The treatment of more sophisticated models, which capture the effect of generator strategic behavior on the determination of these supply functions (in combination with strategic virtual bidding) represents an interesting and open direction for future research.*

Conventional Market Clearing

The approach to market clearing practiced by the majority of North-American ISOs today is inherently myopic in nature. That is to say, the ISO schedules supply in the DA market to minimize the immediate system cost based on a point estimate (forecast) of demand, which we denote by $\widehat{D}$. In doing so, the ISO neglects the subsequent cost of recourse required to compensate imbalances that might arise between supply scheduled in the DA market and realized demand. Needless to say, the cost incurred by a myopic approach to scheduling such as this may far exceed the minimum expected cost of supply, which we formally define as

$$J(x^\star) := \min\{J(x) : x \in \mathbb{R}_+\}. \tag{3.6}$$

A straightforward calculation shows the optimal DA schedule to satisfy[4]

$$x^\star := \arg\min\{J(x) : x \in \mathbb{R}_+\} = \frac{\beta\mu + \gamma}{\alpha + \beta}. \tag{3.7}$$

This optimal DA schedule results in an ex-ante no-arbitrage condition, such that

$$P_{\mathrm{DA}}(x^\star) = \mathbb{E}[P_{\mathrm{RT}}(D - x^\star)]. \tag{3.8}$$

This is equivalent to stating that the expected price spread is equal to zero, $\mathbb{E}[\Delta(x^\star)] = 0$. Myopic scheduling on the part of the ISO will result in a non-zero price spread in expectation

$$\mathbb{E}[\Delta(\widehat{D})] = (\alpha + \beta)\left(x^\star - \widehat{D}\right), \tag{3.9}$$

which can be exploited by speculators for profit. In what follows, we investigate the extent to which the speculative behavior of *virtual bidders* might drive the procurement of supply in the DA market towards the optimal procurement level $x^\star$.

[4]Finding this solution in the more general network case with constraints amounts to solving a two-stage stochastic optimization problem.

Virtual Bidding

Consider a two-settlement electricity market in which a set of virtual bidders, $\mathcal{N} = \{1, \ldots, N\}$, participate. We assume that each virtual bidder is risk-neutral and seeks to maximize the expected profit they derive through price arbitrage between the DA and RT markets. Moreover, we assume that all virtual bids are quantity bids[5], such that the total supply x scheduled by the ISO in the DA market takes the form

$$x = \widehat{D} + \sum_{i=1}^{N} v_i, \tag{3.10}$$

where $v_i \in \mathbb{R}$ denotes the quantity bid of the ith virtual bidder. We adopt the sign convention that $v_i > 0$ ($v_i < 0$) corresponds to a demand bid (supply offer) in the DA market. We denote by $v = (v_1, \ldots, v_N)$ the virtual bid profile, and by $V := \sum_{i=1}^{N} v_i$ the aggregate virtual bid. It follows that the DA and RT prices induced under a virtual bid profile v are given by

$$P_{\text{DA}}(x) = \alpha(\widehat{D} + V), \tag{3.11}$$

$$P_{\text{RT}}(D - x) = \beta(D - \widehat{D} - V) + \gamma, \tag{3.12}$$

respectively, and the RT-DA price spread is equal to

$$\Delta(\widehat{D} + V) = (\alpha + \beta)\left(\frac{\beta D + \gamma}{\alpha + \beta} - \left(\widehat{D} + V\right)\right) \tag{3.13}$$

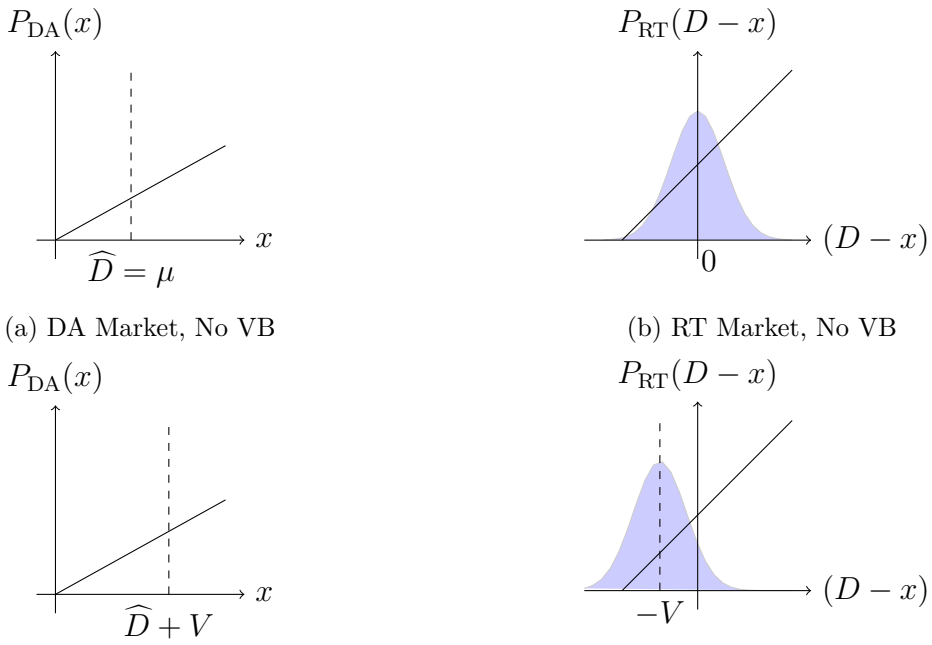
These price functions are illustrated in Figure 3.1. In Figure 3.1a we see that the ISO schedules supply myopically to meet expected demand. In Figure 3.1b, the RT price is determined by the realization of demand. In Figure 3.1c the DA schedule is adjusted due to virtual bidding. In Figure 3.1d, we see that the RT price is still determined by the realization of demand, but is impacted by the virtual bids. In the model we consider, we allow for asymmetry in the beliefs held by individual virtual bidders regarding the market in which they participate. Namely, we assign to each virtual bidder $i \in \mathcal{N}$ a belief defined according to the tuple $(\alpha_i, \beta_i, \gamma_i, \mu_i)$, representing what virtual bidder i believes the DA and RT cost coefficients and mean value of demand to be.[6] The expected payoff to virtual bidder $i \in \mathcal{N}$, under a virtual bid profile v, is therefore defined according to

$$\begin{aligned} \pi_i(v_i, v_{-i}) :&= \mathbb{E}\left[\Delta_i(\widehat{D} + V)v_i\right], \\ &= (\alpha_i + \beta_i)\left(x_i^\star - \left(\widehat{D} + V\right)\right)v_i \end{aligned} \tag{3.14}$$

where $\Delta_i(\cdot)$ is the RT-DA price spread calculated using the beliefs of virtual bidder i, and $x_i^\star := (\mu_i\beta_i + \gamma_i)/(\alpha_i + \beta_i)$ represents the implicit estimate of the optimal DA schedule $x^\star$ by

[5] In practice, virtual bids allow for the specification of both price and quantity, thereby allowing virtual bidders to reveal their willingness to pay (accept) in addition to their quantity bid (offer).

[6] For now it is assumed that the ISO forecast of demand $\widehat{D}$ is common knowledge, although this will not be necessary for the learning dynamics presented in Section 3.4.



(a) DA Market, No VB

(b) RT Market, No VB

(c) DA Market, With VB

(d) RT Market, With VB

Figure 3.1: DA and RT Markets, with and without virtual bidding

virtual bidder i. The collection of payoffs $\pi = (\pi_1, \ldots, \pi_N)$ together give rise to a normal-form (Cournot) game between the virtual bidders, which we refer to as the *virtual bidding game*. We define its equilibrium as follows.

Definition 2 (Nash equilibrium)**.** *The bid profile $v \in \mathbb{R}^N$ defines a* pure strategy Nash equilibrium *of the virtual bidding game if for each $i \in \mathcal{N}$, it holds that*

$$\pi_i(v_i, v_{-i}) \geq \pi_i(\overline{v}_i, v_{-i}) \quad \textit{for all} \quad \overline{v}_i \in \mathbb{R}. \tag{3.15}$$

3.3 Equilibrium Analysis

We proceed with an explicit characterization and analysis of the equilibrium of the virtual bidding game defined in Section 3.2. Before proceeding, it will be convenient to measure the quality of the belief that each virtual bidder $i \in \mathcal{N}$ holds about the market in which they participate according to the quantity

$$\eta_i := x_i^\star / x^\star, \tag{3.16}$$

Naturally, the closer η_i is to one, the more accurate is the belief held by virtual bidder i. We say that virtual bidder i has perfect belief if $\eta_i = 1$. We define the market belief profile according to the vector $\eta := (\eta_1, \ldots, \eta_N)$. With this notation in hand, we present the following characterization of the equilibrium of the virtual bidding game.

Theorem 3. *The virtual bidding game admits a unique pure strategy Nash equilibrium* $v^\star \in \mathbb{R}^N$ *satisfying*

$$v_i^\star = \left(\eta_i - \frac{\sum_{j=1}^N \eta_j}{N+1}\right) x^\star - \left(\frac{1}{N+1}\right) \widehat{D}, \tag{3.17}$$

for each $i \in \mathcal{N}$.

Proof. We begin by considering the payoff-maximizing action $v_i^\star$ of virtual bidder i, given the actions of all other virtual bidders v_{-i}.

$$v_i^\star = \arg\max\{\pi_i(v_i, v_{-i}) : v_i \in \mathbb{R}\} \tag{3.18}$$

Since $\pi_i(v_i, v_{-i})$ is a strongly concave function of v_i, we have that $\nabla \pi_i(v_i^\star, v_{-i}) = 0$ is a neccesary and sufficient condition for optimality. Solving we find that

$$v_i^\star = \left(x_i^\star - \left(\widehat{D} + \sum_{j \neq i}^N v_j + v_i^\star\right)\right) \tag{3.19}$$

We now assume that an equilibrium $v^\star = (v_1^\star, \ldots, v_N^\star)$ exists, and will show that this is indeed the case. Summing over $v_j^\star$, we see that

$$\begin{aligned} \sum_{j=1}^N v_j^\star &= \left(x^\star \sum_{j=1}^N \eta_j - N\widehat{D} - N\sum_{i=1}^N v_j^\star\right) \\ &= x^\star \frac{\sum_{i=1}^N \eta_i}{N+1} - \frac{N}{N+1}\widehat{D} \end{aligned} \tag{3.20}$$

Substituting (3.20) into (3.19), we see that

$$v_i^\star = \left(\eta_i - \frac{\sum_{j=1}^N \eta_j}{N+1}\right) x^\star - \left(\frac{1}{N+1}\right) \widehat{D} \tag{3.21}$$

Since $\nabla \pi_i(v_i^\star, v_{-i}^\star) = 0$, $\forall i$, this is an equilibrium of the virtual bidding game, and is unique due to the strong concavity of the payoff function. □

It is immediate to see that under perfect beliefs (i.e., $\eta_i = 1$ for all $i \in \mathcal{N}$), this unique pure strategy Nash equilibrium $v^\star$ is symmetric, and reduces to

$$v_i^\star = \frac{x^\star - \widehat{D}}{N+1}, \tag{3.22}$$

for all $i \in \mathcal{N}$. All further discussion in this Section refers to the unique pure strategy Nash equilibrium under heterogeneous beliefs in (3.17). We see that the equilibrium action of each virtual bidder is to fill some fraction of the quantity gap between the optimal DA schedule and the myopic ISO schedule. This equilibrium action of each virtual bidder is a function of the quality of their own belief, and the average quality of the beliefs of all virtual bidders.

The Wisdom of Crowds

We wish to consider the effect of virtual bidding on the physical DA schedule at equilibrium, which is given by

$$x_N := \widehat{D} + \sum_{i=1}^{N} v_i^\star. \tag{3.23}$$

It is first useful to characterize the belief quality of the 'crowd' of virtual bidders. One way to model this is to assume that the individual beliefs of virtual bidders are drawn in an independent and identically distributed (IID) fashion from a common probability distribution. That is to say, we model the belief profile $\eta = (\eta_1, \ldots, \eta_N)$ as a collection of IID random variables having mean and variance

$$\mu_\eta := \mathbb{E}[\eta_i] \quad \text{and} \quad \sigma_\eta^2 := \mathrm{Var}(\eta_i),$$

for all $i \in \mathcal{N}$. In addition, we assume the belief profile η to be independent of the demand D.[7]

Definition 3 (Wisdom of the crowd)**.** *We define the crowd of virtual bidders to be* wise *if* $\mu_\eta = 1$*, i.e., their belief is correct on average.*

It is straightforward to show that the DA schedule and price spread which emerge at equilibrium satisfy

$$x_N = \left(\frac{1}{N+1}\sum_{i=1}^{N}\eta_i\right)x^\star + \left(\frac{1}{N+1}\right)\widehat{D} \tag{3.24}$$

and

$$\Delta(x_N) = (\alpha+\beta)\left(\frac{\beta D+\gamma}{\alpha+\beta} - x_N\right). \tag{3.25}$$

We have the following Corollary to Theorem 3, which characterizes their asymptotic values as the the number of virtual bidders grows large.

Corollary 1 (Asympotic Market Efficiency)**.** *Assume that the virtual bidders collectively behave according to the Nash equilibrium* (3.17)*. As the number of virtual bidders participating in the market grows large, it holds that*

$$\lim_{N\to\infty} x_N = \mu_\eta x^\star \tag{3.26}$$

and

$$\lim_{N\to\infty} \mathbb{E}\left[\Delta(x_N)\right] = (\alpha+\beta)(1-\mu_\eta)x^\star. \tag{3.27}$$

[7]This assumption may be strong, as it is not unreasonable to expect that the quality of private estimates and demand may be correlated in some fashion.

Namely, if the crowd is wise (i.e., $\mu_\eta = 1$), the DA schedule, which emerges at the Nash equilibrium, converges to the optimal DA schedule as the number of virtual bidders tends to infinity. As a result, the expected price spread between the RT and DA markets also converges to zero. Such asymptotic market behavior is to be expected, as a large number of virtual bidders will naturally compete away any ex-ante arbitrage opportunity.

We draw the following conclusions from this result, assuming that the desired outcome of implementing virtual bidding is improved market efficiency. First, it is important to have a crowd. The larger the number of virtual bidders, the smaller the expected arbitrage opportunity available to each bidder at equilibrium, and the closer one gets to the optimal DA schedule at equilibrium. In reality, the number of participants is likely to be determined by transaction costs associated with virtual bidding, and the risk premia that risk-seeking or risk-averse virtual bidders will demand or are willing to pay. It is not in the interests of an ISO to restrict access to virtual bidding markets in any way, for example through uplift payments associated with virtual bids as described in [40].

Second, it is important that the crowd is wise. This is not something that can be prescribed per se, but is a phenomenon that has been observed in many contexts. From estimating the weight of an ox, [52], to modern day prediction markets, [53], the crowd average generally outperforms individual estimates. One might also surmise that if participants have skin in the game, they are more likely to be invested in the quality of their own estimate, thus improving the crowd estimate.

Third, Corollary 1 holds for an arbitrary ISO forecast of demand $\widehat{D}$. Of course, the closer $\widehat{D}$ is to $x^\star$, the closer x_N will be to $x^\star$. It remains to be seen whether an aggregate crowd estimate of $x^\star$, could outperform one generated by a central ISO.

Given our distributional interpretation of beliefs, it is also possible to explicitly characterize the variance of the price spread, which results at the Nash equilibrium. Recalling that $\mathrm{Var}(D) = \sigma^2$, the spread variance in the absence of virtual bidding is easily calculated as

$$\mathrm{Var}(\Delta(\widehat{D})) = \beta^2\sigma^2. \tag{3.28}$$

In the presence of virtual bidders, we have

Corollary 2. *The spread variance at the virtual bidding Nash equilibrium is equal to*

$$\mathit{Var}(\Delta(x_N)) = \beta^2\sigma^2 + (\alpha+\beta)\left(\kappa^2\frac{\sigma_\eta^2}{N}(x^\star)^2\right). \tag{3.29}$$

where $\kappa := \frac{N}{N+1}$ is a nondimensional parameter measuring the size of the crowd.

It follows directly that

$$\mathrm{Var}(\Delta(x_N)) \geq \mathrm{Var}(\Delta(\widehat{D})) \tag{3.30}$$

for any number of virtual bidders N. It can also be seen that as the number of virtual bidders grows large, it holds that

$$\lim_{N\to\infty} \mathrm{Var}(\Delta(x_N)) = \mathrm{Var}(\Delta(\widehat{D})) \tag{3.31}$$

This result is independent of the wisdom of the crowd, and states that at equilibrium the variance of the price spread under virtual bidding is lower bounded by the variance of the price spread under the myopic ISO schedule. Under these assumptions, the spread variance *never* decreases after the introduction of virtual bidding. This is due to the underlying variance in demand, that is not addressed at all by virtual bidding. Additionally, a large variance in the distribution of beliefs among virtual bidders only serves to worsen the variance of the spread, although this effect is mitigated as the number of virtual bidders increases.

This theoretical result would seem to be at odds with empirical results presented by [48], which demonstrate that spread variances decreased after the introduction of virtual bidding in the CAISO market. However, it should be noted that they attribute this reduction in variance to the reduction of implicit virtual bidding by physical assets, and the fact that DA physical generation schedules should be closer to their real-time outputs under explicit virtual bidding, thereby reducing the need for costly purchases by the ISO to account for deviations in real-time. In our analysis we have not considered implicit virtual bidding by physical participants, but this would present an interesting avenue for further study. One could conjecture that under implicit virtual bidding the spread variance might increase due to both the false reporting of true physical production schedules, and heterogeneity of beliefs among implicit virtual bidders.

Welfare Analysis

We now investigate the social welfare properties of the virtual bidding Nash equilibrium. As demand is assumed to be inelastic, social welfare is naturally defined according to the expected cost of generation $J(x)$, which we previously defined in (3.3). To simplify the analysis we assume that a myopic ISO takes as its demand forecast $\widehat{D} = \mu$, although the results hold for arbitrary forecasts $\widehat{D}$.

We first see that under a myopic ISO dispatch, $\widehat{D} = \mu$, in the absence of virtual bidding, the generation cost takes the form

$$J(\mu) = \frac{1}{2}\left(\alpha\mu^2 + \beta\sigma^2\right). \tag{3.32}$$

If the ISO adopts the socially optimal dispatch $x^\star$, then the generation cost is

$$J(x^\star) = J(\mu) - \frac{1}{2}\frac{(\gamma - \alpha\mu)^2}{(\alpha + \beta)}. \tag{3.33}$$

As expected, we see that $J(x^\star) \leq J(\mu)$. We note the following identity

$$x^\star - \mu = \frac{\gamma - \alpha\mu}{\alpha + \beta}, \tag{3.34}$$

such that

$$J(x^\star) = J(\mu) - \frac{1}{2}(\alpha + \beta)(x^\star - \mu)^2. \tag{3.35}$$

We now consider the generation cost at the equilibrium of virtual bidders. Since we assume that the individual beliefs of virtual bidders are drawn in an IID fashion from a common probability distribution we have that

$$J(x_N) = \mathbb{E}_\eta \left[C_{\mathrm{DA}}(x_N) + \mathbb{E}_D \left[C_{\mathrm{RT}}(D - x_N) \right] \right] \tag{3.36}$$

where we must take expectations with respect to the random belief profile $\eta = (\eta_1, \ldots, \eta_N)$. Assuming that the crowd is wise (i.e., $\mu_\eta = 1$), it can be shown that

$$J(x_N) = \frac{1}{2}\left(\alpha\mu^2 + \beta\sigma^2\right) \tag{3.37}$$

$$+ \frac{1}{2}(\alpha+\beta)\left(\kappa^2 \frac{\sigma_\eta^2}{N}(x^\star)^2 + \kappa(\kappa-2)(x^\star-\mu)^2\right), \tag{3.38}$$

We see that $J(x_{N=0}) = J(\mu)$, and that $J(x_{N\to\infty}) = J(x^\star)$, as expected. In general, however, virtual bidding may actually increase the generation cost due to the positive contribution from the variance of virtual bidders' beliefs. This can generally be characterized as occuring for a low number of virtual bidders, with high variance in beliefs. For a given variance, this positive term will be offset by a sufficiently large population of virtual bidders, since the final term is a strictly decreasing function of N, on the interval $N \in [0, \infty)$. We can in fact explicitly characterize the number of bidders after which the generation cost is strictly decreasing and less than or equal to the cost under a myopic ISO dispatch, denoted N_{dec}. For a fixed set of parameters $(\alpha, \beta, \gamma, \mu, \sigma_\eta)$, we have

$$N_{\mathrm{dec}} = \sigma_\eta^2 \frac{(x^\star)^2}{(x^\star-\mu)^2} - 2. \tag{3.39}$$

For the generation cost to be strictly decreasing for all $N \geq 1$, we require that

$$\sigma_\eta^2 \leq \frac{3(x^\star-\mu)^2}{(x^\star)^2} \tag{3.40}$$

At equilibrium, for $N \geq N_{\mathrm{dec}}$, virtual bidders never profit at the expense of loads. The expected cost of generation decreases in the presence of virtual bidders, and the marginal cost reduction associated with the addition of a new virtual bidder is always positive.

The exact impact of virtual bidders on social welfare will be a function of the specific market parameters, however these results highlight again the importance of a *crowd* of virtual bidders. Even the effect of a high variance in beliefs can be mitigated by the presence of a large number of virtual bidders.

3.4 Reaching Equilibrium

While the above results hold at the unique Nash equilibrium of the virtual bidders, actually reaching this equilibrium is a more subtle question. We consider simple learning dynamics

for each virtual bidder, assuming that the two-settlement market is a repeated game in a homogeneous environment. In practice this might represent one hour of a day across many weeks, assuming similar patterns of weather and demand.

The best response of virtual bidder i is defined as

$$v_i^{BR} := \arg\max\{\pi_i(v_i, v_{-i}) : v_i \in \mathbb{R}\} \tag{3.41}$$

assuming that the actions of the other virtual bidders v_{-i} are given. It can be shown that this is equal to

$$v_i^{BR} = \frac{1}{2}\left(x_i^\star - \left(\widehat{D} + V_{-i}\right)\right) \tag{3.42}$$

where $V_{-i} = \sum_{j \neq i} v_j$. At equilibrium v_i^{BR} is equivalent to $v_i^\star$ in (3.17).

We consider a smoothed best-response learning dynamic, where at each iteration virtual bidder i plays a weighted sum of their previous action and their best response to the previous actions of all other bidders, with smoothing parameter θ_i, where $0 \leq \theta_i \leq 1$. See [54, 55]. The learning dynamic then takes the form

$$\begin{aligned} v_i(k+1) &= \theta_i v_i(k) + (1-\theta_i) v_i^{BR}(k) \\ &= \theta_i v_i(k) + \frac{(1-\theta_i)}{2}\left(x_i^\star - \left(\widehat{D} + V_{-i}(k)\right)\right) \end{aligned} \tag{3.43}$$

where $v_i(k)$, $V_{-i}(k)$ represents the value of v_i, V_{-i}, respectively at the kth iteration, and $v_i^{BR}(k)$ represents the best response of player i to the actions of all other players at iteration k. We note that virtual bidder i will observe the quantity $(\widehat{D} + V_{-i}(k))$, assuming that the DA ISO dispatch $x(k)$ is published. We also note that $x_i^\star$ is only dependent on the beliefs of virtual bidder i. Thus (3.43) represents a valid learning dynamic, dependent only on the available information at each iteration. We also assume that the myopic ISO dispatch $\widehat{D}$ does not change, and that each virtual bidder does not change their beliefs $(\alpha_i, \beta_i, \gamma_i, \mu_i)$.

If $\theta_i = 0$, then this learning dynamic would constitute naive best response, where the virtual bidder plays their optimal action at each iteration assuming other virtual bidders do not change their actions. It is interesting to note that this same naive best response strategy is obtained if one attempts to solve (3.14) using gradient descent with exact line search, suggesting that a repeated virtual bidding game could in fact represent a form of gradient-descent algorithm that approximates the solution of the ISO problem.

Remark 2. *An alternate interpretation of the learning dynamic in* (3.43) *is as the expected trajectory under a randomized update policy. At each iteration the virtual bidder i adopts their previous action with probability θ_i, and their best response to the previous actions of all other bidders with probability $(1-\theta_i)$. All the following results hold for this stochastic learning dynamic; however, the concept of asymptotic stability is replaced with convergence in expectation.*

Stability and Convergence Analysis

We assume that all virtual bidders adopt the learning dynamic in (3.43). Considering the collective learning dynamics of all virtual bidders it can be shown that

$$(v(k+1) - v^\star) = A_\Theta (v(k) - v^\star) \tag{3.44}$$

where A_Θ is defined as

$$A_\Theta := \frac{1}{2}\left((\mathbf{I} + \Theta) - (\mathbf{I} - \Theta)\mathbf{E}\right) \tag{3.45}$$

where $\Theta = \text{diag}(\theta_1, \ldots, \theta_N)$.

We have the following result

Theorem 4. *As $k \to \infty$, $v(k) \to v^\star$, under the learning dynamics in* (3.43)*, if*

$$\frac{N-3}{N+1} < \theta_i < 1, \quad \forall i = 1, \ldots, N \tag{3.46}$$

Proof. We assume that all virtual bidders adopt the learning dynamic in (3.43), and denoting $\eta = [\eta_1, \ldots, \eta_N]^\top$, and $\chi = \left(\eta x^\star - \widehat{D}\mathbf{1}\right)$, we have the full system update as

$$\begin{aligned} v(k+1) &= \Theta v(k) + \frac{(\mathbf{I} - \Theta)}{2}\left(\chi - (\mathbf{E} - \mathbf{I})v(k)\right) \\ &= \frac{(\mathbf{I} + \Theta) - (\mathbf{I} - \Theta)\mathbf{E}}{2} v(k) + \frac{(\mathbf{I} - \Theta)}{2}\chi \end{aligned} \tag{3.47}$$

It is straightforward to show that $v^\star$ in (3.17) is a unique fixed point of this iteration. It can also be shown that

$$(v(k+1) - v^\star) = A_\Theta (v(k) - v^\star) \tag{3.48}$$

where A_Θ and Θ are as defined in the text. To show asymptotic stability of the unique pure Nash equilibrium $v^\star$ under these learning dynamics, we require that $|\rho(A_\Theta)| < 1$. We cannot characterize $\lambda(A_\Theta)$ analytically, but we simply require that $\lambda_{\max}(A_\Theta) < 1$, and $\lambda_{\min}(A_\Theta) > -1$. We have that

$$\begin{aligned} \lambda_{\max}(A_\Theta) &= \lambda_{\max}\left(\frac{(\mathbf{I} + \Theta) + (\Theta - \mathbf{I})\mathbf{E}}{2}\right) \\ &\leq \lambda_{\max}\left(\frac{\mathbf{I} + \Theta}{2}\right) + \lambda_{\max}\left(\frac{(\Theta - \mathbf{I})\mathbf{E}}{2}\right) \\ &= \frac{1 + \theta_{\max}}{2} - \lambda_{\min}\left(\frac{(\mathbf{I} - \Theta)\mathbf{E}}{2}\right) \\ &\leq \frac{1 + \theta_{\max}}{2} - \lambda_{\min}\left(\frac{(\mathbf{I} - \Theta)}{2}\right)\lambda_{\min}\left(\mathbf{E}\right) \\ &= \frac{1 + \theta_{\max}}{2} \end{aligned} \tag{3.49}$$

and using a similar analysis it can be shown that

$$\lambda_{\min}(A_\Theta) \geq \frac{1+\theta_{\min}}{2} - \frac{(1-\theta_{\min})}{2}N \tag{3.50}$$

Solving the following simple inequalities completes the proof.

$$\frac{1+\theta_{\max}}{2} < 1, \quad \frac{1+\theta_{\min}}{2} - \frac{(1-\theta_{\min})}{2}N > -1 \tag{3.51}$$

□

Theorem 4 is equivalent to stating that the unique pure Nash equilibrium is globally asymptotically stable under the learning dynamics in (3.43), if condition (3.46) is satisfied.

We note that naive best response, *i.e.* $\theta_i = 0$, is only asymptotically stable for $N < 3$. As N grows larger, the feasible range of θ_i, over which the learning dynamics are asymptotically stable, shrinks. As to whether real virtual bidders would adopt smoothing parameters which satisfy (3.46) is unclear.

We now consider the speed of convergence of the learning dynamics. We have that

$$\begin{aligned}
(v(k+1) - v^\star) &= A_\Theta \left(v(k) - v^\star\right) \\
&= \left(A_\Theta\right)^k \left(v(0) - v^\star\right) \\
\|(v(k+1) - v^\star)\| &= \| \left(A_\Theta\right)^k \left(v(0) - v^\star\right) \| \\
&\leq \|A_\Theta\|^k \| \left(v(0) - v^\star\right) \| \\
&= \rho \left(A_\Theta\right)^k \| \left(v(0) - v^\star\right) \|,
\end{aligned} \tag{3.52}$$

where $\rho(A_\Theta)$ denotes the spectral radius of A_Θ. We see that the virtual bidders converge linearly to the Nash equilibrium at the rate of the spectral radius of A_Θ. It can be shown that

$$\rho \left(A_\Theta\right) = \max \left(\frac{1+\theta_{\max}}{2}, \frac{(1-\theta_{\min})}{2}N - \frac{1+\theta_{\min}}{2} \right) \tag{3.53}$$

For fast convergence we want $\rho \left(A_\Theta\right)$ as small as possible. This minimum is achieved if all virtual bidders adopt the same smoothing parameter $\theta_i = \frac{N-2}{N+2}$, $\forall i = 1, \ldots, N$. For large N, this convergence will be slow. More generally, the speed of convergence is limited by the fact that the only information each virtual bidder receives on the actions of the other players is the sum of their bids. This means that if we allowed virtual bidders to update and improve their estimates $(\alpha_i, \beta_i, \gamma_i, \mu_i)$, at each iteration this would not necessarily improve the speed of convergence. In fact it would only serve to shift the Nash equilibrium towards the equilibrium under perfect beliefs. If we assume that the crowd is wise, $\mu_\eta = 1$, and remains wise as virtual bidders improve their private estimates, then the quality of the information received by each virtual bidder, namely the sum of the bids of other bidders, is not improved by updated private beliefs.

3.5 Conclusions

We have analysed a simple model of a two-settlement market under a myopic ISO dispatch, which provides insight into the equilibrium behavior of virtual bidders. The key results are as follows. At equilibrium, if the crowd of virtual bidders is wise, the DA schedule tends to the social optimum, and the expected price spread tends to zero, as the number of virtual bidders grows large. Additionally the variance of the price spread under virtual bidding is always greater than or equal to the variance in the case where there is no virtual bidding, explicit or implicit. We have also proposed simple learning dynamics, which have as their asymptotically stable equilibrium the Nash equilibrium of the virtual bidding game.

It is important to acknowledge the differences between this simplified model and real-world two-settlement power markets. One problem, highlighted by [44], is that typically the DA and RT clearing algorithms are run in different ways. Namely that the DA dispatch must consider commitment costs, and the RT dispatch must consider constraints such as ramping limits of generation. Another issue is that virtual bids are settled at DA hourly prices, but the RT market is typically run every 5-15 minutes. The 'RT price' that is used to settle virtual bids is often the average hourly RT price. These concerns distort the incentives of virtual bidders, and can lead to undesirable behavior. Furthermore, real power markets are run on networks, with generation and load varying from node to node, in addition to requiring DA schedules to satisfy transmission constraints and contingency scenarios. We hope to address this general network problem in future work.

Finally, the environment in which real virtual bidders are speculating and learning is far from homogeneous. It is in fact highly heterogeneous, with network conditions, generation costs, and parameter distributions changing from day to day. This makes learning very difficult, and it is questionable whether virtual bidders can ever reach the equilibrium solutions presented in this chapter. An interesting piece of work for future study would be to understand how far away the actions of real bidders are from equilibrium and the effect that this has on social welfare and price convergence.

Chapter 4

Robust Cournot-Bertrand Equilibria on Power Networks

The previous chapter discussed how uncertainty is addressed in the structure and operation of electricity markets, and the actions of virtual bidders backed by no physical assets. In this chapter we will focus on the actions of real market participants backed by physical assets and how they respond to and manage uncertainty. Specifically we examine how the classical concept of Nash-Cournot equilibrium among generators changes in the presence of uncertainty. This chapter is the result of joint work with Dr. Eric Munsing and the majority of the results were first published in [56] by the author.

4.1 Introduction

As we have seen, electricity is unique among commodities, having highly inelastic demand, very limited storage, network flows determined by Kirchoff's laws, and transmission constraints which can isolate consumers from low-cost suppliers. The deregulation of electricity generation in many regions has left the operation of the power grid in the hands of Independent System Operators (ISOs), who are tasked with collecting bids for supply and demand, clearing the market in such a way as to meet transmission and security constraints, and mitigating the use of market power. However, the characteristics which make electricity unique also make energy markets prone to manipulation, and empirical studies have shown that these markets often operate as oligopolies in which participants affect outcomes by adjusting their bid curves to maximize profits [57, 58]. By modeling strategic equilibria in energy markets, researchers hope to measure social welfare impacts, design regulatory or technical changes which can promote more competitive markets, or identify noncompetitive behavior by comparing models with *ex-post* market outcomes.

A key decision for producers in energy markets is how to respond to uncertainty in supply, demand, and the actions of other producers. Risk-averse producers may hedge their production through long-term contracts, or submit their generation capacity as "must-

take" (accepting any price). There have been a number of studies examining the impact of uncertainty on strategic equilibria in energy spot markets, particularly in two-settlement markets, for example [59–61]. However, these approaches can be intractable due to the computational burden required to model uncertainty on a large network using stochastic models [62, 63].

This work advances prior literature by showing how robust optimization can be used to integrate uncertainty into a convex model of strategic equilibrium in a single-settlement Poolco electricity market with network constraints, allowing us to scale our results to large networks while representing the most common form of spot market operation.

Prior Literature

A number of game theoretic models of strategic competition have been used to model oligopoly behavior in electricity networks, and are reviewed in [62–65]. We highlight four game theoretic models used to identify strategic equilibria in electricity markets: *Cournot* competition assumes that producers adjust their output to maximize profits [66–68], *Bertrand* competition assumes producers adjust prices [62], *Stackelberg* leader-follower games assume that some firms may have greater decision-making power and serve as market leaders [69,70], and *Supply Function Equilibria* models combine features of Cournot and Bertrand models, using differential equations to model the bidding behavior of firms [51]. Of these, Cournot models have gained particular attention for modeling electricity markets due to their mathematical and computational simplicity, as well as their ability to forecast market outcomes [64, 68, 71].

However, since electricity markets are coupled with complex engineered systems these game-theoretic models are not a panacea. Unlike most commodities, electricity markets are built on a transmission network with thousands of nodes [66], have temporal output constraints on generation equipment [72], and are typically structured as a series of sequential markets [73]. To address these issues, a parallel body of literature has sprung up in which engineering models are used to reflect the technical decisions faced by individual producers [74]. These models are often nonlinear, nonconvex, and computationally intractable for modeling the decisions of more than a single producer with a small fleet of generators.

Both game-theoretic and engineering models are challenged by the uncertainty inherent in electricity provisioning: demand is dependent on weather, generation plants may have unplanned outages, and increased penetration of renewable energy sources makes supply uncertain. A variety of techniques from stochastic optimization have been used to model the generators' decision process under these uncertainties [75, 76]: historical energy prices can be used to construct Monte Carlo simulations [77, 78] or to fit parametric probability distribution models to sources of uncertainty [79].

These stochastic approaches can optimize expected profits, but struggle to deal with modeling uncertainty in the hundreds or thousands of nodes which characterize electricity grids. Robust optimization theory [80] and robust game theory [81] provide an alternative approach to integrating uncertainty, by seeking a solution which still performs well under

a 'worst case' scenario. While this approach is anticipated to reduce the expected profits for the operators relative to their non-robust actions [82], it can be attractive for risk-averse players as it guarantees profits against uncertainty. These models are also mathematically appealing as they do not require any distributional assumptions on random variables and can preserve the convexity of the optimization problem, allowing the application of efficient solvers which can scale up to handle uncertainty across thousands of nodes.

Robust optimization has previously been applied to game theoretic problems, allowing the modeling of uncertainty in payoff matrices or competitors' strategies [83,84]. However, robust optimization has only been applied to specific sub-problems in electricity market operation, e.g. the unit commitment problem of the system operator [10, 85–87], nonstrategic bidding as a price-taker [78], strategic equilibrium in a Stackelberg leader-follower game [88], and strategic equilibrium without congestion costs [89].

Novel Contributions

We propose a convex formulation, computing the robust strategic equilibrium in an electricity network with congestion and demand uncertainty, and demonstrate it using a sample network. The contributions are:

- Convex formulation of robust Cournot-Bertrand equilibrium in a single-settlement nodal Poolco electricity market
- Demonstration of the impact of congestion on robust strategic equilibria
- Demonstration of the impact of robust strategies on social welfare outcomes in electricity markets.

To the author's knowledge this is the first attempt to characterize robust Cournot-Bertrand equilibria in electricity markets on transmission networks, extending the work of [90–92] to incorporate uncertainty and risk-averse producers.

Organization

In Section 4.2 we present the ISO and producer problems as Cournot-Bertrand competition on electricity networks. We formulate the resulting equilibrium as a monotone linear complementarity problem (LCP), which can be solved as a convex QP. We apply the results of [93, 94] to develop a robust LCP and formulate the robust counterpart of the corresponding convex QP. In Section 4.3 we present results for a simple example problem. The impact on producer profits, consumer surplus, and net social benefit is discussed, and Section 4.4 concludes.

4.2 Problem Formulation

This formulation builds on the work of [90] and [92], obtaining the equilibrium as the solution of a linear complementarity problem, for which a convex robust counterpart is developed. Background on modeling energy markets with complementarity problems is provided in [62] and background on robust optimization theory can be found in [80].

Network Modeling

We adopt the DC power flow model, its relevant notation, and its associated feasible injection region detailed in (2.14). To be consistent with the literature, for this chapter alone, we adopt the convention that net power withdrawals are positive, and net power injections are negative. We will denote these as nodal imports, $r_i > 0$, and nodal exports, $r_i < 0$, $i = 1, \ldots, n$, respectively. This convention change means that the sign of the shift matrix is flipped, and for this chapter alone we define $H := \begin{bmatrix} -\mathbf{I} \\ \mathbf{I} \end{bmatrix} \hat{H} \in \mathbb{R}^{2m \times n}$, such that we have the following feasible injection region

$$\mathcal{R} := \left\{ r \in \mathbb{R}^n \mid Hr \leq c,\ \mathbf{1}^\top r = 0 \right\} \tag{4.1}$$

We assume there are $|\mathcal{F}|$ firms, owning generation units at nodes $i \in \mathcal{N}_f \subset \mathcal{N}$, $f \in \mathcal{F}$. Each firm f makes production quantity decisions for its generators ($\{q_i\}_{i \in \mathcal{N}_f}$), where $0 \leq q_i \leq \overline{q}_i$. For simplicity, we assume that there is at most one generation unit per node.[1] Generation costs $C(q_i)$ are assumed to be convex.

At each node, we assume a (steeply) decreasing affine inverse demand function $D(x_i)$, where x_i is the quantity demanded at node i. This is a common assumption for relatively inelastic electricity markets [57], and can represent a linearization of a more complicated inverse demand function. It is worth noting that in general x_i is endogenous, and is calculated as $x_i = r_i + q_i$.

The ISO Problem

The ISO controls the import (export) $r_i > 0$ ($r_i < 0$) at each node $i \in \mathcal{N}$ and sets the corresponding locational marginal prices (LMPs). These quantities must satisfy the network feasibility constraints, determined by the set $\mathcal{R}$. The ISO's objective is to maximize social welfare, taken as the aggregated area under the nodal inverse demand functions $D_i(\cdot)$, less the sum of all generation costs $C_i(\cdot)$. Mathematically, the ISO solves the following problem, parametric on the firms' production decisions ($\{q_i\}_{i \in \mathcal{N}}$):

[1] This can be achieved in practice by introducing dummy nodes into the network.

$$\begin{aligned} \underset{r_i}{\text{maximize}} \quad & \sum_{i\in\mathcal{N}} \left(\int_0^{r_i+q_i} D_i(\tau_i)d\tau_i - C_i(q_i) \right) \\ \text{subject to} \quad & \mathbf{1}^\top r = 0, \ : \ \gamma \\ & -Hr \leq c, \ : \ \beta \end{aligned} \tag{4.2}$$

where τ_i, $i = 1, \ldots, n$, is a dummy variable of integration. As in [92], we have excluded the nonnegativity constraints $r_i + q_i \geq 0$, $i \in \mathcal{N}$, by implicitly assuming an interior solution with respect to these constraints. The KKT conditions are as follows

$$0 = D_i(q_i + r_i) - \gamma - \psi_i, \quad i \in \mathcal{N} \tag{4.3a}$$

$$0 = \psi - H^\top \beta \tag{4.3b}$$

$$0 = \mathbf{1}^\top r \tag{4.3c}$$

$$0 \leq \beta \perp T - Hr \geq 0 \tag{4.3d}$$

where ψ is included for convenience, and is defined according to (4.3b). The first KKT condition implies that

$$q_i + r_i = (D_i)^{-1}(\gamma + \psi_i), \quad i \in \mathcal{N} \tag{4.4}$$

And consequently,

$$\sum_{i\in\mathcal{N}} q_i = \sum_{i\in\mathcal{N}} (D_i)^{-1}(\gamma + \psi_i) \tag{4.5}$$

This equation represents the aggregate demand function in the network relating the total consumption quantity to the reference node price γ and the nodal price premiums $\{\psi_i\}_{i\in\mathcal{N}}$, which determine the relative value of LMPs. We denote the LMP vector λ as

$$\lambda = \gamma \mathbf{1} + \psi \tag{4.6}$$

To prevent arbitrage between nodes i and j, the corresponding congestion charge must be $\psi_j - \psi_i$.

The Firm's Problem

We assume that generation firms do not anticipate the impact of their production on the congestion prices set by the ISO. We model this 'bounded rationality' as a game where the ISO and generation firms move simultaneously. Similar to [92] we use a mixed Cournot-Bertrand model, where the ISO behaves a la Bertrand, setting locational price differences, while the generation firms are Cournot players with respect to each other (*i.e.* set quantities), but treat the ISO as a price setter. The reasons for choosing the ISO as a Bertrand player are well discussed in [92].

Each firm chooses its production quantities to maximize profits with respect to the residual demand defined implicitly by (4.5). In this formulation, the reference bus price γ is determined implicitly by the aggregate production decisions of all the generation firms, just

as in a regular Cournot game. However, these production decisions and the implied reference node price also depend on the nodal premiums $\{\psi_i\}$ set by the ISO. The resulting problem solved by each firm $f \in \mathcal{F}$ is

$$\begin{aligned} \underset{q_i : i \in \mathcal{N}_f, \gamma}{\text{maximize}} \quad & \sum_{i \in \mathcal{N}_f} (\gamma + \psi_i) q_i - C_i(q_i) \\ \text{subject to} \quad & 0 \le q_i \le \overline{q}_i, \ : \ \nu_i^-, \nu_i^+, \quad i \in \mathcal{N}_f, \\ & \sum_{i \in \mathcal{N}} q_i = \sum_{i \in \mathcal{N}} (D_i)^{-1}(\gamma + \psi_i), \ : \ \mu_f \end{aligned} \tag{4.7}$$

The KKT conditions are as follows

$$\begin{aligned} 0 &= \gamma + \psi_i - \frac{\partial C_i(q_i)}{\partial q_i} + v_i^- - v_i^+ - \mu_f, \quad i \in \mathcal{N}_f \\ 0 &= \sum_{i \in \mathcal{N}_f} q_i + \mu_f \sum_{i \in \mathcal{N}} \frac{\partial (D_i)^{-1}(\gamma + \psi_i)}{\partial \gamma} \\ 0 &= \sum_{i \in \mathcal{N}} (D_i)^{-1}(\gamma + \psi_i) - \sum_{i \in \mathcal{N}} q_i \\ 0 &\le \nu_i^- \perp q_i \ge 0, \quad i \in \mathcal{N}_f \\ 0 &\le \nu_i^+ \perp \overline{q}_i - q_i \ge 0, \quad i \in \mathcal{N}_f \end{aligned} \tag{4.8}$$

We only consider a single market (e.g. spot market) and do not consider optimization across different energy markets (e.g. forward markets or ancillary services), however we will show that it is possible to represent uncertainty with respect to the outcomes of different energy markets.

These assumptions are consistent with other literature [73, 90] and with the approaches used by most ISOs for scheduling hour-ahead and real-time markets, where the computational benefits of the (convex) lossless DC power flow model are important.

Equilibrium Conditions of the Deterministic Game

Aggregating the KKT conditions for the firms' and the ISO's programs yields the equilibrium conditions, which in general form a mixed nonlinear complementarity problem. It becomes a mixed LCP when both the nodal demand functions and the marginal cost functions are linear, as is assumed henceforth.

Let the inverse demand functions and the cost functions be, respectively

$$D_i(x_i) = a_i - b_i x_i, \quad i \in \mathcal{N} \tag{4.9}$$

$$C_i(q_i) = d_i q_i + \frac{1}{2} s_i q_i^2, \quad i \in \mathcal{N} \tag{4.10}$$

where $a_i, b_i, d_i, s_i \ge 0$. We denote $a = \text{vec}(a_i)$, $B = \text{diag}(b_i)$, $d = \text{vec}(d_i)$, $S = \text{diag}(s_i)$.

We denote $L \in \mathbb{R}^{|\mathcal{N}|\times|\mathcal{F}|}$ as the firm-node assignment matrix, where $L_{ij} = 1$ if node i is owned by firm j, and $L_{ij} = 0$ otherwise. We also denote $\mu \in \mathbb{R}^{|\mathcal{F}|}$, where μ_i is the dual variable associated with firm i. Also denoting $\sum_{i\in\mathcal{N}} \frac{1}{b_i} = \mathbf{1}^\top B^{-1}\mathbf{1} = \rho$, the equilibrium conditions are then

$$0 = \gamma\mathbf{1} + H^\top\beta - d - Sq + v^- - v^+ - L\mu \tag{4.11}$$
$$0 = L^\top q - \mu\rho \tag{4.12}$$
$$0 = \gamma + \frac{\mathbf{1}^\top q}{\rho} - \frac{\mathbf{1}^\top B^{-1}H^\top\mu}{\rho} - \frac{\mathbf{1}^\top B^{-1}a}{\rho} \tag{4.13}$$
$$0 \le \nu^- \perp q \ge 0 \tag{4.14}$$
$$0 \le \nu^+ \perp \overline{q} - q \ge 0 \tag{4.15}$$
$$0 = \mathbf{1}^\top r \tag{4.16}$$
$$0 = a - B(q + r) - \gamma\mathbf{1} - H^\top\beta \tag{4.17}$$
$$0 \le \beta \perp c - Hr \ge 0 \tag{4.18}$$

Here, (4.11)-(4.15) are the aggregated KKT conditions for the firms' problems, and (4.16)-(4.18) are the aggregated KKT conditions for the ISO's problem. Under the assumption of linear demand functions and quadratic convex cost functions, the firms' and the ISO's programs are strictly concave-maximization problems, so (4.11)-(4.18) are also sufficient. Note that (4.13) can be excluded from the preceding market equilibrium conditions because it is implied by (4.16) and (4.17). This set of equations constitutes a mixed linear complementarity program (mLCP).

We wish to turn these set of conditions into a compact LCP. This derivation closely follows that in [92]. We first write out equations (4.16) and (4.17) as follows

$$\begin{bmatrix} a \\ 0 \end{bmatrix} - \begin{bmatrix} B \\ 0 \end{bmatrix} q - \begin{bmatrix} B & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{bmatrix} \begin{bmatrix} r \\ \gamma \end{bmatrix} - \begin{bmatrix} H^\top \\ 0 \end{bmatrix} \beta = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{4.19}$$

Rearranging and solving for γ and r yields

$$r = Qa - QBq - QH^\top\beta \tag{4.20}$$
$$\gamma = \frac{\mathbf{1}^\top B^{-1}}{\rho}a - \frac{\mathbf{1}^\top}{\rho}q - \frac{\mathbf{1}^\top B^{-1}}{\rho}H^\top\beta \tag{4.21}$$

where, denoting $\mathbf{E} = \mathbf{1}\mathbf{1}^\top$

$$Q = B^{-1} - \frac{B^{-1}\mathbf{E}B^{-1}}{\mathbf{1}^\top B^{-1}\mathbf{1}} \tag{4.22}$$

We note that

$$QB = I - \frac{B^{-1}\mathbf{E}}{\rho}, \quad BQ = I - \frac{\mathbf{E}B^{-1}}{\rho} \tag{4.23}$$

We now consider equations (4.11) and (4.12). We have that $\mu = \frac{L^\top q}{\rho}$, and that $LL^\top = \Lambda$, where $\Lambda_{ij} = 1$ if either $i = j$, or if $i \neq j$ but node i and j are owned by the same firm, $\Lambda_{ij} = 0$ otherwise. Using substitution we rewrite equation (4.11) as

$$\begin{aligned} 0 = \mathbf{1}(\frac{\mathbf{1}^\top B^{-1}}{\rho}a - \frac{\mathbf{1}^\top}{\rho}q - \frac{\mathbf{1}^\top B^{-1}}{\rho}H^\top\beta) + H^\top\beta \\ -d - Sq + \nu^- - \nu^+ - \frac{\Lambda q}{\rho} \end{aligned} \tag{4.24}$$

Collecting terms, using (4.23), and solving for ν^- we get

$$\nu^- = (BQ - I)a + d + (S + \frac{Y}{\rho} + \frac{\mathbf{E}}{\rho})q - BQH^\top\beta + \nu^+ \tag{4.25}$$

We denote $N = (S + \frac{Y}{\rho} + \frac{\mathbf{E}}{\rho})$, where $N \in \mathbb{S}_+$ is positive semi-definite, and has the following properties

$$N_{ij} = \begin{cases} \frac{2}{\rho} + s_i, \text{ if } i = j, \\ \frac{2}{\rho}, \text{ if } i \neq j, \text{ and the units at nodes } i \text{ and } j \\ \qquad \text{belong to the same firm,} \\ \frac{1}{\rho}, \text{ otherwise} \end{cases} \tag{4.26}$$

We can now write out the following LCP

$$w = \begin{bmatrix} \overline{q} - q \\ \nu^- \\ c - Hr \end{bmatrix}, \qquad z = \begin{bmatrix} \nu^+ \\ q \\ \beta \end{bmatrix},$$

$$t = \begin{bmatrix} \overline{q} \\ (BQ - I)a + d \\ c - HQa \end{bmatrix}, \qquad M = \begin{bmatrix} 0 & -I & 0 \\ I & N & -BQH^\top \\ 0 & HQB & HQH^\top \end{bmatrix}$$

where $w = t + Mz$, $w \geq 0$, $z \geq 0$, $w^\top z = 0$. We notice that M is square but not symmetric. Since it is square we can write M as the sum of a symmetric matrix P and a skew symmetric matrix K, such that $M = P + K$, where $P \succeq 0$.

$$M = \begin{bmatrix} 0 & 0 & 0 \\ 0 & N & 0 \\ 0 & 0 & HQH^\top \end{bmatrix} + \begin{bmatrix} 0 & -I & 0 \\ I & 0 & -BQH^\top \\ 0 & HQB & 0 \end{bmatrix} \tag{4.27}$$

Due to the fact that $z^\top Mz = \frac{1}{2}z^\top(M + M^\top)z = z^\top Pz$, we can solve the LCP by solving the following convex QP

$$\begin{aligned} \underset{z \geq 0}{\text{minimize}} \quad & h(z) = z^\top Pz + t^\top z \\ \text{subject to} \quad & Mz + t \geq 0 \end{aligned} \tag{4.28}$$

with any solution z^* solving the LCP(M, t), iff $h(z^\star) = 0$ [62].

Formulating a Robust Counterpart

We wish to identify strategies for the producers that are robust to uncertainty. Three potential sources of uncertainty for a generator are: the parameters of the inverse demand function, the quantity of zero marginal cost renewable generation in the network, and the volume of forward contracts signed by other generation firms. All of these sources can be represented as aggregate uncertainty in the residual demand curve faced by a producers, however we will see that this formulation can additionally capture more general sources of uncertainty.

We seek a robust equilibrium where producers maximize their profits, robust to uncertainty in the residual demand curve, while assuming that other producers are adopting robust strategies. A robust optimization problem takes the form

$$\min_{x \in X} \max_{u \in \mathcal{U}} f(x; u) \tag{4.29}$$

which determines the best possible action $x^\star$ under a worst case realization of uncertainty $u \in \mathcal{U}$. As seen in (4.28), the equilibrium solution of the deterministic problem is an LCP which can be formulated as a convex QP. We obtain a robust equilibrium solution by considering a robust LCP, and formulating the robust counterpart to the equivalent convex QP.

A nominal LCP(M, t) has the form

$$0 \leq z \perp Mz + t \geq 0 \tag{4.30}$$

The function $h(z) = z^\top (Mz + t)$ is known as the residual of LCP(M, t), with $h(z) = 0$ iff z solves LCP(M, t). Applying the results of [93] we define an uncertain LCP(u) as

$$0 \leq z \perp M(u)z + t(u) \geq 0 \tag{4.31}$$

where $M(u), t(u)$ are parametric on the realization of a random variable $u \in \mathcal{U}$. A robust solution to the LCP seeks to find a feasible solution z^* which minimizes the residual function $h(z; u)$ under a worst case uncertainty realization u^*. This takes the form

$$\begin{aligned} \min_{z \geq 0} \max_{u \in \mathcal{U}} \quad & z^\top (M(u)z + t(u)) \\ \text{subject to} \quad & \min_{u \in \mathcal{U}} M(u)_i z + t_i(u) \geq 0, \ \forall i \end{aligned} \tag{4.32}$$

In [94], the authors show that this problem is tractable for affine uncertainty sets of the form

$$\begin{aligned} t(u) &= t_0 + \sum_{l=1}^{L} u_l t_l \\ M(u) &= M_0 + \sum_{k=1}^{K} u_k M_k, \ M_0 \succeq 0, \ M_k \succeq 0, \ \forall k \\ u &\in \mathcal{U} \subseteq \mathbb{R}^{L+K} \end{aligned} \tag{4.33}$$

where L and K are general scalars defining the affine uncertainty set, and $\mathcal{U}$ can take any of the following forms[2]

$$\begin{aligned} \mathcal{U}_1 &= \{u \ : \ \|u\|_1 \leq 1\}, \ \mathcal{U}_2 = \{u \ : \ \|u\|_2 \leq 1\} \\ \mathcal{U}_\infty &= \{u \ : \ \|u\|_\infty \leq 1\} \end{aligned} \tag{4.34}$$

While the formulation is general, for exposition we restrict our attention to uncertainty in $t(u)$, such that $M(u) = M$, $\forall u \in \mathcal{U}$. If $\mathcal{U} = \mathcal{U}_\infty$, then (4.32) takes the form

$$\begin{aligned} \min_{z \geq 0} \quad & z^\top (Mz + t_0) + \sum_{l=1}^{L} \|z^\top t_l\|_1 \\ \text{subject to} \quad & M_i z + t_0 - \sum_{l=1}^{L} \|(t_l)_i\|_1 \geq 0, \ \forall i \end{aligned} \tag{4.35}$$

As stated previously we consider uncertainty in the residual demand function, which we treat as interval uncertainty in the intercept of the inverse demand functions at each node. We consider functions of the form

$$\begin{aligned} D_i(x_i; \zeta_i) &= a_i(\zeta_i) - b_i x_i, \ i \in \mathcal{N} \\ a_i(\zeta_i) &= a_0 + \zeta_i a_{il}, \ \|\zeta_i\|_\infty \leq 1 \end{aligned} \tag{4.36}$$

Where $a_{il} \geq 0$ is a common belief among producers regarding the bounds of the uncertainty interval at node i. We implicitly assume that all firms have the same belief regarding the uncertainty in the inverse demand function at each node. This assumption is required for the robust LCP model we have used here. Incorporating different beliefs regarding uncertainty to capture the presence of firms with different risk preferences should be possible, although will require a much closer treatment of the individual robust optimization that each firm faces.

Since a only appears in t in the deterministic LCP, it is straightforward to translate uncertainty in a to uncertainty in t, with the resulting robust LCP having the form of (4.35). This is a convex optimization problem, and thus a robust equilibrium solution exists but may not be unique. The problem can be solved using standard convex optimization solvers.

4.3 Example and Results

The model is demonstrated on an example network, and outcomes are compared with conventional (non-robust) Nash-Cournot equilibrium. Networks with up to 300 buses were simulated and were all found to demonstrate qualitatively similar behavior, thus for expositional clarity the simple 3-node network shown in Fig. 4.1 is used here, similar to the network modeled in [95].

[2]For the case when $\mathcal{U} = \mathcal{U}_2$, the M_0, M_k are restricted to be symmetric matrices.

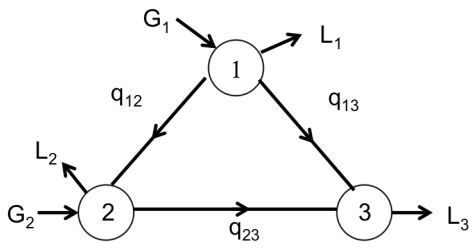


Figure 4.1: 3 Node Network

Example Network

The three buses $i = 1, 2, 3$ have customers with inverse demand functions $D_i(x_i) = 40 - 0.08q_i$, $i = 1, 2$, and $D_3(x_3) = 35 - 0.05q_3$ \$/MWh (node 3 has greater demand elasticity). Each pair of buses is connected by a single transmission line, and all three lines have equal impedance. The market has two firms $f = 1, 2$ each with a single generator in its fleet; Firm 1's generator is sited at $i = 1$, while Firm 2's generator is at $i = 2$. Both generators have a maximum capacity of $q_i = 1000$MW. Each generator has a constant marginal cost: $d_1 = \$15$/MWh for firm 1, and $d_2 = \$20$/MWh for firm 2. We simulate a congested scenario by imposing a 20MW constraint on the line between nodes 1 and 2, and a 35MW constraint on the line between nodes 1 and 3.

In the examples that follow, we assume that $a_{il} = a_l$, $\forall i$. That is the uncertainty at each node is independent but lies in the same interval. The uncertainty a_l is swept over a range of \$0-15/MWh.

For comparison, the non-robust quantity decisions (i.e. classic Nash-Cournot equilibrium) are also modeled. For each scenario, the non-robust quantity is intercepted with the realized demand curve to produce the price at each node. This simulates the scenario where uncertainty is present, but firms behave as if there is no uncertainty. The system is modeled within Matlab using CVX and the Gurobi solver.

Results and Discussion

The profits for the two generation firms are shown in Fig. 4.2 for the case when the transmission lines are unconstrained and there is no congestion in the network. We plot the results for the range of potential uncertainty realizations as a shaded region. These shaded regions are bounded by the profits achieved at the maximum and minimum limits of the uncertainty interval, plotted as thick lines. The results for the nominal value of the uncertainty are plotted as a third solid line through each shaded region. By construction, the uncertainty interval spans all potential realizations of demand and thus all possible market outcomes.

Compared with classic Nash-Cournot equilibrium, we see that the primary goal of the robust optimization is met: the robust strategy always results in higher profits for the worst case realization of demand. For small ranges of uncertainty, we see that the firms actually make *greater* profits at the robust equilibrium than at the non-robust equilibrium, regardless

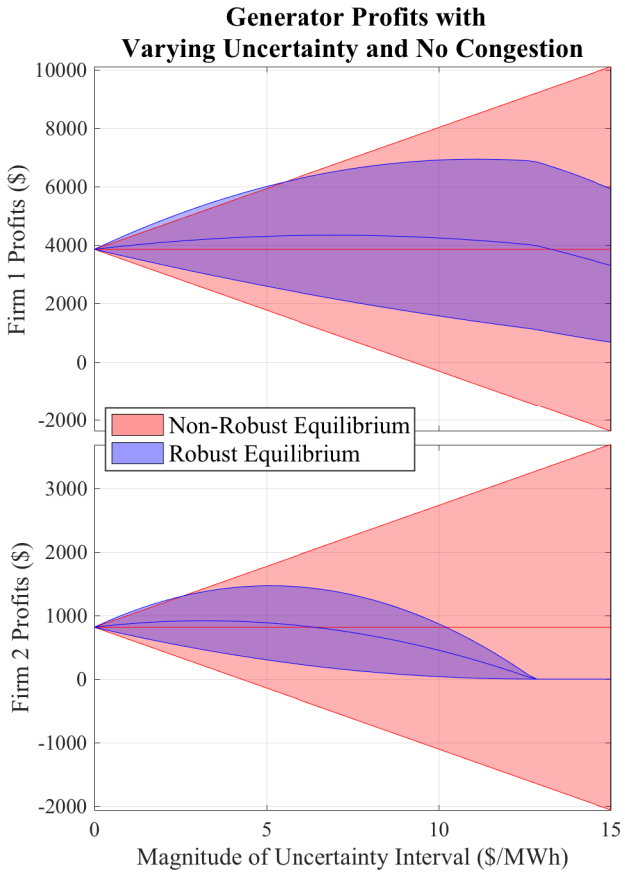


Figure 4.2: Profits of generation firms under a range of uncertainty intervals, with no network congestion. Both robust and non-robust strategies are shown, and the three lines indicate high, nominal, and low realizations of demand.

of the level of demand. This can be explained by observing that each firm restricts its output in order to protect itself from low prices, contracting the net supply curve and driving up prices.

Eventually the reduction in demand due to the higher prices offsets the initial gain in profits, and we see that the Nash-Cournot equilibrium results in higher profits in nominal and high demand scenarios. At low demand the robust scenario still guarantees that the generators will not incur a loss, whereas a Nash-Cournot equilibrium can actually result in a net loss for generators as realized prices fall below the marginal cost of generation.

It is important to emphasize that these results assume that all firms follow the same robust optimization behavior, *i.e.* that they have the same belief about uncertainty and the same sensitivity to risk. However, for a less risk-sensitive firm, there is an incentive to increase production in order to increase expected profits. The final equilibrium would be dependent on the firms' risk acceptance, with greater risk aversion driving firms towards the robust optimization, and lower risk sensitivity driving them towards Cournot optimization.

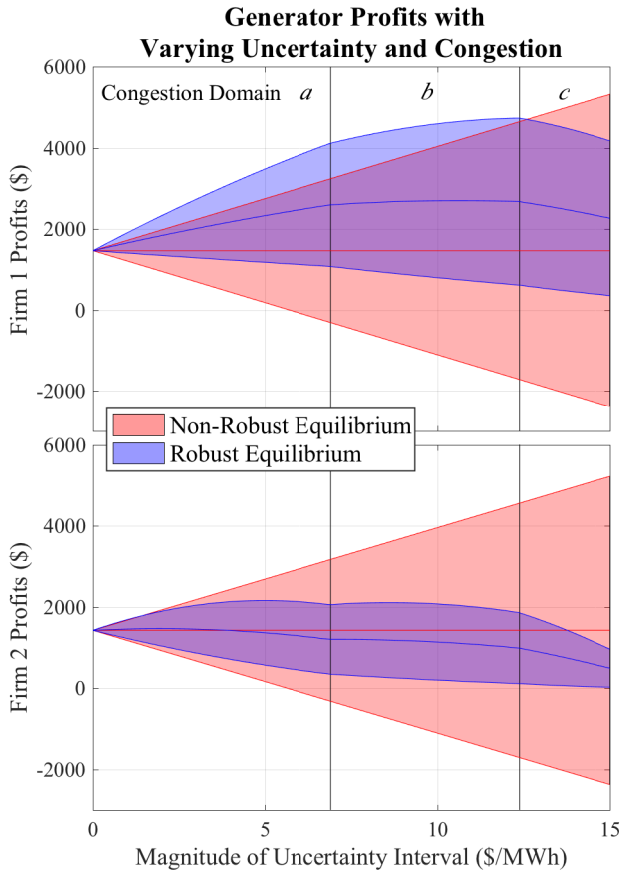


Figure 4.3: Profits of each generating firm when congestion is present on the network. Note that the shape of the curves changes over distinct domains, dictated by the congestion on the network: in domain a line 1-3 is congested; in domain b lines 1-3 and 1-2 are congested, and in domain c only line 1-2 is congested.

This is explored in greater detail in [81].

Network Effects

As firms restrict their production in response to uncertainty, network flows change and can shift the congestion patterns on the network. We can divide the uncertainty range into distinct domains with unique congestion patterns, highlighted in Figure 4.3. Within each domain the residual demand curves for each generator stay constant, and equilibrium follows the principles outlined above for the uncongested case. However as uncertainty increases and the congestion pattern changes, there is a discontinuous shift in the residual demand curve each firm faces, seen as a change of curvature in Figure 4.3.

In our example, comparing Figures 4.2 and 4.3 shows that congestion reduces the profits of Firm 1, but also makes a robust strategy more attractive for low and moderate uncertainty.

Firm 2 benefits from congestion rents, but sees profits more threatened by uncertainty in domain a. As uncertainty increases and the line between nodes 1 and 2 becomes congested in domain b, Firm 2 sees greater benefit from uncertainty and the profits of Firm 1 are eroded. This effect is repeated more dramatically as uncertainty increases into domain c. These effects are dependent on the exact network structure, and were found to be particularly complex for larger networks. We will not delve further into these here, but as we will see in later chapters, this effect is indicative of *critical regions*, defined by the network congestion pattern, with the equilibrium behavior of producers following a different closed form equation in each region.

Welfare Effects

The impact of the robust strategy on consumers is less nuanced. The consumer surplus is calculated as the area above the market clearing price and below the demand curve, representing the surplus value which consumers would have been willing to pay for electricity [96]:

$$CS = \frac{1}{2}(a - \lambda)^\top x \tag{4.37}$$

The total consumer surplus for the market is shown in Figure 4.4 for competitive, Nash-Cournot, and robust equilibria. As we assume that producers offer a fixed quantity of power into the market, the consumer surplus is invariant to the realized inverse demand function for a given uncertainty interval. When firms restrict their output to be robust to low realizations of demand, prices rise above competitive levels, demand decreases, and consumer surplus drops below the Cournot oligopoly level.

The total efficiency of the market can be measured by its net social benefit: the sum of consumer surplus, producer profits, and merchandising surplus[3] [96]. Building on (4.37), this can be written as

$$NSB = CS + (\lambda^\top q - C(q)) + \beta^\top c \tag{4.38}$$

Since robust behavior restricts supply below Nash-Cournot equilibrium levels, the net social benefit decreases monotonically [96] as shown in Figure 4.5.

4.4 Conclusion

Electricity markets are particularly susceptible to non-competitive behavior, making it important to understand strategic equilibria in order to inform better market design and policies. The unique structure of electricity networks, and the many sources of uncertainty in

[3]This is the rent collected by the system operator in the presence of congestion, and can be shown to be equal to $\mu^\top T$.

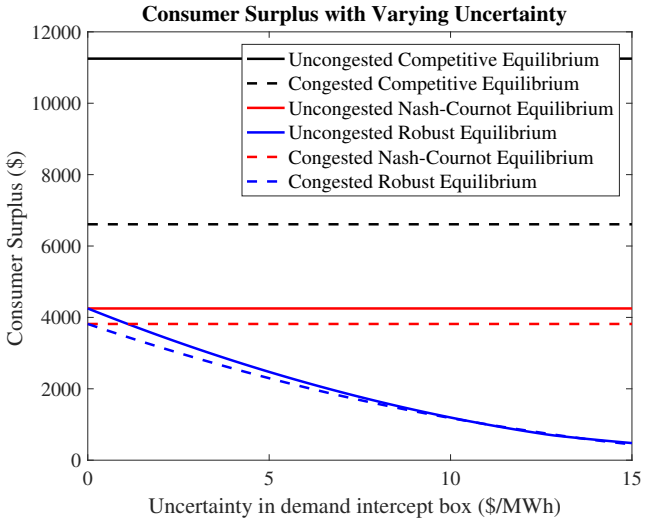


Figure 4.4: Consumer surplus under a number of equilibrium models: perfect competition, Nash-Cournot equilibrium, and Nash-Cournot robust equilibrium. When producers restrict production to be robust to uncertainty, consumers are clearly impacted. Congestion further reduces consumer surplus by introducing congestion charges.

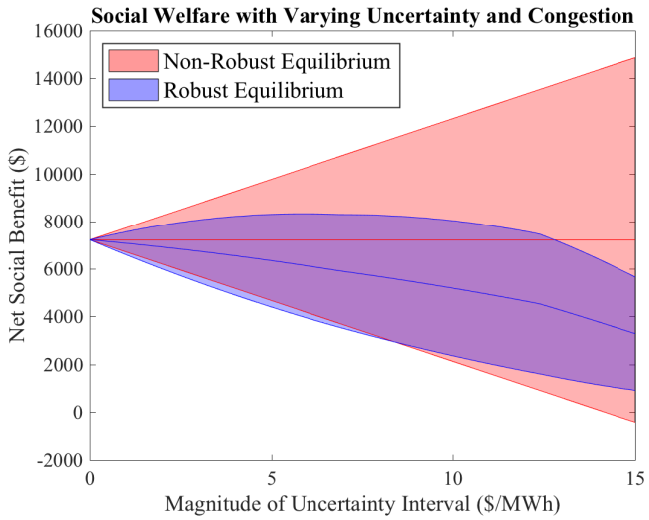


Figure 4.5: Net Social Benefit (sum of consumer surplus, producer profits, and merchandising surplus) under both robust and non-robust equilibrium.

supply and demand, also make it important to have scalable tools for studying the impact of uncertainty on energy markets.

We extend a model of strategic equilibria in electricity networks to include robustness to uncertain demand, reflecting the behavior of risk-averse generation firms. The robust optimization model remains convex, allowing it to be scaled to large power networks. The model is not intended to describe the optimal bidding strategies or bid curves of individual producers, however it provides an efficient way of simulating the impact of uncertainty on market outcomes.

Whereas robust optimization in competitive markets may reduce profits for producers, we see that robustness with small uncertainty intervals *uniformly increases* profits for the generating firms relative to Nash-Cournot equilibrium. The impact of the robust equilibrium on consumers is uniformly negative, as firms restrict their output leading to an increase in prices, similar to that which would be seen under collusive behavior. Thus the "price of robustness" is seen in a reduction of the net social benefit of the market.

Congestion affects different firms unevenly, as it simultaneously creates congestion rents and increases market power for some generators. We show that uncertainty can affect congestion patterns in robust equilibrium, with the exact relationship between generator profits and congestion being dependent on network topology.

These results can be applied to reflect uncertainty in supply due to intermittent renewable electricity generation, by modeling the uncertainty in net load (demand less must-take renewables). The results can also represent uncertainty in the forward contracts signed by other firms, which will contract the residual supply curve in the spot market. By incorporating robustness into the strategic equilibrium, we offer additional insight that producers, utilities, and regulators can use to understand outcomes in real markets.

Chapter 5

Bilevel Problems in Electricity Markets

In the previous chapter we examined the impacts on market outcomes of participants robustly anticipating the effects of uncertainty on their profits. While useful for examining market outcomes, Cournot models in general are not intended to tell us about the way a participant would actually *bid* into the market. This question will be the focus of this next chapter, where we consider the strategic bidding problem of private participants in transmission-constrained electricity markets. We will describe a new solution approach to the bilevel single firm problem using multiparametric programming. Moreover, we will use this methodology to explicitly characterize and compute equilibria in the multi-firm problem. The techniques and methods from multiparametric programming used in this chapter will inform the work in the following two chapters.

5.1 Introduction

Multi-Leader-Follower Games and Bilevel Optimization Problems

The *Multi-Leader-Follower* game is an extension of the concept of the Stackelberg game described in the previous chapter. In a classical Stackelberg game (single-leader-follower), there is a *leader* who maximizes their payoff subject to all other players, the *followers*, being in a competitive equilibrium [97]. This model is typically applied to markets or games where there is a single dominant firm and a competitive fringe, where the dominant firm can be modeled as choosing their action first, and the competitive fringe then responds to this action. The multi-leader-follower game generalizes this idea to the case with multiple dominant firms or leaders.

This concept proves to be extremely useful in analyzing electricity markets, as the bidding and market clearing processes fit the structure of a multi-leader-follower game. In a single-stage Poolco market, the payoffs of market participants depend on their own and

other participants' actions (bids), and the market clearing outcomes determined by the ISO (cleared quantities and prices). The market participants (leaders) decide their bids simultaneously in advance, moving first, and submit these to the ISO (single follower) which moves second, determining cleared quantities and prices from participants' bids. For participants to maximize their payoffs, this leads to a set of *bilevel* optimization problems, where the first level is a participant's private payoff maximization, subject to a second level which is the ISO's market clearing optimization. In other words the participants anticipate their and other participants' impact on the market outcome, and optimize their bids accordingly.

Typically the second level of a bilevel problem is substituted for the equilibrium conditions of the follower's problem, and, assuming the actions of other participants are fixed, this leads to the formulation of the private participant payoff maximization problem as a *mathematical problem with equilibrium constraints* (MPEC). The multi-firm problem requires the simultaenous solution of a number of MPECs, and is commonly referred to in the literature as an *Equilibrium Problem with Equilibrium Constraints* EPEC. In this chapter we will use the theory of multiparametric programming to provide new solution methods to particular subclasses of these problem types in the context of strategic behavior in elextricity markets.

Prior Literature

Solution methodologies for MPEC and EPEC problems are discussed in [98]. For MPECs these include solving sequential non-linear programs (NLPs), the most widely used of which is the PATH algorithm. An alternative solution method is to reformulate the MPEC as an MILP using the big-M method to transform the complementarity constraints to integer constraints. Finally some work has been done on SDP relaxations of the complementary constraints, and for some problem instances these compare favorably to the MILP reformulation. As to EPECs, these are effectively a set of coupled MPECs, and inherit all of their 'bad' properities. In general no global solution exists for an EPEC, so methods tend to find local optima. The most popular approach is the *Diagonalization* method, which involves iteratively solving the MPEC of each player sequentially, while holding the strategies of the other players fixed. As will be discussed in Section 5.5, this method offers no convergence or solution guarantees, is very sensitive to initial conditions, and may be a poor choice of solution method given the structure of the problem. Another approach is the *Simultaneous solution method*, which involves simultaneously solving for the strong stationarity conditions of all MPECs at once. Finally, given the non-convex nature of EPECs, a variety of approximate and exotic algorithms have been proposed to find solutions, including: a predator-prey model [99], genetic algorithm [100, 101], particle swarm optimization [61, 102], bacteria foraging optimization [103], and a bat-inspired algorithm [104]. These algorithms provide no solution or convergence guarantees in general, but can be useful in modelling situations with nonlinear costs and constraints.

The more specific strategic bidding problem is very well studied in the literature, particularly around the turn of the millenium as deregulation became a prominent feature of US power markets. As discussed in the previous chapter many equilibrium models exist

to describe this kind of strategic behavior: Competitive, Cournot, Bertrand, Stackelberg, Conjectured Supply Function, and Supply Function Equilibrium (SFE). An excellent reference describing the salient features of each model, and examples in the literature of such approaches can be found in [64], and [105] provides a comprehensive review of equilibrium models and solution methodologies.

The SFE model is closest to that studied in this Chapter. The SFE model was first introduced in [106], and applied by [107], where each firm bids a 'supply function', relating its quantity to its price, and the equilibrium is found by solving a set of first-order differential equations. The SFE model has since been expanded to include capacity constraints [108], but suffers from an inability to deal with transmission constraints in any systematic way. [51] provides an overview of the advantages and shortcomings of the SFE method. The use of this approach was generally used to model market outcomes and market power, rather than true bidding strategies of generation firms. In recent years, the 'SFE' designation has come to include the class of multi-leader follower strategic bidding problems, where firms bid supply functions to an ISO. These approaches typically only consider the supply function parameters, whereas we shall also consider capacity witholding of generation.

For the case of the single-firm, a parametric approach to this problem was studied in [109], where a single markup parameter was introduced for each generator. The authors observed that the parameter space could be partitioned, and proposed an iterative scheme to discover the optimal markup parameter, rather than pre-specifying the critical regions of the problem, as we shall see later in this Chapter. In [110], the MPEC of the single-firm problem is reformulated as an MILP and then use a convex relaxation to improve solution times. In this Chapter we propose an alternative reformulation of the MPEC problem using multiparametric programming.

The general multi-leader follower problem is comprehensively discussed in [111] and [112], both providing examples of problem formulations, existence results, its relation to variational inequalites, and solution methodologies. For the specific case of strategic generation, the most complete work is [113], which uses EPECs to model bilevel games in restructured electricity markets, providing existence results in more simple cases without transmission constraints. In the more complex case they solve for local Nash equilibria, a weaker concept than the global generalized Nash equilibria we will study in this Chapter, using a diagonalization approach. In [114], an EPEC formulation is used, with a novel methodology to discover multiple equilibria by sequentially inserting 'holes' into the feasible set around previously discovered equilibria. Our methodology also finds all pure strategy generalized Nash equilibria of the problem.

The work in this Chapter is closest in spirit to [115], and to [116]. In [115] a single markup parameter for both the affine and quadratic bid parameters is proposed, and shown that this is sufficient to capture the optimal space of bid functions of this type. The authors propose an *Individual Welfare Maximization Algorithm* to solve both the single and multi-firm problems, which resembles diagonalization. The authors also demonstrate that the best repsonse function of a firm is not continuous in geeneral, meaning that the problem can have zero, one, or many equilibria. In [116] an *analytical* solution method to the multi-firm problem

is proposed, using similar parametric ideas to those proposed in this Chapter. However this approach involves complete enumeration of every active constraint set of the ISO problem, and only finds local Nash equilibria, with no solution of this approach guaranteed to be a global equilibrium. Our approach provides more structural insight into the problem and guarantees finding all pure strategy generalized Nash equilibria for a given problem instance. We additionally consider the witholding of generation capacity by strategic firms.

Novel Contributions

In this chapter we present the following novel contributions. In all cases for all classes of problem instance the methods detailed below can handle both strategic affine parameter bidding (a-bidding) and strategic capacity witholding, in the presence of transmission and generation capacity constraints.

- Multiparametric solution approach to the single-firm MPEC problem.
- Algorithm to exactly compute all pure strategy generalized Nash equilibria of the multi-firm EPEC problem, if any exist.
- A conjecture on the existence of mixed strategy equilibria along discontinuities of the multi-firm problem.

Organization

In Section 5.2 we present the problem formulation, Section 5.3 demonstrates the solution methodology for the single firm problem, Section 5.4 describes the multi-firm problem and an algorithm to compute its equilibria, Section 5.5 discusses some other interesting features of this class of problems, and Section 5.6 concludes.

5.2 Problem Formulation

We focus here on the strategic bidding problem of generation firms. This formulation builds off that described in Section 2.3. We consider a single period problem, *i.e.* $T = 1$, and as such we drop the subscript t for the rest of this chapter. We also assume there is no storage on the network. We begin by describing the participant model, then formulate the ED problem, and finally describe the participants' private optimization problems.

Generation Firms

We assume a set of generation firms $\mathcal{F} := \{1, \ldots, N\}$ firms, indexed by $f \in \mathcal{F}$, owning generation units at nodes $i \in \mathcal{N}_f \subset \mathcal{N}$, $f \in \mathcal{F}$. For simplicity, we assume that there is

at most one generation unit per node,[1], and that this unit is owned by a single firm, *i.e.* $\mathcal{N}_j \cap \mathcal{N}_k = \emptyset$, $\forall j,k \in \mathcal{F}$, $\bigcup_{f\in\mathcal{F}} \mathcal{N}_f = \mathcal{N}$. We denote a firm's generation vector $g \in \mathbb{R}^{n_f}$, where n_f is the number of generating units owned by firm f, and denote a generator-node assignment matrix for each firm as $\Lambda_f \in \mathbb{R}^{n\times n_f}$, where

$$[\Lambda_f]_{ij} := \begin{cases} 1, & \text{if the } j\text{th generating unit of firm } f \text{ is located at node } i \\ 0 & \text{otherwise} \end{cases}, \quad f \in \mathcal{F} \tag{5.1}$$

The nodal generation vector for the whole network, $g \in \mathbb{R}^n$ is recovered as $g = \sum_{f\in\mathcal{F}} \Lambda_f g_f$, and the firm generation vector can be recovered from the nodal generation vector as $g_f = \Lambda_f^\top g$

We now distinguish between the true cost function of a generation firm, and the cost function that they actually bid to the ISO. We denote the true cost function of a generation firm as $C_f(g_f)$ as notated previously, and the bid cost function as $\widehat{C}_f(g_f)$, where

$$C_f(g_f) = \frac{1}{2} g_f^\top Q_f g_f + a_f^\top g_f, \quad f \in \mathcal{F} \tag{5.2}$$

$$\widehat{C}_f(g_f) = \frac{1}{2} g_f^\top \widehat{Q}_f g_f + \widehat{a}_f^\top g_f, \quad f \in \mathcal{F} \tag{5.3}$$

where $Q_f \in \mathbb{R}^{n_f\times n_f}$, $\widehat{Q}_f \in \mathbb{R}^{n_f\times n_f}$, $Q_f, \widehat{Q}_f \succeq 0$, $a_f \in \mathbb{R}^n_f$, $\widehat{a}_f \in \mathbb{R}^n_f$, $a_f, \widehat{a}_f \geq 0$. In the case of linear generation costs, *i.e.* constant marginal cost, the quadratic term is omitted in both functions. In general there are no exogenous constraints on bid parameters, however many markets enforce price caps, which make sense in the case of purely linear costs, however are more difficult to characterise in the quadratic case. Additionally it is unlikely that generators would ever want to bid below their marginal cost as they would be losing money by default. As such, we define the following constraints on bid parameters

$$Q_f \preceq \widehat{Q}_f \preceq \overline{Q}_f, \quad f \in \mathcal{F} \tag{5.4}$$

$$a_f \leq \widehat{a}_f \leq \overline{a}_f, \quad f \in \mathcal{F} \tag{5.5}$$

where $\overline{Q}_f \in \mathbb{R}^{n_f\times n_f}$, $\overline{a}_f \in \mathbb{R}^{n_f}$ are arbitrary, finite, upper limits on the quadratic and affine bid cost parameters respectively. In the case of purely linear costs $\overline{a}_f$ would have the interpretation as the market price cap, and would be common across all firms, *i.e.* $\overline{a}_f = \overline{a}$, $f \in \mathcal{F}$.

The aggregate bid cost function for the network $\widehat{C}(g)$ is recovered as

$$\widehat{C}(g) = \frac{1}{2} g^\top \widehat{Q} g + \widehat{a}^\top g \tag{5.6}$$

where

$$\widehat{Q} = \sum_{f\in\mathcal{F}} \Lambda_f \widehat{Q}_f \Lambda_f^\top \tag{5.7}$$

$$\widehat{a} = \sum_{f\in\mathcal{F}} \Lambda_f \widehat{a}_f \tag{5.8}$$

[1]This can be achieved in practice by introducing dummy nodes into the network.

We assume that each firms generation is bounded by a convex generation set $\mathcal{G}_f$, where

$$\mathcal{G}_f := \{g_f : 0 \leq g_f \leq \overline{g}_f\}, \quad f \in \mathcal{F} \tag{5.9}$$

where $\overline{g}_f \in \mathbb{R}^{n_f}$ is the true vector of upper capacity limits of firm f's generating units. We also assume that generators may not truthfully reveal their actual capacity to the ISO, and may strategically withold some capacity from the market. We define the vector of upper capacity limits that is bid to the ISO by firm f as $\widehat{g}_f \in \mathbb{R}^{n_f}$, where $0 \leq \widehat{g}_f \leq \overline{g}_f$. The nodal generation upper capacity limit, $\widehat{g}$, is recovered as $\widehat{g} = \sum_{f \in \mathcal{F}} \Lambda_f \widehat{g}_f$. We denote the bid feasible set of generation as $\widehat{\mathcal{G}}$, where

$$\mathcal{G} := \{g : 0 \leq g \leq \widehat{g}\} \tag{5.10}$$

ED Problem

We consider a single period economic dispatch problem with elastic demand, employing the DC power flow model, parameterized by the bid parameters of generation, denoted $\mathbf{ED}(\widehat{a}, \widehat{Q}, \widehat{g})$.

$$\begin{aligned}
\mathbf{ED}(\widehat{a}, \widehat{Q}, \widehat{g}): \ \min_{g,d} \ & J(g, d, \widehat{a}, \widehat{Q}) && \text{(5.11a)}\\
\text{s.t. } & \gamma : \mathbf{1}^\top p = 0 && \text{(5.11b)}\\
& \beta : Hp \leq c && \text{(5.11c)}\\
& \lambda : p = g - d && \text{(5.11d)}\\
& g \in \widehat{\mathcal{G}} && \text{(5.11e)}\\
& d \geq 0 && \text{(5.11f)}
\end{aligned}$$

where $J(g, d, \widehat{a}, \widehat{Q}) = \widehat{C}(g) - B(d)$, and $B(d)$ is defined as in (2.28). Since the objective is convex, and constraints are continuously differentiable and affine, the KKT or equilibrium conditions of the problem are necessary and sufficient for optimality. We denote the optimal objective function, generation optimizers and LMPs as $J^\star(\widehat{a}, \widehat{Q}, \widehat{g})$, $g^\star(\widehat{a}, \widehat{Q}, \widehat{g})$, $\lambda^\star(\widehat{a}, \widehat{Q}, \widehat{g})$, respectively. We represent that these optimal solutions satisfy the equilibrium conditions of the problem as

$$g^\star(\widehat{a}, \widehat{Q}, \widehat{g}) \in \text{EQBM}(\mathbf{ED}(\widehat{a}, \widehat{Q}, \widehat{g})) \tag{5.12}$$

$$\lambda^\star(\widehat{a}, \widehat{Q}, \widehat{g}) \in \text{EQBM}(\mathbf{ED}(\widehat{a}, \widehat{Q}, \widehat{g})) \tag{5.13}$$

Private Profit Maximization of the Firm

Each generation firm seeks to maximize its payoff, the *first* level, subject to the equilibrium conditions of the ISO's ED problem, the *second* level. We will use x as a shorthand to represent all the decision parameters of all firms, *i.e.* $x = (\widehat{a}, \widehat{Q}, \widehat{g})$, and as a generic parameter

vector $x \in \mathbb{R}^{n_p}$, where n_p is the number of parameters. We represent the actions that relate to firm f, as x_f, and the actions that relate to all other firms as x_{-f}, such that $x = [x_f, x_{-f}]$, where this representation does not indicate any reordering of the vector x. We denote the payoff of a firm f as $\pi_f(x_f, x_{-f})$, and the private problem of a firm f is written as

$$\pi_f^\star(x_{-f}) = \max_{\lambda, g, x_f} \ \lambda_f^\top g_f - C_f(g_f) \tag{5.14a}$$

$$\text{s.t. } \lambda_f = \Lambda_f^\top \lambda \tag{5.14b}$$

$$g_f = \Lambda_f^\top g \tag{5.14c}$$

$$g \in \text{EQBM}(\mathbf{ED}(x_f, x_{-f}) \tag{5.14d}$$

$$\lambda \in \text{EQBM}(\mathbf{ED}(x_f, x_{-f}) \tag{5.14e}$$

$$x_f \in \mathcal{X}_f \tag{5.14f}$$

where $\mathcal{X}_f$ is a general feasible set representing (5.4), (5.5), and (5.10) for the firm f. As can be seen this problem is non-convex due to the inclusion of the equilibrium constraints, which would manifest as complementarity constraints if written out in full. In the subsequent sections we will address how to solve this problem for a single firm, and how to compute equilibria of the multi-firm problem, using techiques from multiparametric programming.

We will additionally make the assumption that each firm only optimizes one of its cost function parameters, *i.e.* only $\widehat{a}$, or $\widehat{Q}$. This is due to the fact, as we shall see in (6.38), that both the generation optimizers and LMPs have a bilinear dependence on $\widehat{a}$, or $\widehat{Q}$, which further complicates the problem. As such we will assume that each firm f optimizes only the affine term in its bid function $\widehat{a}_f$, and the quadratic term is fixed, with $\widehat{Q}_f = Q_f$. This means that, from here out $x = (\widehat{a}, \widehat{g})$. This is in contrast to [115] and [116], which use a single parameter for each firm multiplying both the affine and quadratic bid terms, but in line with [113], where this approach is reffered to as a-bidding.

5.3 Optimal Action of the Single Firm

We first consider the case where a single firm is attempting to optimize its actions, subject to the anticipated actions of other firms. We assume that these actions x_{-f} are constant and fixed. This problem is non-convex as stated above, and belongs to the MPEC class for which solution methods exist, discussed in [117]. We propose an alternative solution framework using multiparametric programming, and will show how these non-convex problems can be reduced to a finite set of QPs, or a single tractable MIQP, using similar concepts to those presented in [118]. The approach has both advantages and disadvantages relative to MPEC solution methods, but will also help us address the multi-firm equilibrium problem exactly in the next section. Using the analytical formulae, we will

From the results in Section 2.4, we can make the following deductions about the ED problem in (5.11), using the notation of that Section.

Theorem 5. *Solution properties of ED parametric on affine bid coefficients and generation capacities*

i *The feasible parameter set $\mathcal{K}^\star$ is a polyhedron.*

ii *The optimizer function $z^\star(x) : \mathcal{K}^\star \to \mathbb{R}^{2n}$ is:*

- *continuous, if the bid function is quadratic, i.e. $\widehat{Q} \succ 0$.*
- *discontinuous, if the bid function is affine, i.e. $\widehat{Q} = 0$.*
- *polyhedral piecewise affine (PPWA) over $\mathcal{K}^\star$. In particular it is affine in each critical region $\mathcal{CR}_j$, and every $\mathcal{CR}_j$ is a polyhedron.*

iii *The value function $J^\star(x) : \mathcal{K}^\star \to \mathbb{R}$ is:*

- *continuous*
- *polyhedral piecewise affine (PPWA) over $\mathcal{K}^\star$, in particular affine in each $\mathcal{CR}_j$, if the bid function is affine and demand is inelastic.*
- *polyhedral piecewise quadratic (PPWQ) over $\mathcal{K}^\star$, in particular quadratic in each $\mathcal{CR}_j$, if the bid function is quadratic and demand is elastic or inelastic, or if the bid function is affine and demand is elastic.*

iv *The LMP vector λ is a PPWA function over $\mathcal{K}^\star$, since it is an affine function of the dual variable function $u^\star(x) : \mathcal{K}^\star \to \mathbb{R}^m$ which is:*

- *discontinuous, if the bid function is affine, and demand is inelastic*
- *continuous, if the bid function is quadratic and demand is elastic or inelastic, or if the bid function is affine and demand is elastic.*
- *PPWA over $\mathcal{L}^\star$, in particular affine in each $\mathcal{CR}_j$.*

It should be noted that these results diverge slightly from those in Section 2.4, in particular with regard to the optimizer function, due to the fact here that we have the product of the parameter and the optimization variable in the objective function. The intuition for this difference is that one can imagine varying the affine cost coefficient of a generic LP, and at some point the optimal solution will jump to a different vertex of the feasible set. This results in the discontinous optimizer function in the case of affine costs.

These insights allow us to reformulate the generation optimizers and LMPs as PPWA functions over a set of critical regions. We assume that the ED problem has been solved parametrically, using the methods described in Section 2.4, over a feasible parameter set $\mathcal{K}^\star$, resulting in a set of full dimensional polyhedral critical regions, CR_j, $j = 1, \ldots, N_{CR}$. We define the $\mathcal{H}$ representation of the jth critical region as

$$CR_j := \{x : P_{j,x}x < P_{j,c}\}, \quad j = 1, \ldots, n_{CR} \tag{5.15}$$

where $P_{j,x}$ and $P_{j,c}$ are a matrix and vector respectively, of appropriate dimension, defining the bounding halfspaces of the polyhedral region CR_j. Each critical region CR_j has associated optimizer and dual variable functions such that

$$\lambda = F_j x + h_j, \text{ if } x \in CR_j, \ \forall j \tag{5.16}$$
$$g = R_j x + v_j, \text{ if } x \in CR_j, \ \forall j \tag{5.17}$$

where $F_j, R_j \in \mathbb{R}^{n \times n_p}$, $\forall j$, and $h_j, v_j \in \mathbb{R}^n$.

We can substitute these functions into (5.14) to obtain

$$\pi_f^\star(x_{-f}) = \max_{x_f} \ \lambda_f^\top g_f - C_f(g_f) \tag{5.18a}$$
$$\text{s.t. } \lambda_f = \Lambda_f^\top \lambda \tag{5.18b}$$
$$g_f = \Lambda_f^\top g \tag{5.18c}$$
$$\lambda = F_j x + h_j, \text{ if } x \in CR_j, \ \forall j \tag{5.18d}$$
$$g = R_j x + v_j, \text{ if } x \in CR_j, \ \forall j \tag{5.18e}$$
$$x_f \in \mathcal{X}_f \tag{5.18f}$$

We can substitute these expressions in and reformulate the profit function of firm f in the jth critical region as

$$\pi_{f,j}(x) = \lambda_f^\top g_f - \frac{1}{2} g_f^\top Q_f g_f - a_f^\top g_f \tag{5.19}$$
$$= \lambda^\top \Lambda_f \Lambda_f^\top g - \frac{1}{2} g^\top \Lambda_f Q_f \Lambda_f^\top g - a_f^\top \Lambda_f^\top g \tag{5.20}$$
$$\begin{aligned} &= x^\top \left(F_j^\top \Lambda_f \Lambda_f^\top R_j - \frac{1}{2} R_j^\top \Lambda_f Q_f \Lambda_f^\top R_j \right) x \\ &\quad + \left(h_j^\top \Lambda_f \Lambda_f^\top R_j + v_j^\top \Lambda_f \Lambda_f^\top F_j - v_j^\top \Lambda_f Q_f \Lambda_f^\top R_j - a_f^\top \Lambda_f^\top R_j \right) x \\ &\quad + \left(h_j^\top \Lambda_f \Lambda_f^\top v_j - \frac{1}{2} v_j^\top \Lambda_f Q_f \Lambda_f^\top v_j - a_f^\top \Lambda_f^\top v_j \right) \end{aligned} \tag{5.21}$$

It can be shown using a similar analysis to that in Section **??** that the quadratic term in (5.21) is negative semi-definite, and thus that $\pi_j(x)$ is concave, $\forall j$. We now observe that (5.18) can be recast as the solution of N_{CR} QPs.

$$\pi_f^\star(x_{-f}) = \max_{j=1,\ldots,N_{CR}} \left(\max_{x_f} \pi_j(x_f, x_{-f}) \ : \ x \in CR_j \right) \tag{5.22a}$$

Each QP is concave over a critical region of the ED problem, and the solution corresponds to the maximum payoff obtained among these QPs, with the optimal action being the corresponding optimal solution of the maximum payoff. Each QP can be solved individually and then compared, or the entire problem can be solved in a single optimization using epigraph

form and slack variables, or alternatively using integer variables and the big-M method to represent set membership, as will be seen in (5.28). Since each QP is concave, a maximum exists, however this may not be unique, meaning that the optimizer or optimal action is not necessarily unique. Additionally, since the critical regions are open sets, care must be taken if a solution is found which lies on a region boundary where the payoff function is discontinuous. We will see this in the following example.

Example

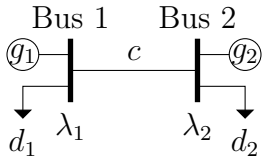


Figure 5.1: Two bus network

We consider a simple two bus network with a generator and load at each bus, as shown in Figure 5.1, with line capacity $c = 5$MW, and upper generation limits $\overline{g}_1 = 20$MW, $\overline{g}_2 = 10$MW. We assume that generators bid affine demand functions, such that $Q_1 = Q_2 = 0$. We will refer to this as *price-quantity bidding* or (P,Q) bidding. The true affine cost of generator 1 is $a_1 = 10$\$/MW. We assume that generator 2 bids a fixed quantity at a fixed price, such that $\widehat{a}_2 = 20$\$/MW, $\widehat{g}_2 = 10$MW. We also assume that demand is equally elastic at both nodes, such that $B(d_i) = 40d_i - d_i^2$. This is a sufficiently general case displaying all the typical behaviors of other examples. For illustration we let the market price cap $\overline{a}_1 = 30$\$/MW.

We first illustrate the multiparametric solution of ED for this example. The plots are in 3D and some have been rotated to better visualize the results, so refer to the axes when comparing across plots. In Figure 5.2 we observe the critical regions of the problem over the feasible parameter space, and the active constraints in each region are listed in Table 5.1. We also illustrate the generation optimizers in Figure 5.4, the demand optimizers in Figure 5.5, the line flow in the $1 \to 2$ direction in Figure 5.6, and the LMPs in Figure **??**.

Critical Region	$g_1 = \hat{g}_1$	$g_2 = \hat{g}_2$	$p_{12} = c$
1	1	0	0
2	1	1	0
3	1	0	1
4	0	1	0
5	0	0	1

Table 5.1: Active Constraints of Congestion Regions in the ED Problem

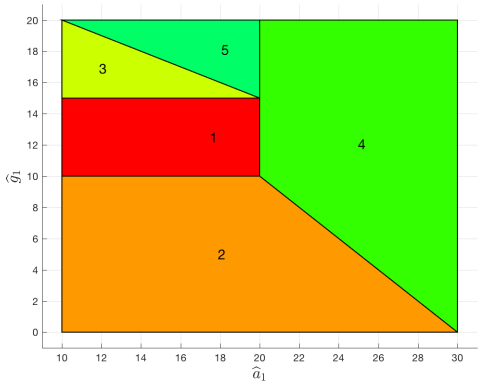


Figure 5.2: Critical Regions of ED for (P,Q) bidding

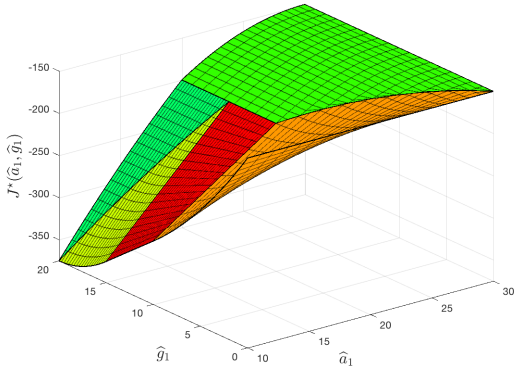
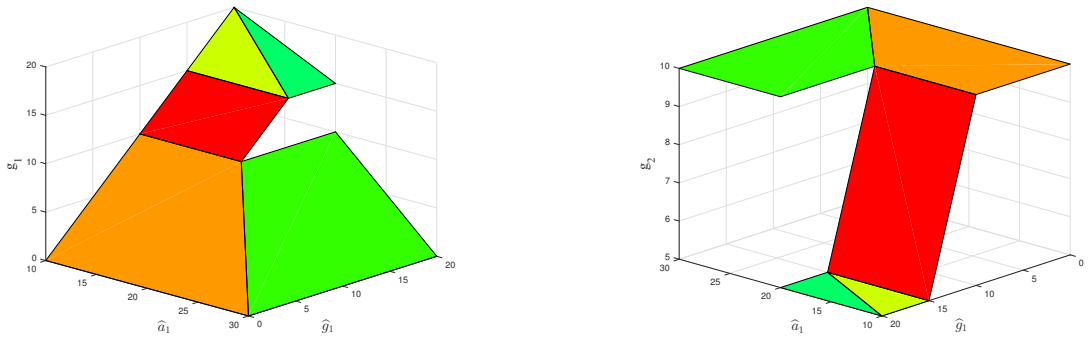


Figure 5.3: Value Function of ED over the feasible parameter space for (P,Q) bidding

We now consider the profit surface of generator 1 over the parameter space, (5.21), shown in Figure 5.8. We see that the function is PPWQ and discontinuous. We also observe that the optimal parameters are not unique and generator 1 maximizes its profit for $\widehat{a}_1 \in [10, 20]$, $\widehat{g}_1 = 15$. This maximum is attained on the boundary of critical regions 1 and 3, however the maximum is valid as the profit function is equal on both sides of the boundary. This boundary represents degeneracy in the transmission line constraint and does not affect the profits of the generator. This behavior where it is in the interest of generators to *marginally congest* transmission lines has been observed previously in [119]. It would appear that a maximum is also attained on $\widehat{a}_1 = 20$, $\widehat{g}_1 \in [15, 20]$, however this falls on the boundary between critical regions 4 and 5 where the profit function is discontinuous. This represents primal degeneracy where both generators bid the same marginal cost. If this occurs then in practice either demand would be split equally between generators, or a tie-breaking rule would be used. An acceptable bid for generator 1 would be $\widehat{a}_1 = 20 - \epsilon$, $\widehat{g}_1 \in [15, 20]$, however this results in a lower payoff than the maximum described previously. It is also interesting to compare Figure 5.8 to Figures 5.3 and 5.5, and see how the private incentives of the generator compare to the social optima. A set of profit maxima among which the generator is indifferent results in vastly different social costs and demand optimizers.

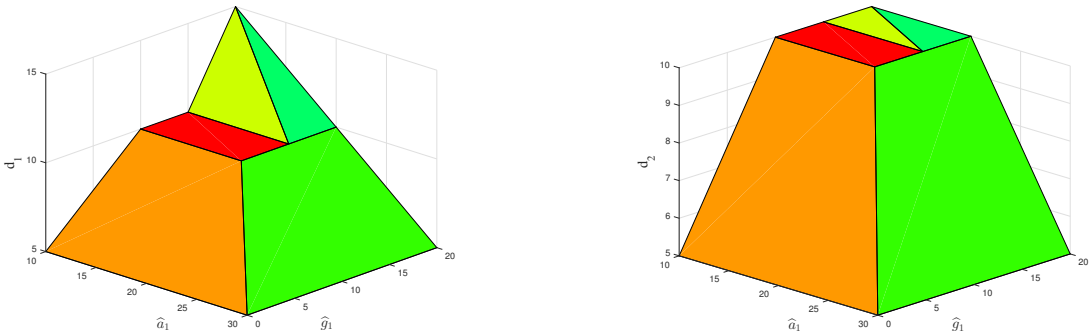
Tractability and Scalability

In the worst case, the number of critical regions is equal to the number of possible combinations of active sets of the optimization problem. However, in many problems only a few active constraints sets generate full-dimensional critical regions inside the feasible parameter set. To keep the number of critical regions low and computable within a reasonable time, it is desirable to keep the dimension of the parameter space as small as possible. Alternatively the volume of the parameter space can be kept 'small' by restricting the range of feasible



(a) g_1 (b) g_2

Figure 5.4: Generation optimizers of ED over the feasible parameter space for (P,Q) bidding



(a) d_1 (b) d_2

Figure 5.5: Demand optimizers of ED over the feasible parameter space for (P,Q) bidding

parameters in each dimension.

Relative to simply solving an MPEC, this parametric methodology has some advantages and disadvantages. It requires an intensive offline computation to calculate critical regions and associated functions, with a lightweight online computation to solve a relevant bilevel problem. This makes it useful for situations in which many instances of the problems need to be solved in quick succession, or in offline planning situations where computation time is less important. An advantage is that one automatically has the sensitivity functions dictating how the solution will change in response to deviations of the parameters. This is useful for real-time prediction and planning. This also enables the single firm to anticipate a range of actions from its competitiors and maximize their own profit in either a stochastic or robust manner. A shortcoming of this approach is that is somewhat 'brittle', requiring complete recomputation of all parametric solutions if any major changes occur on the system that are not modelled parametrically. For example, the generation cost function, or the network topology. These variables can be modelled parametrically but will result in an increase in

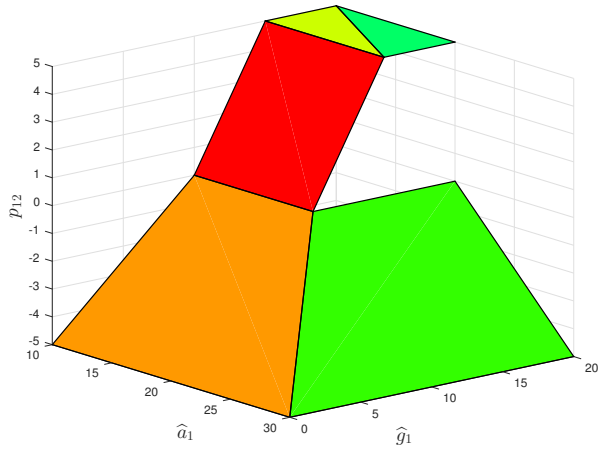
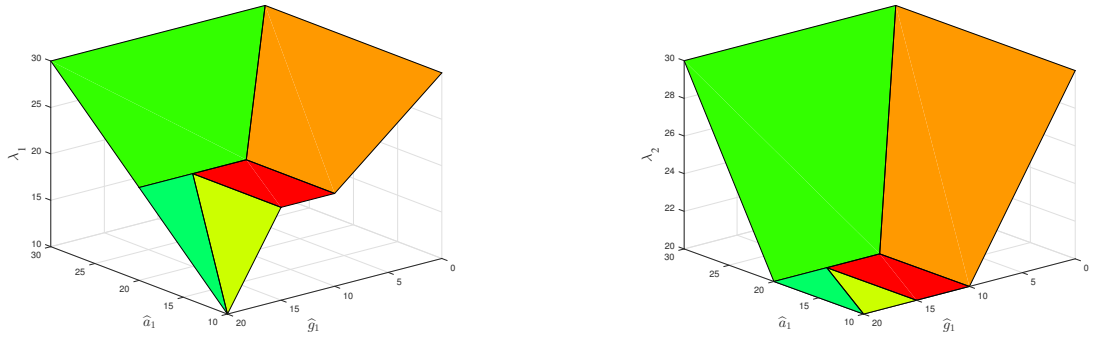


Figure 5.6: Line flow in the $1 \rightarrow 2$ direction over the feasible parameter space for (P,Q) bidding



(a) λ_1 (b) λ_2

Figure 5.7: LMPs of ED over the feasible parameter space for (P,Q) bidding

the dimension of the parameter set, incurring the scaling problems described previously. Depending on the intended use of the bilevel problems, operational or planning, a design tradeoff must be made between the set of variables that are modelled parametrically and the dimension and size of the parameter set.

5.4 Solving the Multi-Firm Problem

We have addressed the problem of a single firm's strategic bid, which can be thought of as the action of a monopolist. We now turn to the case of multiple strategic firms, which is a multi-leader follower game as described previously. The multi-leader follower game is

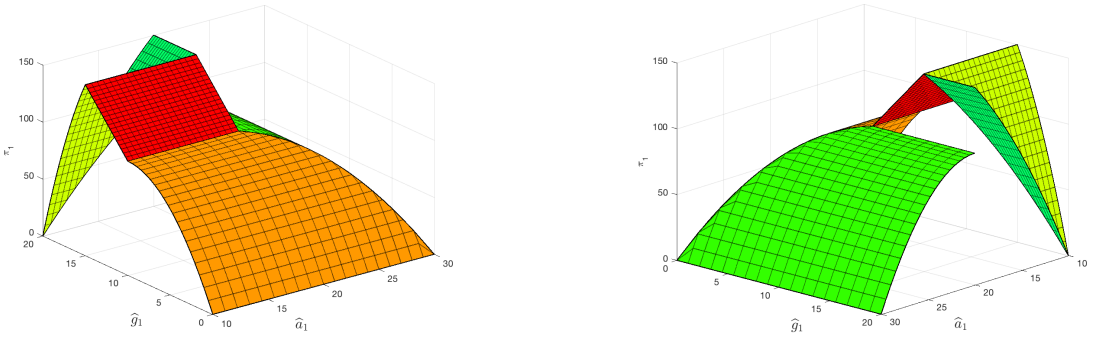


Figure 5.8: Profit surface for generator 1 over the feasible parameter space from two angles.

an example of a *Generalized Nash Game* [111], which extends the notion of a Nash Game by allowing the strategy sets of each player to depend on the actions of other players. For completeness we first state the definition of a Nash game and an existence theorem for its equilibrium, and then the definition of a generalized Nash game and an existence theorem for its equilibrium.

We will maintain the notation used heresofar in this Chapter, which deviates a little from the literature, referring interchangably between firms and *players*.[2]. We follow the definitions in [120]. Let N be the number of players. The strategy set of player f is denoted by $\mathcal{X}_f \subset \mathbb{R}^{n_f}$ and their payoff function by $\pi_f : \mathcal{X} \to \mathbb{R}$ (to be maximized), where $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_N$. Player f's (pure) strategy is denoted by $x_f \in \mathcal{X}_f$ while $x_{-f} \in \mathcal{X}_{-f}$ denotes the other players' action, *i.e.* $x_{-f} = (x_1, \ldots, x_{f-1}, x_{f+1}, \ldots, x_N)$ and $\mathcal{X}_{-f} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_{f-1} \times \mathcal{X}_{f+1} \times \cdots \times \mathcal{X}_N$. A game is thus described by $(N, \mathcal{X}_f, \pi_f(\cdot))$.

Definition 4. *A Nash equilibrium is a strategy point $x^\star \in \mathcal{X}$ such that no player has an incentive to deviate, i.e. for all $f \in \{1, \ldots, N\}$,*

$$\forall x_f \in \mathcal{X}_f, \pi_f(x_f, x^\star_{-f}) \leq \pi_f(x^\star_f, x^\star_{-f}) \tag{5.23}$$

An existence theorem for infinite games from [121], reported in [120] is presented here.

Theorem 6. *Let N agents be characterized by an action space $\mathcal{X}_f$ and an objective function π_f. If $\forall f \in \{1, \ldots, N\}$, $\mathcal{X}_f$ is nonempty, convex and compact; $\pi_f : \mathcal{X} \to \mathbb{R}$ is continuous with $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_N$ and $\forall x_{-f} \in \mathcal{X}_{-f}$, $x_f \to \pi_f(x_f, x_{-f})$ is concave on $\mathcal{X}_f$, then there exists a Nash equilibrium.*

Turning to the generalized Nash game, we let $2^{\mathcal{X}_f}$ be the family of subsets of $\mathcal{X}_f$. We let $C_f : \mathcal{X}_{-f} \to 2^{\mathcal{X}_f}$ be a constraint correspondence of player f, *i.e.* a function mapping a point in $\mathcal{X}_{-f}$ to a subset of $\mathcal{X}_f$. Thus, $C_f(x_{-f})$ defines the fth player's action space given

[2]In the literature players are generally indexed with i, and the payoff function of player i is generally denoted $\theta_i(\cdot)$.

other players' actions x_{-f}. A generalized game is described by $(N, \mathcal{X}_f, C_f(\cdot), \pi_f(\cdot))$. where N is the number of players, $\mathcal{X}_f$ the strategy set of player f, and $\pi_f(\cdot)$ the payoff function of player f.[3]

Definition 5. *The generalized Nash equilibrium (GNE) for a generalized game $(N, \mathcal{X}_f, C_f(\cdot), \pi_f(\cdot))$, is defined as a point $x^\star$ solving for all $f \in \{1, \ldots, N\}$*

$$x_f^\star \in \underset{x_f \in C_f(x^\star_{-f})}{\arg\max}\ \pi_f(x_f, x_{-f}) \tag{5.24}$$

We present an existence theorem for a GNE from [122], reported in [120].

Theorem 7. *Let N players be characterized by an action space $\mathcal{X}_f$, a constraint correspondence C_f, and an objective function $\pi_f : \mathcal{X} \to \mathbb{R}$. Assume for all players, we have*

i $\mathcal{X}_f$ is nonempty, convex and compact subset of a Euclidean space,

ii C_f is both upper-semi continuous and lower-semi continuous in $\mathcal{X}_{-f}$,

iii $\forall x_{-f} \in \mathcal{X}_{-f}$, $C_f(x_{-f})$ is nonempty, closed, convex,

iv π_f is continuous on the graph $Gr(C_f)$,

v $\forall x \in \mathcal{X}, x_f \to \pi_f(x_f, x_{-f})$ is quasiconcave on $C_f(x_{-f})$,

Then there exists a generalized Nash equilibrium.

Further existence theorems for generalized Nash Equilibria can be found in [120] using alternative formulations of the generalized game.

In our case, this dependence of the action set of a player on the actions of the other players, while not immediately obvious from the single-firm problem in (5.18), arises due to the complementarity constraints introduced in (5.14). As noted previously, the single-firm problem is an example of an MPEC. To find the equilibrium among all firms, one must simultaneously solve a number of MPECs, leading to an instance of an *Equilibrium Problem with Equilibrium Constraints* (EPEC). This is noted in [111] as "a class of mathematical programs of significant difficulty in general". The authors of [111] note that this results in a Nash problem with nonconvex subgame equilibria. This means that the resulting Nash equilibrium may not exist, due to the nonconvexity in each player's problem; and that often such a nonconvex Nash game is computationally intractable.

We present a methodology to compute all pure equilibria of the multi-firm problem, and identify if no pure equilibria exist. We cannot say as much for mixed equilibria, however we will point to a specific example of a mixed equilibria which gives an interesting direction for further study.

[3] This is also referred to as an abstract economy in reference to Debreu's work in Economics.

Optimal Reaction Set

We denote the *optimal reaction set* of a firm f, as $r_f(x_{-f}) : \mathcal{X}_{-f} \to \mathcal{X}_f$, which constitutes the optimal action of player f for a given action x_{-f} of the other players. An alternative characterization of the generalized Nash equilibrium is given as a point $x^\star$, solving for all $f \in \{1, \ldots, N\}$

$$x_f^\star \in r_f(x_{-f}^\star) \tag{5.25}$$

We can compute the optimal reaction set for a firm by solving its corresponding single-firm problem parametrically on the actions of other players. As we shall see this consitutes solving a *Multiparametric Mixed-Integer Quadratic Program* (mpMIQP), arguably one of the most challenging classes of multiparametric program.

We also note here that a necessary but insufficient condition for a GNE is the first order optimality condition of the single-firm problems. That is for $x^\star$ to be a GNE in pure strategies

$$\frac{\partial \pi_f}{\partial x_f}(x^\star) = 0, \; \forall f = 1, \ldots, N \tag{5.26}$$

Additionally, we note a necessary but insufficient condition for the GNE to be stable. $x^\star$ is a stable GNE in pure strategies if

$$\frac{\partial^2 \pi_f}{\partial x_f^2}(x^\star) \leq 0, \; \forall f = 1, \ldots, N \tag{5.27}$$

We formulate the problem of a single strategic bidder as a *Mixed-Integer Quadratic Program* (MIQP). We assume that its bid function is quadratic, and thus its payoff function is PPWQ over N_{CR} critical regions in $\mathcal{X}$, and continuous. This can be seen by considering Theorem 5, and seeing that its payoff function is the composition of a continous function and the product of two continuous functions.[4] The following MIQP problem has the bids of other participants x_{-f} as a parameter, and the optimal solution is the rational response $x_f^\star = r_f(x_{-f})$ for participant f and the optimal payoff is $\pi_f^\star(x_{-f})$. We use the big-M method

[4] In the case of linear generation costs and bids, and inelastic demand, the problem forms an MILP and can be solved using techniques from Section 7.4 of [38]. The case of linear bids and elastic demand is more complicated as this is an MIQP with a discontinuous cost. It cannot be assumed to be lower-semi continuous and thus a maximum does not necessarily exist. We can get around this if we ignore solutions on critical region boundaries for now, and say that a maximum exists almost everywhere. We will pursue this idea further in Section 5.5.

to reformulate set membership constraints with integer variables.

$$\begin{aligned}
\pi_f^\star(x_{-f}) = \max_{x_f,w,u}\ & \mathbf{1}^\top w && \text{(5.28a)}\\
\text{s.t. } & x_f \in \mathcal{X}_f && \text{(5.28b)}\\
& \pi_{f,j}(x_f, x_{-f}) + M(1-u_j) \geq w_j, \quad \forall j && \text{(5.28c)}\\
& P_{j,x}[x_f^\top x_{-f}^\top]^\top - M(1-u_j) \leq P_{j,c}, \quad \forall j, && \text{(5.28d)}\\
& 0 \leq w \leq Mu && \text{(5.28e)}\\
& \mathbf{1}^\top u = 1 && \text{(5.28f)}\\
& u_j \in \{0,1\}, \quad \forall j && \text{(5.28g)}
\end{aligned}$$

where $w \in \mathbb{R}^{N_{CR}}$ is a vector of slack variables, $u \in \mathbb{R}^{N_{CR}}$ is a vector of binary variables, and $M \in \mathbb{R}$ is a large scalar. Each slack and binary variable is associated with a critical region, and this formulation ensures that only one critical region and its associated payoff function is feasible at the optimal solution.

This MIQP is then solved parametrically on x_{-f} over $\mathcal{X}_{-f}$, giving an mpMIQP. The mpMIQP can be solved using techniques described in Chapter 18 of [38], in particular Theorem 18.1, Section 18.6.1, Example 18.4, and Lemma 18.4. This problem instance is an example of a (*one-to-r problem*), as described in Lemma 18.4 of [38], where r refers to the number of critical regions over which the original problem (5.28) is defined. In other words (5.28) can be reformulated as

$$\pi_f^\star(x_{-f}) = \max \left\{ \begin{array}{l} \max\limits_{x_f} \pi_{f,1}(x_f, x_{-f}) \\ \text{s.t. } x \in CR_1 \\ \vdots, \\ \max\limits_{x_f} \pi_{f,N_{CR}}(x_f, x_{-f}) \\ \text{s.t. } x \in CR_{N_{CR}} \end{array} \right\} \tag{5.29}$$

As stated in this Lemma, the mpMIQP can be solved by N_{CR} mpQPs. This involves solving a separate mpQP for each critical region j of the original single-firm problem of firm f, which will result in a new set of critical regions $CR_{f,j,k} \subset CR_j$, $k = 1, \ldots, N_{CR_{f,j}}$, and $CR_{f,j,k} \subset \mathcal{X}_{-f}$. From the results on mpQPs in Section 2.4, we see that the objective function $\pi_{f,j}$ is PPWQ, in particular quadratic in each $CR_{f,j,k}$, and the optimizer function $x_f^\star$ is PPWA, in particular affine in each $CR_{f,j,k}$. Having solved all N_{CR} mpQPs for firm f, it may be the case that some of the resulting critical regions intersect, so that is not clear which optimizer function to use in this intersection. This conflict is resolved by examining the respective objective functions of each intersecting critical region over the intersection. In the region where one objective function is greater, its associated optimizer function is used in this region. This can be done in a principled way for any number of intersecting critical regions to create an *ordered region single valued function*, as per Definition 18.1 of [38]. This can of course result in regions which are non-euclidean, for example with boundaries defined

by the intersections of quadratic functions. In practice this is not an issue as Section (iii) of Example 18.4 in [38] details a way to avoid the storage of non-convex and non-euclidean regions. The presence of non-euclidean regions means that the optimal reaction function is not PPWA, merely piecewise-affine (PWA). In addition the optimal reaction function is not continuous in general, as was observed in [115].

Following this procedure we are left with a new set of critical regions associated with the parametric solution of the single-firm problem of firm f, which we will denote $\mathcal{R}_{f,j}$, $j = 1, \ldots, N_f$. Associated with each region is an affine optimal reaction function

$$x_f^\star = F_{f,j} x_{-f} + h_{f,j}, \quad x_{-f} \in \mathcal{R}_{f,j},\ j = 1, \ldots, N_f \tag{5.30}$$

where $F_{f,j} \in \mathbb{R}^{n_f \times n_{-f}}$, $\forall j$, and $h_{f,j} \in \mathbb{R}^{n_f}$, $\forall j$. We repeat this procedure for all firms $f \in \mathcal{F}$, and have the following Theorem

Theorem 8. *If $\pi_f(x_f, x_{-f})$ is a continuous PPWQ function $\forall f = 1, \ldots, N$, then the optimal reaction function, $x_f^\star = r_f(x_{-f})$, is a PWA function over $\mathcal{X}_{-f}$, in particular affine within each region $\mathcal{R}_{f,j}$, $j = 1, \ldots, N_f$. That is*

$$x_f^\star = F_{f,j} x_{-f} + h_{f,j},\ \textit{if}\ x_{-f} \in \mathcal{R}_{f,j},\ \forall j = 1, \ldots, N_f,\ \forall f = 1, \ldots, N \tag{5.31}$$

Then, a necessary condition for a GNE in pure strategies, $x^\star$ is that

$$x^\star = \mathbf{F} x^\star + \mathbf{h} \tag{5.32}$$

or, solving for $x^\star$,

$$x^\star = (\mathbf{I} - \mathbf{F})^{-1} \mathbf{h} \tag{5.33}$$

where

$$\mathbf{F} = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_N \end{bmatrix} \in \mathbb{R}^{\sum_f n_f \times \sum_f n_f}, \quad \mathbf{h} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_N \end{bmatrix} \in \mathbb{R}^{\sum_f n_f} \tag{5.34}$$

where for $x_{-f}^\star \in \mathcal{R}_{f,j}$, $F_f = [\mathbf{0}_f, F_{f,j}]$, where $\mathbf{0}_f \in \mathbb{R}^{n_f \times n_f}$, represents a matrix of zeros in the fth position, and $h_f = h_{f,j}$.

Proof. This follows from the results detailed above for mpQPs from multiparametric programming. We need to find a point $x^\star$ that is a fixed point of the equations

$$x_f^\star = F_{f,j} x_{-f} + h_{j,f},\ \text{if}\ x_{-f} \in \mathcal{R}_{f,j},\ \forall j = 1, \ldots, N_f,\ \forall f = 1, \ldots, N \tag{5.35}$$

This is satisfied if (5.33) holds for $x^\star$. Such a fixed point is a GNE since it constitutes the optimal response of any player to the vector of actions played by the other players. □

Computing Equilibria

For each firm f we now have a look-up table of affine functions (5.30), associated to a set of regions, defined over $\mathcal{X}_{-f}$. Rather than as affine functions, we can think of these functions as defining lower-dimensional regions in $\mathcal{X}$ space. We denote these regions as $\overline{\mathcal{R}}_{f,j}$, and they are defined as

$$\overline{\mathcal{R}}_{f,j} := \{x : x_f = F_{f,j}x_{-f} + h_{f,j},\ x_{-f} \in \mathcal{R}_{f,j}\}, \quad j = 1, \ldots, N_f,\ f = 1, \ldots, N \tag{5.36}$$

It can be seen that a fixed point, as defined in Theorem 8, if it exists, constitutes the intersection of N of these regions, one from each of $f = 1, \ldots, N$. Clearly then computing any and all equilibria of the multi-firm problem consists of finding any and all intersections of the sets $\overline{\mathcal{R}}_{f,j}$, $f = 1, \ldots, N$, $j = 1, \ldots, n_f$.

In the worst case, if we assume that for each firm $n_f = r$, then this computation would involve checking r^N intersections of these sets. Being exponential in the number of firms is clearly undesirable! However, in general, many of these combinations can be discarded, as they are disproved as potential equilibria by checking the intersections of other sets. This can be seen by considering the fact that if N sets have non-zero intersection, then all subsets of these N sets consisting of at least two or more sets must also have non-zero intersection. This motivates the recursive algorithm, Algorithm 1, to calculate any and all equilibria in pure strategies of the multi-firm problem, where **Intersect**$(\mathcal{A}, \mathcal{B})$ is a function computing the intersection of two sets $\mathcal{A}, \mathcal{B}$.

A potential improvement to this algorithm comes from noticing that each $\overline{\mathcal{R}}_{f,j}$ is a subset of at most one critical region of the original ED problem CR_k. We therefore need only check intersections of sets which belong to the same original critical region, or of sets which belong to neighboring original critical regions, *i.e.* they could intersect at the boundary of the original critical regions. Each critical region and its neighbors could be processed in a principled and sequential manner to minimize the number of intersections that would be necessary to consider.

The algorithm as stated will find all pure strategy GNEs for the multi-firm problem. These intersections may be singletons, representing a unique equilibrium, or indeed a set, representing an infinity of equilibria.

Examples

We now provide two examples employing this approach. For ease of visualization, we restrict our attention to the two bus example, with two strategic firms, a duopoly. We observe that in both the symmetric and asymmetric case, for elastic demand, both problems exhibit a unique pure strategy Nash equilibrium. In these examples we will demonstrate the asymmetric case.

We consider a two bus network with a single generating firm at each bus, each with a single unit, as shown in Figure 5.1. We assume symmetric elastic loads at each bus, with $B(d_1) = B(d_2) = 40d_i - d_i^2$. We assume assymetric generation cost functions, with $Q_1 = Q_2 = 1\$/\text{MWh}^2$, $a_1 = 10\$/\text{MWh}$, and $a_2 = 20\$/\text{MWh}$. We assume equal generation

Data: N firms, with $\{N_f\}_{f=1}^{N}$ regions
Result: $\mathcal{I}_{store}$, is the set of all intersections of the sets $\overline{\mathcal{R}}_{f,j}$, $\forall j$, $forall f$.
Initialization: $\mathcal{I} = \emptyset$, $\mathcal{I}_{\mathbf{store}} = \emptyset$, $f = 1$;
MFE$(\mathcal{I}, f)$
begin
  for $j = 1, \ldots, N_f$ **do**
    if $f = 1$ **then**
      $\mathcal{I} = \overline{\mathcal{R}}_{f,j}$
    end
    $\mathcal{A} =$ **Intersect**$(\mathcal{I}, \overline{\mathcal{R}}_{f,j})$
    if $\mathcal{A} = \emptyset$ **then**
      break
    else
      if $f = N$ **then**
        $\mathcal{I}_{\mathbf{store}} = \mathcal{I}_{\mathbf{store}} \cup \{\mathcal{A}\}$
      else
        MFE$(\mathcal{A}, f + 1)$
      end
    end
  end
end

Algorithm 1: Computation of Multi-Firm Equilibria

capacity, such that $\overline{g}_1 = \overline{g}_2 = 20$MW. The transmission line has a capacity constraint $c = 5$MW, and we set an arbitrary price cap of $\overline{a}_i = 50$\$/MW.

Duopoly Price Competition

We first examine the case when both firms compete on price with an equal fixed capacity bid. The fixed quantity bids were set at $\widehat{g}_1 = \widehat{g}_2 = 20$MW.

We first show the set of critical regions and objective function over the feasible parameter set, in Figures 5.9 and 5.10 respectively. We now examine the profit function surface for each producer in Figure 5.11, and observe as expected, that it is continuous, and concave in each critical region. Note that some figures have been rotated for a better angle of the 3D surface, so refer to axes when comparing across figures. Following the analysis of the previous section we can compute the optimal payoff functions and best response functions of each firm, parametric on the action of the other firm, seen in Figures 5.12 and 5.13 respectively. In this case we see that the best response functions are discontinuous, and intersect at $x^\star = (15.5, 22.5)$, meaning that we have a GNE in pure strategies.

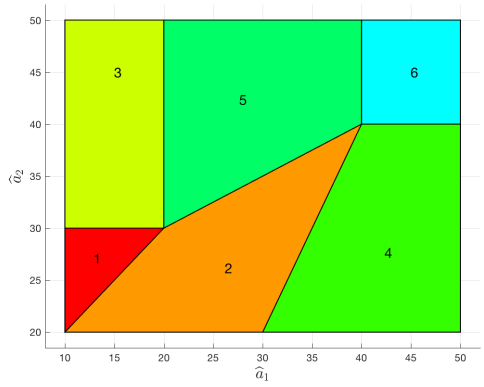


Figure 5.9: Critical Regions of ED for asymmetric price competition

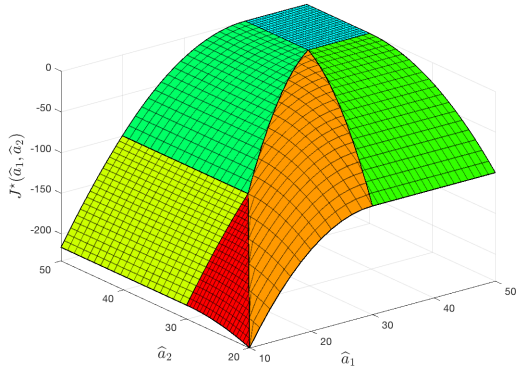
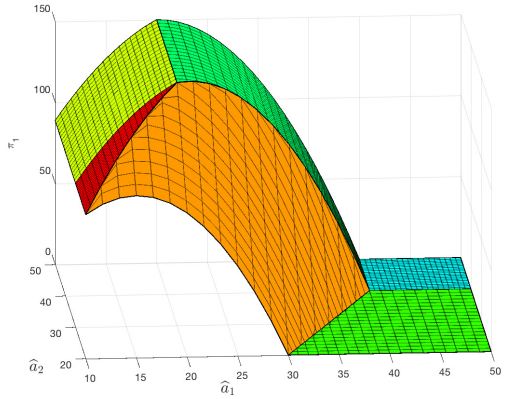
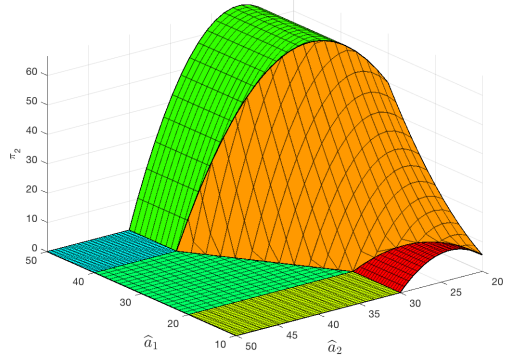


Figure 5.10: Value Function of ED for asymmetric price competition



(a) π_1



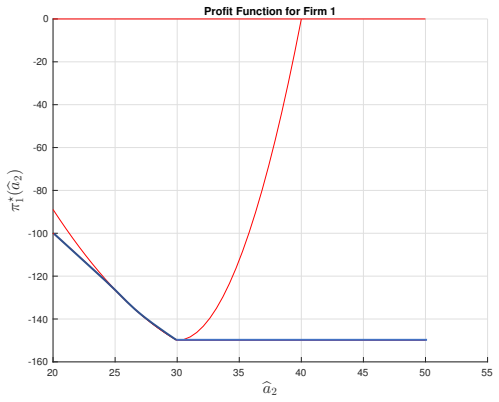
(b) π_2

Figure 5.11: Profit functions of individual producers for asymmetric price competition

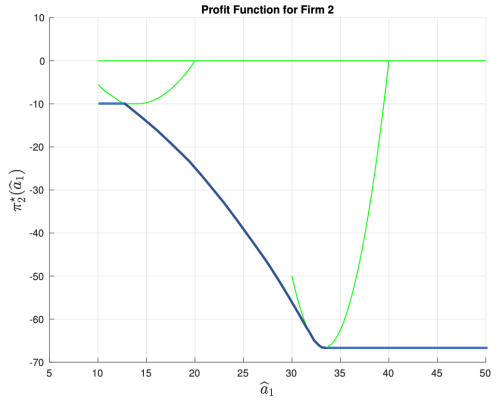
Duopoly Quantity Competition

We now consider the case where the firms compete on quantity with fixed price bids. The situation is eactly the same as before, except now $\widehat{g}_f$ is the decision parameter of firm f. We let $\widehat{a}_1 = 10\$/\text{MWh}$, and $\widehat{a}_2 = 20\$/\text{MW}$, equal to the generators' marginal cost.

We first show the set of critical regions and objective function over the feasible parameter set, in Figures 5.14 and 5.15 respectively. We now examine the profit function surface for each producer in Figure 5.16, and observe as expected, that it is continuous, and concave in each critical region. Note that some figures have been rotated for a better angle of the 3D surface, so refer to axes when comparing across figures. Following the analysis of the previous section we can compute the optimal payoff functions and best repsonse functions of each firm, parametric on the action of the other firm, seen in Figures 5.17 and 5.18 respectively. In this

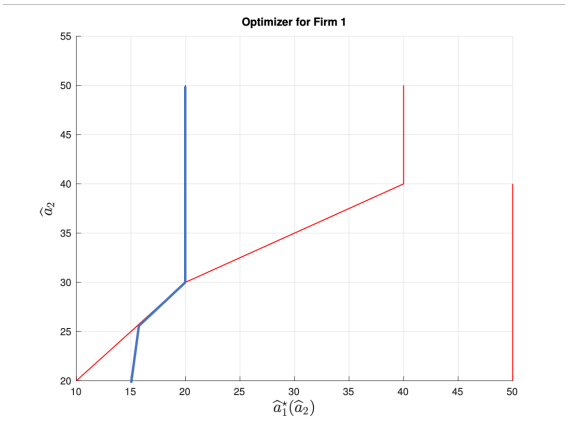


(a) $\pi_1^\star(\widehat{a}_2)$

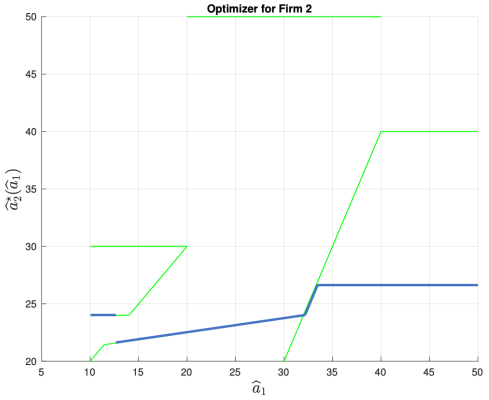


(b) $\pi_2^\star(\widehat{a}_1)$

Figure 5.12: Optimal (negative) payoff of each firm, in blue, parametric on the action of the other firm. Payoffs in other critical regions shown in red and green.



(a) $r_1(\widehat{a}_2)$



(b) $r_2(\widehat{a}_1)$

Figure 5.13: Best response function of each firm in blue. Sub-optimal actions from other critical regions shown in red and green.

case we see that the best response function of Firm 1 is continuous, and the best response function of Firm 2 is discontinuous. The two functions intersect at $x^\star = (8.8, 3.6)$, meaning that we have a GNE in pure strategies.

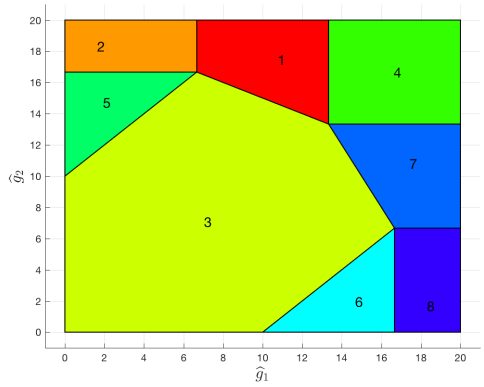


Figure 5.14: Critical Regions of ED for asymmetric capacity competition

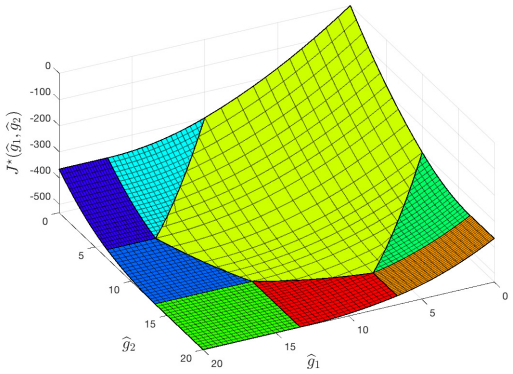
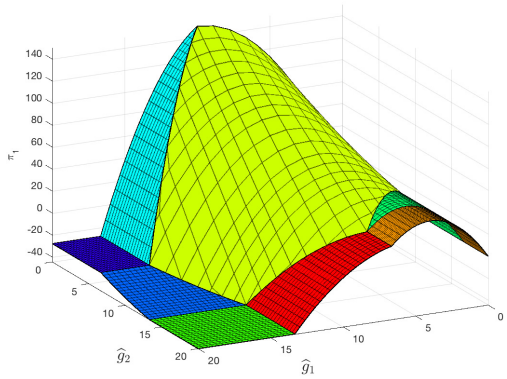
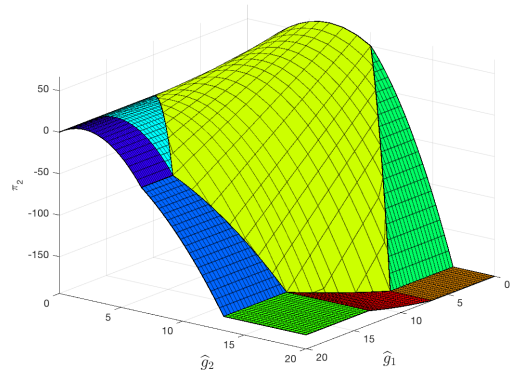


Figure 5.15: Value Function of ED for asymmetric capacity competition



(a) π_1



(b) π_2

Figure 5.16: Profit functions of individual producers for asymmetric capacity competition

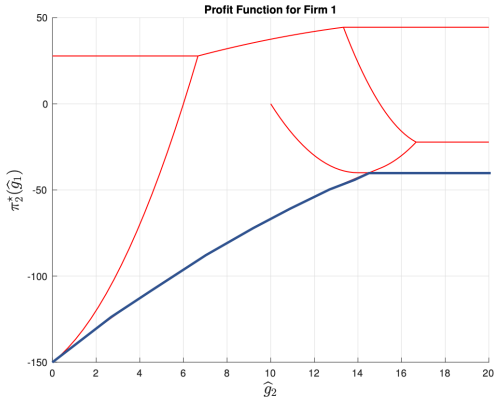
5.5 Additional System and Equilibrium Phenomena

Best Response as a Hybrid Linear System

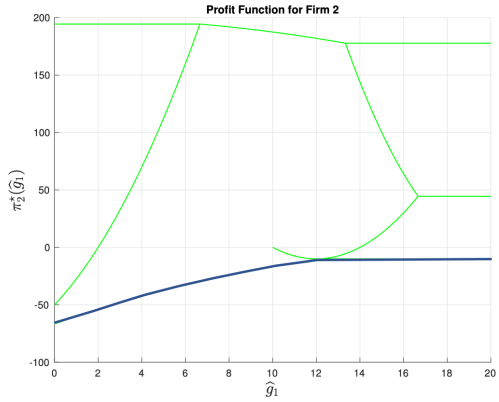
We consider the dynamical system defined by the best response functions of each firm, where we index each iteration by k, and the initial action of all agents is given by $x(0)$. This represents the situation where firms play a repeated game, using their best reponse to the previous action of the other firms as their action at the next iteration.

$$x_f(k+1) = r_f(x_{-f}(k)), \quad \forall f \tag{5.37}$$

By examining (5.30) and (5.36) we see that the best response dynamics form a hybrid linear system, more specifically a PWA system, and equilibria of this system correspond to

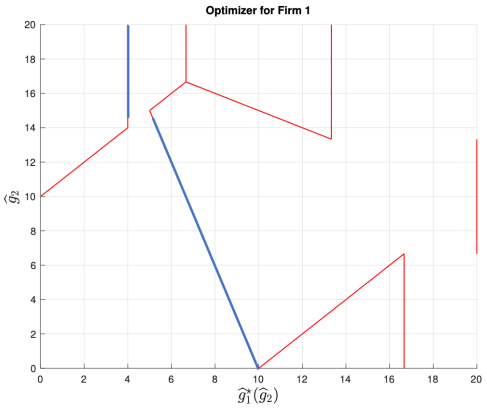


(a) $\pi_1^\star(\widehat{a}_2)$

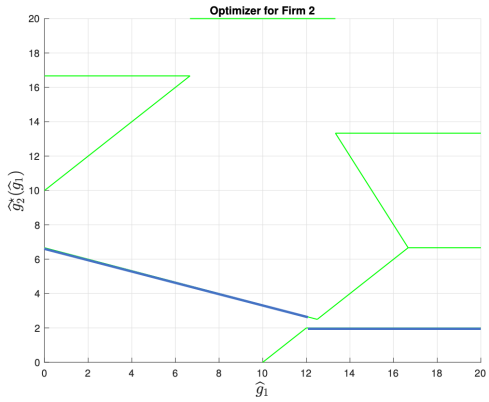


(b) $\pi_2^\star(\widehat{a}_1)$

Figure 5.17: Optimal (negative) payoff of each firm, in blue, parametric on the action of the other firm. Payoffs in other critical regions shown in red and green.



(a) $r_1(\widehat{a}_2)$



(b) $r_2(\widehat{a}_1)$

Figure 5.18: Best response function of each firm in blue. Sub-optimal actions from other critical regions shown in red and green.

generalized Nash equilibria of the multi-firm problem. More explicitly

$$x_f(k+1) = F_{f,j}x_{-f}(k) + h_{f,j}, \quad \textbf{if } x_{-f}(k) \in \overline{\mathcal{R}}_{f,j}, \quad \forall f \tag{5.38}$$

This is interesting as an aside, but it also has implications for the computation of equilibria in the literature. As mentioned previously, diagonalization is a very common technique for computing equilibria in the literature, where the computation effectively follows the best response dynamics. For diagonalization to be successful, it requires that the best response dynamics be stable. This will generally not be the case, as the stability of PWA systems is notoriously difficult to guarantee. Even the interconnection of two stable systems is not stable in general. The best that can be hoped for in general is a characterization of 'local stability'

in some region of the state space. This casts doubt on the effectiveness and usefulness of diagonalization as a technique for computing equilibria.

Discontinuities and Mixed Equilibria

We finish with a final example that illustrates an interesting phenomenon in the case of affine bid functions. From Theorem 5, we see that if the bid function is affine, the optimizer function will be discontinuous in general, and thus the payoff function of each generation firm will also be discontinuous in general. We present an example of this scenario here. We assume symmetric generation, with identical affine costs and affine bids. Here $Q_1 = Q_2 = 0$, $a_1 = a_2 = 10\$/\text{MWh}$, $\overline{g}_1 = \overline{g}_2 = 20\text{MW}$. We assume symmetric elastic demand with the same demand function as the previous two examples. We assume that generators strategically bid only their affine costs, and submit fixed capacity bids equal to their true capacities.

We first show the set of critical regions and objective function over the feasible parameter set, in Figures 5.19 and 5.20 respectively. We now examine the profit function surface for each firm in Figure 5.21, and observe as expected, that it is discontinuous, concave in each critical region, and symmetric for each firm. The discontinuity lies along the line $\widehat{a}_1 = \widehat{a}_2$, and reflects primal degeneracy in the ISO problem. We now assume that the critical regions are in fact closed, so we allow the region boundary to be part of the best response calculation, and we see these in Figure 5.23. We see that the best response functions track and intersect along the discontinuity, *i.e.* if there was a tie-breaking rule which favored one generator over the other in the case of primal degeneracy, then this would in fact be their optimal action.

Following the work of [123] and [124], we observe that this problem resembles a Bertrand-Edgeworth game, *i.e.* a game of Bertrand price competition with finite supply capacity and elastic demand. In [123], and extended to a weaker set of conditions in [124], it is proved that a mixed strategy Nash equilibria exists in these games along the discontinuity where $\widehat{a}_1 = \widehat{a}_2$, if a well defined payoff is introduced along the discontinuity. For example, the ISO could split demand evenly between firms if such primal degeneracy were to occur. More work is needed to verify this conclusion in this specific case, but we make the following conjecture.

Conjecture *If $\pi_f(x_f, \mathbf{x}_{-f})$ is a discontinuous PPWQ function over a set of critical regions, $\forall f$, and the discontinuity is identical $\forall f$, then a mixed strategy generalized nash equilibrium of the multi-firm game exists along the discontinuity.*

This conjecture says nothing about actually calculating what the mixed strategy would be, and [123] and [124] only address existence results in their papers. Proving this existence conjecture for the specific case of affine bid functions with elastic demands remains the preserve of future work.

5.6 Conclusions

In this Chapter we have examined the strategic generation bidding problem on constrained transmission networks. We have used techniques from multiparametric programming to pro-

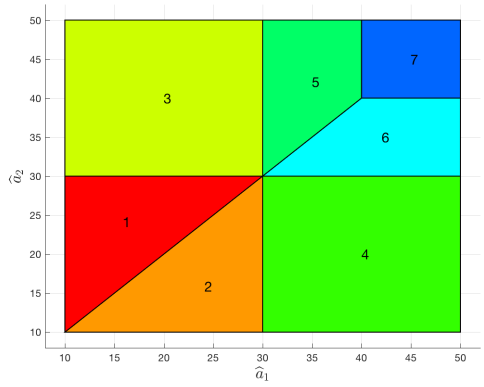


Figure 5.19: Critical Regions of ED for symmetric linear price competition

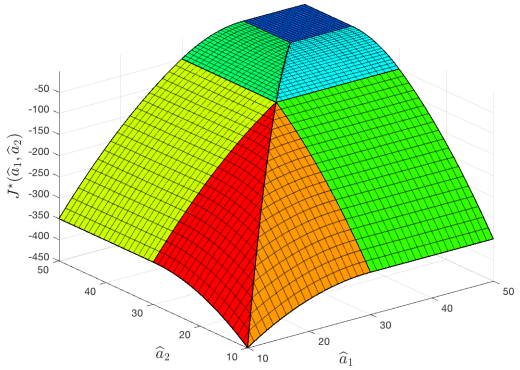
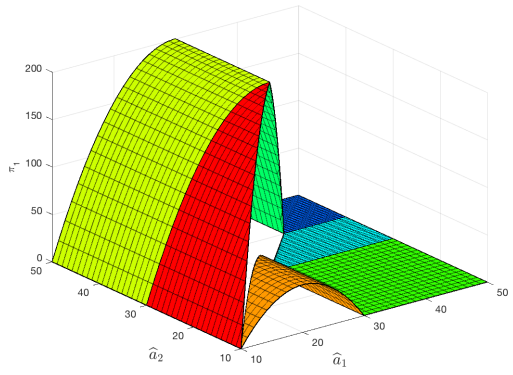
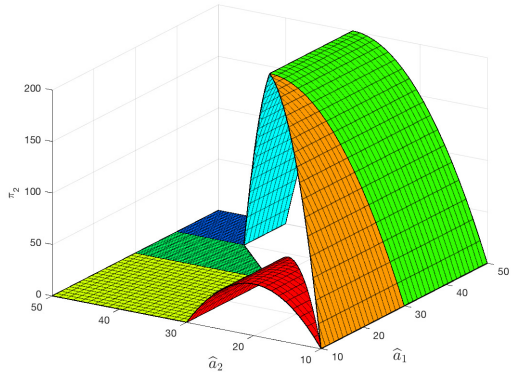


Figure 5.20: Value Function of ED for symmetric linear price competition



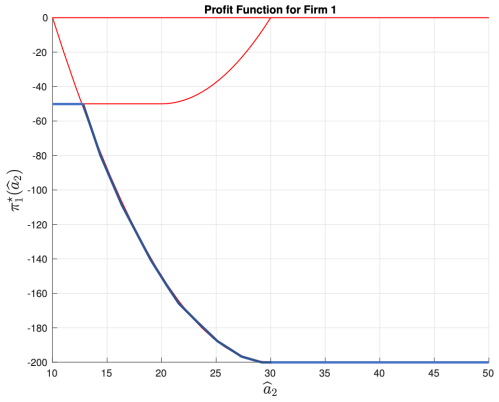
(a) π_1



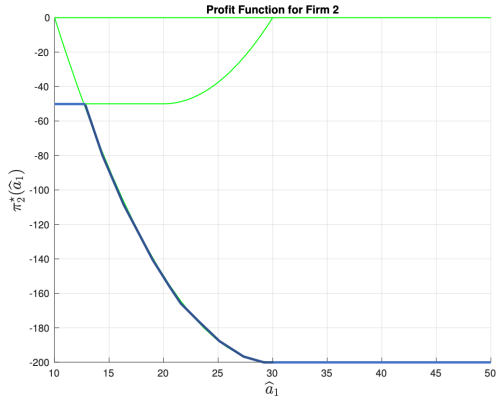
(b) π_2

Figure 5.21: Profit functions of individual producers for symmetric linear price competition

vide solution methodologies for the single-firm MPEC problem, and the exact computation of equilibria in the multi-firm problem. While the mathematical techniques presented here are sound, further computational studies on larger problem instances are required to validate the effectiveness and scalability of this approach. Of particular interest would be the computation of robust solutions to the single and multi-firm problems using this style of approach, as the multiparametric method is a natural fit for handling uncertainty or variation in some parameters of the problem. Further work on the stability and properties of the PWA system defined by the best response dynamics would be of interest, characterizing when we can expect diagonalization techniques to succeed. Additionally examining the dynamical system produced by smoothed best response dynamics as presented in Chapter 3 would be an interesting direction of future work.

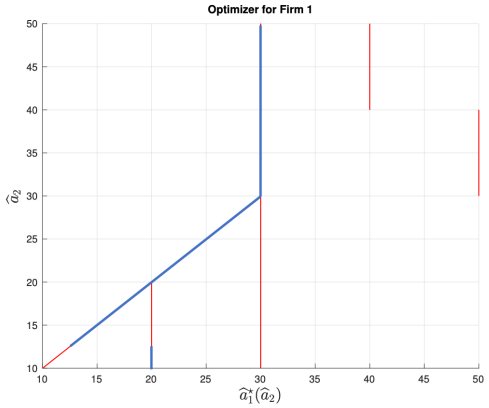


(a) $\pi_1^\star(\widehat{a}_2)$

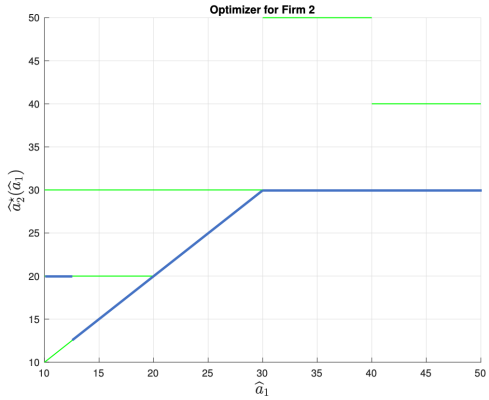


(b) $\pi_2^\star(\widehat{a}_1)$

Figure 5.22: Optimal (negative) payoff of each firm, in blue, parametric on the action of the other firm. Payoffs in other critical regions shown in red and green.



(a) $r_1(\widehat{a}_2)$



(b) $r_2(\widehat{a}_1)$

Figure 5.23: Best response function of each firm in blue. Sub-optimal actions from other critical regions shown in red and green.

Chapter 6

Open Access Storage: Should I Stay or Should I Go?

In the previous Chapter we addressed the question of strategic generation bidding in transmission-constrained electricity markets using techniques from multiparametric programming. In this Chapter we will extend these ideas to a new resource class, energy storage. In particular we are interested in the optimal operation of a given storage capacity, under a variety of regulatory frameworks and ownership models, and its consequences for private and social outcomes. This is a timely discussion in the context of the integration of flexible resources into existing electricity market structures. Throughout this chapter we will refer only to energy storage, but this should be taken to encompass both physical energy storage and virtual energy storage in the form of aggregations of flexible loads, in the sense of Section 2.3.

6.1 Introduction

Given its increasing penetration, the integration, operation and ownership of energy storage has been the subject of much debate. Two schools of thought predominate and can be broadly characterized as 1). *Market-Based Operation* (MBO), and 2). *Open Access Storage* (OAS). MBO would see energy storage independently owned and operated, and treated much like existing supply and demand side market participants are today, bidding into wholesale markets to maximize profits. OAS refers to the treatment of energy storage on a network as a *transmission asset*. That is, similar to transmission lines, energy storage is treated as part of the open access transport network, whose operation is optimized by the ISO to maximize social welfare. Transmission lines move energy through space, and energy storage moves energy through time, albeit in a single forward direction. In practice storage owners would submit their storage capability to the ISO, who would optimize the storage operation as part of MPED, and create a storage dispatch schedule, as is currently done for generators. The storage owners are then remunerated through some mechanism for adhering

to this dispatch schedule. Potential models for OAS have been proposed in [125–128]. These models all involve variations on *Financial Storage Rights* (FSRs), financially binding rights that entitle the owner to some portion of the revenue or surplus created by the storage. Storage owners are compensated in much the same way that transmission owners are though *Financial Transmission Rights* (FTRs) and *Flowgate Rights* (FGRs), through the proceeds of centralized auctions run by the ISO. It should be noted that these models are typically energy-only models, in that they do not consider ancillary services or alternative uses of storage other than energy arbitrage. Concerns also remain as to whether this mechanism provides suitable incentives for appropriate investment in energy storage, much as congestion revenue from FTRs is generally insufficient to cover the capital and investment costs of transmission.

The benefits of having storage operation optimized centrally is that storage can provide economic and system benefits to the whole market beyond the private owner, which might not materialize in the event of private strategic operation of storage. Centrally optimized storage also allows the ISO to take advantage of all the private information it has at its disposal, which in general is not available to market participants. In markets with ancillary services, the ISO could also potentially better co-optimize storage capacity across multiple applications.

In this chapter we shall study the incentives faced by storage owners under different regulatory frameworks for storage operation, the optimal operation of storage under each of these frameworks, and its impact on social welfare. We accomplish this through an analytical multiparametric solution of the MPED problem, and the solution of bilevel problems in a similar manner to the previous chapter. In particular we examine two variants of OAS, one in which storage owners must relinquish all storage capacity to the ISO, and one in which storage owners are allowed to strategically withold some capacity from the market. We also examine a case of MBO where the full storage capacity is operated and optimized privately by its owner, and appears in the market as a deviation in the net demand. This could alternatively be considered as storage discharge being bid into the market as zero marginal cost generation, and storage charging being bid as inelastic demand. We consider three ownership structures, merchant storage, generation-owned storage, and load-owned storage. We only consider each ownership structure in isolation, modelled as a single firm acting as a monopoly, and do not consider the multi-firm equilibrium problem.

Prior Literature

The concept of OAS was introduced in [125] where storage owners sell physically binding rights to their storage capacity through sequential auctions coordinated by the ISO. This notion was extended in [126] to financially binding rights to storage capacity, which represent entitlements to portions of the merchandising surplus collected by the ISO, analagous to existing FGRs for transmission lines. This work was in some sense completed by [127], broadening the applicability of financially binding rights to FSRs, which can take a form analagous to either FTRs or FGRs. These FSRs can be used in combination with FTRs

as a perfect hedge in a contract for difference when the buyer and seller exhibit differing intertemporal and spatial demand and supply characteristics. FSRs are defined formally in Section 6.2. In [128], the regulatory issue of competitively priced vs. unpriced services for the use of energy storage is discussed in the context of ratebased cost-recovery, and it is demonstrated how a storage capacity rights auction model with FSR style rights can overcome this issue.

Existing work on the integration of energy storage into transmission networks falls broadly into two categories. The first focuses on the optimal siting, sizing, and operation of energy storage. The second focuses on the welfare impacts of storage ownership. In this chapter we do not address the sizing or placement problem, assuming an existing installation of storage. We will however, consider questions of optimal operation under different regulatory frameworks and the subsequent welfare impacts. To the author's knowledge this is the first work to examine such concepts under different regulatory frameworks and indeed in networks with transmission constraints.

For the optimal siting and sizing problem, [129] introduced the concept of the locational marginal value of storage, using sensitivity analysis of the MPED problem, and showed how the optimal location and marginal value for the next increment of storage capacity could be determined from empirical price data for a given network. In [130], the infinite horizon investment and control problem for energy storage with a capacity budget was studied, and for convex and non-decreasing generation costs the optimal storage capacity allocation is found to assign zero capacity to generation-only buses that connect to the rest of the network via single links. In [34], the optimal storage placement problem is formulated as a discrete optimization problem, and a $(1 - \frac{1}{e})$-optimal placement algorithm is presented when the placement value function is submodular. The authors use a multiparametric programming approach to characterize sufficient conditions for submodularity and to develop the placement algorithm. We use many similar concepts in this chapter, in particular the parametric solution of MPED obtained by the authors, however we use these results to analytically characterize savings for participants under different regulatory frameworks. Additionally it should be noted that all of these papers consider the optimal storage placement problem in the context of *social* value, rather than from the perspective of a private participant. This is a perfectly reasonable formulation, however the private problem certainly merits further study. The results in this chapter give some insight as to this question and how the preferences of a private participant would vary under different regulatory frameworks. We also note the work of [131] which uses the MILP reformulation of the single-firm MPEC problem, detailed in the previous Chapter, in the context of price-maker storage bidding in electricity markets. This is an alternative methodology that is valid for computing solutions to the private problems described in this Chapter.

In [132] the authors estimate the value of energy storage in PJM over a six year period, and identify different drivers of value as well as the welfare impacts of storage. They also briefly discuss the implications of ownership and suggest that a regulated storage owner may have greater incentive than a private owner to invest and own storage due to its valuing the external social benefits of energy storage. The question of ownership is addressed more fully

in [133], where a single bus network with a linear price model is used to compute typical behavior for isolated ownership classes, and to compute equilibria among heterogeneous storage owners, for private storage operation. The author finds that if storage is owned by a finite number of symmetric generation or symmetric merchant storage firms, then storage will be underused relative to the social optimum, and if storage is owned by a finite number of symmetric consumers then storage will be overused relative to the social optimum. In the heterogenous ownership case, it is found that a combination of merchant and consumer ownership of storage maximizes potential welfare gains from storage. We will see that these results are echoed in the work in this chapter. In contrast, in [134] using a dynamic market model, the authors demonstrate that load-owned storage and merchant storage have an incentive to under-dimension their capacity relative to the social optimum, but suppliers do not. In [135], using a similar model to [133], the author examines when energy storage reduces social welfare, demonstrating that generator-owned storage or standalone storage in a market with strategic generating firms can reduce welfare compared to the no-storage case. We again see that these results are echoed using the more detailed model of this chapter. In all the papers here, network constraints are not considered allowing a great simplification of the problem and the easy computation of private and equilibrium solutions. Here we attempt to advance this analysis to the case of real networks, uniting the insights from these simpler models with the analytical framework of the placement problem in [34].

Other relevant work includes [136] who examine the effect of strategic bidding by load aggregators on the market equilibrium using a bilevel model, and a diagonalization approach to calculate the equilibrium. The authors demonstrate that this strategic bidding may mitigate the market power of generation firms, but also lead to reductions in social welfare. In [137] the authors use an MPEC formulation to numerically simulate the impact of different storage ownership structures on system welfare and the price of anarchy. They conclude that even when operated strategically, storage generally leads to welfare gains relative to the no-storage case, and that network congestion increases the lost welfare from strategic storage operation. In this chapter we provide an analytical method of treating this problem and characterize results exactly. Our conclusions echo the numerical simulations of [137].

Novel Contributions

The contributions of this chapter are as follows:

- We derive analytic expressions for the parametric solution of MPED.
- We present a methodology for solving the non-convex bilevel problem of strategic storage using parametric programming.
- We characterize participant savings both within and outside the small storage set.
- We demonstrate that loads can lose money with socially optimal storage, even within the small storage set.

- We demonstrate that the socially optimal storage control is not necessarily optimal for merchant storage, even within the small storage set.
- We provide conditions for the nonnegativity of load saving, and nonpositivity of generator savings.
- We provide conditions under which social and private optima are equal for loads.

We note that we address the strategic operation of storage, but not the strategic bidding of storage, *i.e.* if storage is forced to act alternatively as a generator or a load in the market, what price would they bid, in a problem more akin to the previous Chapter. This question of strategic price bidding, while important, is only relevant when storage is a price-maker in the market. This represents an interesting avenue of future work but we do not address it here.

Organization

The remainder of the chapter is organized as follows. In Section 6.2 we define ownership and regulatory models for storage operation, and mathematically formulate the storage model and MPED problem. In Section 6.3 we characterize the solution of MPED parametrically and provide explicit formulae for LMPs, generation optimizers, storage controls, and optimal cost. In Section 6.4 we demonstrate how the private storage optimization problem can be formulated as an MPEC and solved using multiparametric techniques in a similar fashion to the previous chapter. In Section 6.5 we characterize system and participant savings under different regulatory models both within and outside the small storage set. We additionally provide a counterexample to prove that load-owning storage can lose money, even within the small storage set. Section 6.6 concludes.

6.2 Problem Formulation

Network Modelling

We adopt the DC power flow model described in Section 2.2, and consider the operation of the system over a finite horizon of time periods, $\mathcal{T} := \{1, \dots, T\}$, indexed by t. This gives the feasible injection region of the power network as

$$\mathcal{P} = \{p : \mathbf{1}^\top p_t = 0, Hp_t \leq c, t \in \mathcal{T}\} \tag{6.1}$$

Storage Modelling

We adopt the storage model presented in Section 2.3, similar to that of [33] and [34]. For each bus i, the storage's state of charge (SOC) $x_i(t)$ evolves as

$$x_{i,t+1} = x_{i,t} - u_{i,t}, \quad t = 1, \dots, T-1 \tag{6.2}$$

where $u_{i,t}$ is the amount of energy discharged (if $u_{i,t} > 0$) or charged (if $u_{i,t} < 0$) in time period t, and the initial SOC is assumed to be $x_{i,t} = 0$. Each storage device has an energy capacity $s_i \geq 0$, such that

$$0 \leq x_{i,t} \leq s_i, \quad t \in \mathcal{T} \tag{6.3}$$

where $s_i = 0$ if there is no storage connected to bus i. Equations (6.2) and (6.3) can be compactly expressed in the following vector form

$$0 \leq Lu_i \leq s_i \mathbf{1}, \quad i = 1, \ldots, n \tag{6.4}$$

where $u_i \in \mathbb{R}^T$ is the vector of storage controls over T periods at bus i, and $L \in \mathbb{R}^{T \times T}$ is a lower triangular matrix with entries $L_{ij} = -1$ for $i \geq j$.[1] We denote the general set of feasible storage controls given a storage capacity s as

$$\mathcal{U}(s) := \{u \in \mathbb{R}^{n \times T} : 0 \leq Lu_i \leq s_i \mathbf{1}, \ \forall i\} \tag{6.5}$$

Multiperiod Economic Dispatch

We consider the MPED problem presented in Section 2.3, with inelastic demand and no generation constraints.[2]

$$J^\star(s, d) = \min_{g,p,u} \ \sum_{t=1}^{T} C_t(g_t) \tag{6.6a}$$

$$\begin{aligned}
\text{s.t. } \gamma_t &: \mathbf{1}^\top p = 0, && t \in \mathcal{T} && \text{(6.6b)}\\
\beta_t &: Hp \leq c, && t \in \mathcal{T} && \text{(6.6c)}\\
\lambda_t &: p_t = g_t + u_t - d_t, && t \in \mathcal{T} && \text{(6.6d)}\\
\mu_i &: Lu_i \leq s_i \mathbf{1}, && i \in \mathcal{N} && \text{(6.6e)}\\
\nu_i &: Lu_i \geq 0, && i \in \mathcal{N} && \text{(6.6f)}
\end{aligned}$$

where all notation is identical to that of Section 2.3.

In this chapter we will consider *MPED* as being parametrized by the storage vector, s, and the apparent demand, $d = \{d_t\}_{t \in \mathcal{T}}$, hence the notation for the optimal cost, $J^\star(s, d)$. We also distinguish between notation for the storage optimized by the ISO, s, and the apparent demand, d, and the storage capacity owned by market participants, and the nominal demand, denoted $\overline{s}$, $\overline{d}$, respectively.

We recall that the *Locational Marginal Prices* (LMPs) are the dual variables, $\lambda_t \in \mathbb{R}^n, t \in \mathcal{T}$ associated with the net injection constraint in (6.6d), and are equal to

$$\lambda_t^\star(s, d) = \gamma_t^\star(s, d)\mathbf{1} - H^\top \beta_t^\star(s, d), \quad t \in \mathcal{T} \tag{6.7}$$

[1] This model can easily be extended using a more detailed storage model with ramping constraints, charging efficiency, and SOC decay. For the sake of simplicity, we use the idealized model.

[2] Generation constraints can be included without changing any of the results of this Chapter, however they complicate the presentation of an already difficult formulation.

Regulatory Frameworks for Storage Integration

This formulation can capture a number of operational and regulatory models for storage:

1. *Open Access Storage (OAS)*: Under OAS, storage is treated as a transmission asset. Storage owners submit relevant operational parameters such as SOC and storage capacity s to the ISO, and then relinquish control of their devices, allowing their operation to be optimized by the ISO for maximal social benefit. Storage owners can be remunerated through the auction of financial storage rights, described in the following subsection. Within OAS there are two regulatory frameworks possible:

   a) *Capacity Withholding is Illegal*: In today's markets with open access transmission, transmission line owners are generally not permitted to withhold line capacity as this would give them a degree of market power and allow them to extract greater congestion rents. Similarly, treating storage as a transmission asset, one could envision a regulatory framework where storage owners are not allowed to withhold available capacity from the market, barring maintenance or fault. This would result in storage owners submitting their full capacity, $\overline{s}$, to the ISO under OAS, such that $s = \overline{s}$. We will denote this framework as *competitive open access storage (COAS)*.

   b) *Capacity Withholding is Legal*: An alternative point of view holds that storage owners are welcome to decide the level of capacity they wish to release to the ISO under OAS, with any additional capacity remaining unusable. This would result in storage owners submitting a capacity $s^p \in \mathbb{R}^n$, $0 \leq s^p \leq \overline{s}$, to the ISO under OAS, such that $s = s^p$. We will denote this framework as *strategic open access storage (SOAS)*.

2. *Private Storage Operation (PSO)*: Under private operation, storage owners retain control of their devices, operating them as they see fit. In this context, the ISO has no storage to optimize, $s = 0$, $u = 0$, and the effect of storage enters through a modifed apparent demand vector $d_t = \overline{d}_t - u_t^p$, where $\overline{d}_t$ represents the nominal inelastic demand at time t, and u_t^p represents the storage injection by a private participant, and is defined identically to u_t. The constraints (6.6e), (6.6f), are incorporated into the private optimization of the storage owner. Depending on the regulatory regime, u^p can be treated as must-take generation or load, or treated as its own specific storage asset class. In the case of load flexibility, u^p can show up behind the meter in the load vector submitted by the LSE to the ISO.

For some given storage capacity s, each framework will result in some feasible set of storage controls, defined as follows

$$\mathcal{U}_{COAS}(s) := \{u \in \mathbb{R}^{n \times T} : u = u^\star(s, d)\} \tag{6.8a}$$

$$\mathcal{U}_{SOAS}(s) := \{u \in \mathbb{R}^{n \times T} : u = u^\star(s^p, d), 0 \leq s^p \leq s\} \tag{6.8b}$$

$$\mathcal{U}_{PSO}(s) := \mathcal{U}(s) \tag{6.8c}$$

Assuming some non-zero storage capacity, the feasible sets have the following relation:

$$\mathcal{U}_{COAS}(s) \subseteq \mathcal{U}_{SOAS}(s) \subseteq \mathcal{U}_{PSO}(s) \tag{6.9}$$

where equality holds only when $s = 0$.

A third potential framework would include features of all of the above paradigms, where OAS and PSO are allowed to exist in parallel. A storage owner could submit a private capacity $0 \leq s^p \leq \overline{s}$ of their choice to the ISO, and additionally operate any remaining capacity with a private storage control $u^p \in \mathcal{U}(\overline{s} - s^p)$. While it is interesting to consider how a storage owner might allocate their capacity between OAS and PSO, this framework has the same feasible set of storage controls as PSO, $\mathcal{U}_{PSO}$. This is because any superposition of storage controls resulting from the capacity released to the ISO and the storage owner's private operation, could also be chosen as a purely private storage control. Given this fact, we focus on the pure PSO case and assume that OAS and PSO are mutually exclusive, such that either a participant submits a storage capacity to the ISO for OAS, or they operate the storage privately.

Merchandising Surplus and Financial Storage Rights

We recall that the *Merchandising Surplus* (*MS*) is the surplus collected by the ISO after receiving payments from all loads and disbursing payments to all generators, and is equal to

$$MS(s,d) = \sum_{t=1}^{T} -\lambda_t^\star(s,d)^\top (g_t^\star(s,d) - d_t) \tag{6.10}$$

Following the derivation in [33], the merchandising surplus can be decomposed into the sum of a *transmission congestion surplus* (*TCS*) and *storage congestion surplus* (*SCS*).

$$MS = TCS + SCS \tag{6.11}$$

where

$$TCS = \sum_{t=1}^{T} -\lambda_t^\star(s,d)^\top p_t^\star(s,d) \tag{6.12a}$$

$$= \sum_{t=1}^{T} \beta_t^\star(s,d)^\top c \tag{6.12b}$$

$$SCS = \sum_{t=1}^{T} \lambda_t^\star(s,d)^\top u_t^\star(s,d) \tag{6.12c}$$

$$= \sum_{t=1}^{T} \mu_t^\star(s,d)^\top s \tag{6.12d}$$

It is clear to see that the *MS* is nonnegative, and positive when any congestion occurs in transmission lines or energy storage.

In many current power markets, shares of the merchandising surplus associated with transmission congestion are auctioned off through *financial transmission rights*. These can take either the form of *point-to-point financial transmission rights* (*FTRs*), entitling the holder to the price difference between two nodes in a given time interval, or of *flowgate rights* (*FGRs*), entitling the holder to the congestion rent associated with a given line in a given time interval. *FTRs* represent a perfect hedge for participants engaged in bilateral contracts that are left out of the money due to transmission congestion. The revenue paid out to a set of simultaneously feasible *FTRs* is always less than or equal to the *TCS*, and it can be shown that there is always a set of simultaneously feasible *FTRs* that exactly recovers the *TCS*.

For the case of *OAS*, [33, 126] define *financial storage rights*, entitling the holders to a share of the *MS* associated with storage congestion. These can take either the form of a *Financial Storage Right* (*FSR*), composed of an hourly power profile at a node entitling the holder to the hourly price profile at that node, or the form of an *Energy Capacity Right* (*ECR*), composed of a storage capacity at a given node entitling the holder to the storage congestion rent associated with the storage unit at that node. *FSRs* and *ECRs* for storage are analogous to *FTRs* and *FGRs* for transmission lines, respectively. Similarly the revenue paid out to a set of simultaneously feasible *FSRs* is always less than or equal to the *SCS*, and it can be shown that there is always a set of simultaneously feasible *FSRs* that exactly recovers the *SCS*.[3]

In our case we will assume that under *OAS*, a simultaneously feasible set of *FTRs* and *FSRs* that fully recover the *MS* has been auctioned off at their true value, and the resulting share of auction revenue equal to the *SCS* is returned to storage owners.[4]

Storage Ownership Structures and Payoffs

We consider three classes of market participants: *merchant storage*, *load serving entities* (*LSEs*), and *generation*. For reasons that will subsequently become clear we consider each class of participant as a collective, modelling the aggregate payoff of each class. An alternative interpretation is that there are only three firms in the market: one firm that owns all merchant storage, one firm that owns all *LSEs*, and one firm that owns all generation.

We then consider three scenarios for storage ownership, with each participant collective owning storage in isolation *e.g.* one firm owns all load and storage on the network, one firm owns all generation and no storage. Merchant storage only exists when storage is not

[3]For a more rigorous treatment of *FSRs*, the reader is referred to [33], where simultaneous feasibility of financial rights is defined, and an example is presented showing the perfect hedge provided to a bilateral contract between a generator and a load using a collection of *FTRs* and *FSRs*.

[4]In current markets *FTRs* are auctioned off months in advance, so generally recoup their expected value at auction, which may be higher or lower than the true merchandising surplus which results at delivery. Additionally the auction revenue from *FTRs* is not necessarily returned to transmission owners, in some cases being distributed among market participants.

owned by either *LSEs* or generation. We will then consider the payoff to each storage-owning collective under each of the regulatory frameworks described previously, COAS, SOAS, PSO.

We also assume that only storage is ever operated strategically, and all other participants behave competitively. In practice this means generators bid their true cost functions to the ISO, and do not mark up their bids in any way.

The specific payoffs to participant collectives, when owning storage in isolation, under any of the regulatory frameworks are defined as follows:

$$MP(s,d,u) = \sum_{t=1}^{T} \lambda_t^\star(s,d)^\top u_t \tag{6.13a}$$

$$LP(s,d,u) = \sum_{t=1}^{T} -\lambda_t^\star(s,d)^\top (\overline{d}_t - u_t) \tag{6.13b}$$

$$GP(s,d,u) = \sum_{t=1}^{T} \lambda_t^\star(s,d)^\top (g_t^\star(s,d) + u_t) - C_t(g_t^\star(s,d)) \tag{6.13c}$$

where $MP(s,d,u)$ is the profit made by merchant storage, $LP(s,d,u)$ is the negative payment made by storage-owning load[5], $GP(s,d,u)$ is the profit made by storage-owning generation, $s \in \mathbb{R}^n$ is the storage capacity released to the ISO under *OAS*, $d \in \mathbb{R}^{n\times T}$ is the apparent demand seen by the ISO, and $u \in \mathbb{R}^{n\times T}$ is the storage control under any of the regulatory frameworks. We can write the payoffs in this way, since we assume that the storage congestion surplus is completely returned to storage owners and recalling that

$$SCS = \sum_{t=1}^{T} \mu_t^\star(s,d)^\top s = \sum_{t=1}^{T} \lambda_t^\star(s,d)^\top u_t \tag{6.14}$$

We assume that each storage-owning participant collective wishes to maximize this payoff. We denote a general participant collective payoff function $\pi(s,d,u)$, where π can refer to any of MP, LP, or GP. Thus the payoffs will take the following forms under each regulatory framework

$$\pi_{COAS}(\overline{s},\overline{d},u^\star(\overline{s},\overline{d})), \tag{6.15a}$$

$$\pi_{SOAS}(s^p,\overline{d},u^\star(s^p,\overline{d})), \tag{6.15b}$$

$$\pi_{PSO}(0,\overline{d}-u^p,u^p) \tag{6.15c}$$

Denoting the set of primal and dual variables of *MPED* as $z := (g,p,u,\lambda,\gamma,\beta,\mu,\nu)$, and given the feasible sets of storage controls described in (6.9), the optimization to calculate the optimal payoff under each regulatory framework takes the following general form:

$$\pi_x^\star = \max_{s,u} \ \pi_x(s,d,u) \tag{6.16a}$$

$$\text{s.t. } u \in \mathcal{U}_x(\overline{s}) \tag{6.16b}$$

$$z \in \text{EQBM}\,(MPED(s,d)) \tag{6.16c}$$

[5]Since loads are inelastic, their 'payoff' here is the just the negative payment they make to consume energy.

where EQBM $(MPED(s,d))$ is the equilibrium solution of MPED, and x is a dummy subscript representing one of COAS, SOAS, or PSO. Given the relation described in (6.9), it is trivial to see that for all participant collectives

$$\pi^\star_{COAS} \leq \pi^\star_{SOAS} \leq \pi^\star_{PSO} \tag{6.17}$$

Following from participant payoffs, the main quantity of interest in this chapter will be the *savings* to participants from storage, alternatively the improvement in their payoff. These savings are the increase in payoff a participant will achieve using storage relative to a base case with no storage, and take the general form

$$\widehat{\pi}_x(s,d,u) = \pi_x(s,d,u) - \pi(0,\overline{d},0) \tag{6.18}$$

where $\widehat{\pi}_x(s,d,u)$ is the saving under a regulatory regime, x, calculated relative to a base case of MPED with no storage. For a participant, maximizing payoff is the same as maximizing savings, since $\pi(0,\overline{d},0)$ is a constant, and we will generally refer to participants maximizing their payoff throughout this chapter. We note that

$$0 \leq \widehat{\pi}^\star_{SOAS} \tag{6.19a}$$
$$0 \leq \widehat{\pi}^\star_{PSO} \tag{6.19b}$$

since if there is no utilization of storage that yields a positive saving, participants can choose to simply leave their storage unused, *i.e.* completely withold all their capacity. Similarly, we denote the system savings from storage, alternatively the welfare improvement, as

$$\widehat{J}(s,d,u) = J^\star(0,\overline{d},0) - J^\star(s,d,u) \tag{6.20}$$

Note this is defined inversely to participant savings, as a lower social cost represents a positive saving to the system.

Aside from the COAS case which is not optimized by the participant, the other regulatory cases present a non-convex bilevel optimization program for the participant in (6.16), where the second level is the equilibrium conditions of MPED. As discussed previously, these problems represent MPECs and can be solved by existing solution methods discussed in this Chapter and the previous Chapter. However these solution methods provide little analytical or structural insight. We seek to understand the conditions under which participants will increase their payoff (make savings) with storage, and under what conditions they might prefer one regulatory regime over another.

To further understand the structure of these problems, we will use *multiparametric programming* to analytically characterize the solution of MPED, parametric on the *OAS* storage capacity s, and the apparent demand d.

6.3 Multiparametric Approach to MPED

Recalling the theory of multiparametric programming, for a multiparametric quadratic program with linear constraints (mpQP), as we have in (6.6), the value function, optimizers,

and dual variables can be characterized explicitly as functions of the parameters over a finite set of polyhedral *critical regions* in the parameter space. For every value of the parameters within each critical region the *binding constraints* at the optimum of the optimization problem *do not change*.

Solving MPED parametrically is useful as it allows structural insight into the savings that participants make with the addition of storage under each regulatory regime. It also offers an alternative method of solving bilevel problems of the kind described in (6.16), using similar methods to those discussed in the previous Chapter.

Multiparametric Solution of MPED

MPED was previously analyzed using multiparametric programming in [34], however the solution is incomplete in that it implicitly assumes that no two consecutive LMPs are equal. This condition corresponds to the case where the storage control equalizes the load and consequently the price between a number of consecutive periods. This condition is difficult to resolve as it means the reduced dual function in (6.24) cannot be decoupled across time as claimed in [34]. This was amended in the Journal version of this paper, following discussions with the author, [138], where a full solution of MPED was derived. However this solution does not lend itself to compact analytic formulae, as the time-decoupling of the solution is lost. Here we extend the original formulation of [34], to the case where any number of consecutive periods can have identical LMPs at each node across time. This limits the applicability of the following analytic work to cases satisfying this condition, however from extensive simulation and the author's experience, this will cover the vast majority of real world cases. Additionally, it should be noted that the full parametric solution, applying in all cases, is only analytically problematic, and not numerically intractable, and can be found recursively for any instance of MPED using techniques described in [38].

We aim to find the parametric solution of the mpQP MPED(s, d) using multiparametric programming. We begin by considering the dual of (2.35) as in [34].

$$\max_{\lambda,\gamma,\beta,\mu,\nu} \ \phi(\lambda,\gamma,\beta,\mu,\nu) \tag{6.21a}$$

$$\text{s.t. } \lambda_t = \gamma_t \mathbf{1} - H^\top \beta_t, \quad t \in \mathcal{T} \tag{6.21b}$$

$$\lambda_i = L^\top(\mu_i - \nu_i), \quad i \in \mathcal{N} \tag{6.21c}$$

$$\beta, \mu, \nu \geq 0 \tag{6.21d}$$

where the Lagrange dual function is given by

$$\begin{aligned}\phi(\lambda,\gamma,\beta,\mu,\nu) = \sum_{t=1}^{T} -\frac{1}{2}(\lambda_t - a_t)^\top Q_t^{-1}(\lambda_t - a_t) \\ + d_t^\top \lambda_t - c_t^\top \beta_t - s^\top \mu_t\end{aligned} \tag{6.22}$$

As observed in [34], due to complementary slackness and the structure of the matrix L,

$$\mu_t = (\lambda_{t+1} - \lambda_t)^+, \quad \text{and} \quad \nu_t = (\lambda_{t+1} - \lambda_t)^-, \ t \in \mathcal{T} \tag{6.23}$$

where for convenience $\lambda_{T+1} := 0 \in \mathbb{R}^n$. This allows the dual to be reduced to the following form

$$\max_{\lambda,\gamma,\beta} \tilde{\phi}(\lambda, \beta) \tag{6.24a}$$

$$\text{s.t. } \lambda_t = \gamma_t \mathbf{1} - H^\top \beta_t, \quad t \in \mathcal{T} \tag{6.24b}$$

$$\beta \geq 0 \tag{6.24c}$$

where

$$\begin{aligned} \tilde{\phi}(\lambda, \beta) = \sum_{t=1}^{T} & -\frac{1}{2}(\lambda_t - a_t)^\top Q_t^{-1}(\lambda_t - a_t) \\ & + d_t^\top \lambda_t - c^\top \beta_t - s^\top (\lambda_{t+1} - \lambda_t)^+ \end{aligned} \tag{6.25}$$

Within a critical region of the problem, the dual becomes an equality constrained QP that under certain conditions can be solved analytically. To define the critical regions, the power balance equalities in (6.6b), (6.6d) always hold, and for different values of s, d, different transmission limits (6.6c), and storage constraints (6.6e), (6.6f), will bind at the optimum. Similarly to [34] we characterize the network and storage congestion patterns as follows.

For each $(i, t) \in \mathcal{N} \times \mathcal{T}$, let $\chi_{i,t}(s, d) = 1$ if the constraint $(Lu_i)_t \leq s_i$ is binding at the optimum and $\chi_{i,t}(s, d) = 0$ otherwise. In other words, χ represents the storage congestion pattern. From (6.23) we can write

$$\chi_{i,t} := \chi_{i,t}(s, d) \begin{cases} 1 & \text{if } \lambda^\star_{i,t+1}(s, d) - \lambda^\star_{i,t}(s, d) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{6.26}$$

We also define Δ_{χ_t} as the diagonal matrix with vector χ_t on the diagonal, with $\chi_0 := 0$.

We let $\mathcal{E}_{Ct} \subseteq \{1, \ldots, 2m\}$ denote the set of oriented lines that are congested at the solution in period t, and $m_t := |\mathcal{E}_{Ct}|$ denote the number of congested lines. We define a selection matrix $W_t \in \mathbb{R}^{m_t \times 2m}$ such that for $l = 1, \ldots, m_t$ and $l' = 1, \ldots, 2m$,

$$(W_t)_{l,l'} := (W_t(s, d))_{l,l'} \begin{cases} 1 & \text{if the } l\text{th element in } \mathcal{E}_{Ct} \text{ is } l' \\ 0 & \text{otherwise} \end{cases} \tag{6.27}$$

and the shift factor matrix for congested lines as

$$H_t := W_t H \tag{6.28}$$

We assume the following constraint qualification

Assumption 1. *(Flow LICQ): For each $t \in \mathcal{T}$, H_t, as defined in* (6.28)*, is of full row rank for all $t \in \mathcal{T}$.*

We now diverge from the proof in [34]. Analagous to χ_t, we define σ_t as follows: For each $(i,t) \in \mathcal{N} \times \mathcal{T}$, let $\sigma_{i,t}(s,d) = 1$ if both $(Lu_i)_t < s_i$, and $(Lu_i)_t > 0$, (*i.e.* $\mu_{i,t} = 0$, and $\nu_{i,t} = 0$), and $\sigma_{i,t}(s,d) = 0$ otherwise. In other words, σ represents the times when LMPs are equal across time periods. In summary:

$$\sigma_{i,t} := \sigma_{i,t}(s,d) \begin{cases} 1 & \text{if } \lambda^\star_{i,t+1}(s,d) = \lambda^\star_{i,t}(s,d) \\ 0 & \text{otherwise} \end{cases} \tag{6.29}$$

We also define $\Delta_{\sigma_t} \in \mathbb{R}^{n \times n}$ as the diagonal matrix with σ_t on the diagonal. The inclusion of this congestion pattern for the cases when LMPs were equal across time periods was omitted from [34], and is the substantive difference between the two derivations.

We can now confine the reduced dual to a critical region, where the congestion patterns are given by W_t, χ_t and σ_t giving an equality constrained QP.

$$\max_{\lambda,\gamma,\tilde{\beta}} \ \psi(\lambda,\gamma,\tilde{\beta}) \tag{6.30a}$$

$$\text{s.t. } \lambda_t = \gamma_t \mathbf{1} - H_t^\top \tilde{\beta}_t, \quad t \in \mathcal{T} \tag{6.30b}$$

$$\Delta_{\sigma_t}(\lambda_{t+1} - \lambda_t) = 0, \quad t \in \mathcal{T} \tag{6.30c}$$

where

$$\begin{aligned} \psi(\lambda,\gamma,\tilde{\beta}) = \sum_{t=1}^{T} & -\frac{1}{2}(\lambda_t - a_t)^\top Q_t^{-1}(\lambda_t - a_t) + d_t^\top \lambda_t \\ & - c_t^\top \tilde{\beta}_t - s^\top(\Delta_{\chi_t} - \Delta_{\chi_{t-1}})\lambda_t \end{aligned} \tag{6.31}$$

Unfortunately, we see that this problem is not separable across time in general, which makes the analytic solution more difficult. The key idea is that λ_t which are coupled by (6.30c) must be optimized simultaneously. We examine a number of special cases below.

For situtations where $\Delta_{\sigma_t} = 0$ and $\Delta_{\sigma_{t-1}} = 0$, the analytic solution for $\lambda^\star_t(s,d)$ is the same as that described in [34]. In this case $\lambda^\star_t(s,d)$ is not coupled across time to any other LMPs.

For situations where $\Delta_{\sigma_t} = \mathbf{I}$, for some t, the solution can be derived analytically. This corresponds to the case where $\lambda_{t+1} = \lambda_t$, arising when the LMPs are identical across periods when there is no line congestion, or very special cases when there is line congestion. This can in general extend over a number of consecutive time periods, which we will denote $\tilde{T}$, such that $\lambda_t = \lambda_{t+1} = \cdots = \lambda_{t+\tilde{T}-1}$. We also define $\sigma(t) := \{t_1, \ldots, t_{\tilde{T}}\}$ as the set of consecutive time periods for which $\Delta_{\sigma_j} = \mathbf{I}$, $j \in \{t_1, \ldots, t_{\tilde{T}-1}\}$, and $\Delta_{\sigma_{t_{\tilde{T}}}} = 0$. It is always true that $t \in \sigma(t)$. As an example, say there are three periods, and that $\Delta_{\sigma_1} = \mathbf{I}$, $\Delta_{\sigma_2} = \Delta_{\sigma_3} = 0$. implying that $\lambda_2 = \lambda_1$, and λ_3 is distinct. Then $\sigma(1) = \sigma(2) = \{1,2\}$, and $\sigma(3) = \{3\}$.

This condition allows the problem to decouple in time across segments where $\lambda^\star_t$ is distinct, amd can be solved analytically, the derivation of which is given in Appendix A.1.

For convenience we also define a reduced time index set $\mathcal{K}$, which is assumed to have distinct elements, and $\mathcal{K} \subseteq \mathcal{T}$.

$$\mathcal{K} := \{k : k = \max(\sigma(t)), \ \forall t \in \mathcal{T}\} \tag{6.32}$$

$\mathcal{K}$ corresponds to a set of time indices that have non-identical consecutive LMPs, where for any set of time indices with consecutively identical LMPs, $\sigma(t)$, the final time index of that set is chosen to represent that sequence. In the brief example given above, $\mathcal{T} = \{1, 2, 3\}$, and $\mathcal{K} = \{2, 3\}$. $\mathcal{K}$ is a surjective map of $\mathcal{T}$, with each $t \in \mathcal{K}$ associated with a unique $k \in \mathcal{K}$, and each $k \in \mathcal{K}$ associated to the set $\sigma(k) \in \mathcal{T}$. In the event that there are no consecutive identical LMPs, then $\mathcal{K} = \mathcal{T}$.

We now have the following characterizations of the value function, optimizers, and dual variables of (6.6). If $\mathcal{K} = \mathcal{T}$, then these results reduce to those derived in [34].

Theorem 9. *In the critical region where the storage and network congestion patterns are represented by χ_t and W_t, $t \in \mathcal{T}$, and $\Delta_{\sigma_j} = \mathbf{I}$ or $\Delta_{\sigma_j} = 0$, $\forall j \in \sigma(t)$, $\forall t \in \mathcal{T}$, the generation dispatch $g^\star$, and locational marginal prices $\lambda^\star$ are unique and affine in s and d, and can be written as*

$$\lambda_t^\star(s, d) = A_t(W_t)\left(\Delta_t s + d_{\sigma(t)}\right) + \overline{\lambda}_t(W_t) \tag{6.33}$$

$$g_t^\star(s, d) = Q_t^{-1}(\lambda_t^\star(s, d) - a_t) \tag{6.34}$$

where

$$A_t(W_t) := Q_{\sigma(t)} M_t R_t M_t Q_{\sigma(t)} + \rho_t \mathbf{1}\mathbf{1}^\top, \tag{6.35}$$

$$\Delta_t := \sum_{j \in \sigma(t)} (\Delta_{\chi_j} - \Delta_{\chi_{j-1}}) \tag{6.36}$$

$$d_{\sigma(t)} := \sum_{j \in \sigma(t)} d_j \tag{6.37}$$

$$\overline{\lambda}_t(W_t) := A_t(W_t)\left(\sum_{j \in \sigma(t)} Q_j^{-1} a_j\right) + |\sigma(t)| B_t(W_t) W_t c, \tag{6.38}$$

$$B_t(W_t) := Q_{\sigma(t)} M_t H_t^\top K_t^{-1} \tag{6.39}$$

and

$$Q_{\sigma(t)} := \left(\sum_{j \in \sigma(t)} Q_j^{-1}\right)^{-1} \tag{6.40}$$

$$\rho_t := \frac{1}{\mathbf{1}^\top Q_{\sigma(t)}^{-1} \mathbf{1}}, \tag{6.41}$$

$$M_t := Q_{\sigma(t)}^{-1} - \frac{Q_{\sigma(t)}^{-1} \mathbf{1}\mathbf{1}^\top Q_{\sigma(t)}^{-1}}{\mathbf{1}^\top Q_{\sigma(t)}^{-1} \mathbf{1}}, \tag{6.42}$$

$$K_t := H_t M_t H_t^\top, \tag{6.43}$$

$$R_t := H_t^\top K_t^{-1} H_t \tag{6.44}$$

When there is no line congested in period t, all the expressions above hold with $R_t := 0$ and $B_t(W_t) := 0$.

Proof. The proof is included in Appendix A.1. The expressions for $\mu^\star(s,d)$ and $\nu^\star(s,d)$ are uniquely defined by the expression for $\lambda^\star(s,d)$. □

We observe the following useful lemmas:

Lemma 1. *The matrix $A_t(W_t)$ is positive semi-definite. That is $A_t(W_t) \succeq 0$. Additionally, $M_t \succeq 0$, $K_t \succ 0$.*

Proof. The proof is included in Appendix A.2. □

Lemma 2. *The matrix $M_t \succeq 0$ has exactly one eigenvalue at 0, associated with the eigenvector $\mathbf{1}$.*

Proof. The proof is included in Appendix A.3. □

We make additional remarks on properties of the analytical solution of MPED, and some useful identities involving the matrices of Theorem 9 in Appendix A.4.

We now turn to the storage control optimizers $u^\star(s,d)$, and observe that in general they are non-unique. This will occur when there are multiple storage units located at the same bus, or at different buses connected by an uncongested line. For the purpose of solving *MPED* any optimal set of storage controls is good enough, however in the event that multiple parties own storage on the network, the allocation of storage control will have an effect on the distribution of storage revenues among participants.

To characterize the optimal solution set for the storage controls we determine the binding constraints related to storage at the optimum. We introduce the matrix $\xi(s,d)$, defined similarly to $\chi(s,d)$, such that for each $(i,t) \in \mathcal{N} \times \mathcal{T}$, let $\xi_{i,t}(s,d) = 1$ if the constraint $(Lu_i)_t \geq 0$ is binding at the optimum and $\xi_{i,t}(s,d) = 0$ otherwise. In other words, ξ represents the congestion pattern of the lower limit of storage. From (6.23) we can write

$$\xi_{i,t} := \xi_{i,t}(s,d) \begin{cases} 1 & \text{if } \lambda^\star_{i,t+1}(s,d) - \lambda^\star_{i,t}(s,d) < 0 \\ 0 & \text{otherwise} \end{cases} \tag{6.45}$$

The binding storage constraints at the optimum can then be written as

$$\mathbf{1}^\top (g_t^\star + u_t^\star - d_t) = 0, \qquad t \in \mathcal{T} \tag{6.46}$$

$$H_t(g_t^\star + u_t^\star - d_t) = c_t, \qquad t \in \mathcal{T} \tag{6.47}$$

$$\Delta_{\chi_i} L u_i^\star = \Delta_{\chi_i} s_i, \qquad i \in \mathcal{N} \tag{6.48}$$

$$\Delta_{\xi_i} L u_i^\star = 0, \qquad i \in \mathcal{N} \tag{6.49}$$

where $\Delta_{\xi_i} \in \mathbb{R}^{T \times T}$ is the diagonal matrix with vector ξ_i on the diagonal. Any $u^\star$ that satisfies these equations is a valid storage control profile that satisfies the optimality conditions of MPED within a given critical region.

While the storage control optimizers might not be unique, the storage congestion surplus is in fact unique. As proved in [34] the LMPs, $\lambda^\star(s,d)$, are unique, and consequently the

upper storage congestion multiplier, $\mu^\star(s,d)$, is unique. Therefore the SCS defined in (6.12d) is unique. This allows us to demonstrate a candidate solution for the storage control which attains the unique SCS, and thus corresponds to a vaild optimal storage control.

Theorem 10. *$u_t^\star(s,d) = -\Delta_t s$, $t \in \mathcal{T}$, is a valid optimal storage control in a given critical region as characterized in Theorem 9.*

Proof. We show that the candidate storage control attains the unique SCS and is therefore optimal. Let $u_t^\star = -\Delta_t s$. From (6.12c) and (6.12d)

$$SCS = \sum_{t=1}^{T} \lambda_t^\star(s,d)^\top u_t^\star(s,d) \tag{6.50}$$

$$= \sum_{t=1}^{T} -\lambda_t^\star(s,d)^\top \Delta_t s \tag{6.51}$$

$$= \sum_{t=1}^{T} (\lambda_{t+1}^\star - \lambda_t^\star)^{+\top} s \tag{6.52}$$

$$= \sum_{t=1}^{T} \mu_t^\star(s,d)^\top s \tag{6.53}$$

□

Finally the optimal cost can be characterized analytically as

$$J^\star(s,d) = \sum_{k\in\mathcal{K}} J_k(s, d_{\sigma(k)}) \tag{6.54}$$

where

$$\begin{aligned} J_k(s, d_{\sigma(k)}) = \frac{1}{2} s^\top \Delta_k A_k(W_k) \Delta_k s + \frac{1}{2} d_{\sigma(k)}^\top A_k(W_k) d_{\sigma(k)} \\ + \overline{\lambda}_k^\top (\Delta_k s + d_{\sigma(k)}) + d_{\sigma(k)}^\top A_k(W_k) \Delta_k s + \overline{J}_k \end{aligned} \tag{6.55}$$

and

$$\overline{J}_k = \frac{1}{2} \overline{\lambda}_k^\top Q_{\sigma(k)}^{-1} \overline{\lambda}_k - \sum_{j\in\sigma(k)} \frac{1}{2} a_j^\top Q_j^{-1} a_j \tag{6.56}$$

We have now analytically characterized the value function, optimizers and dual variables of MPED within a critical region, under the condition where only entire consecutive LMP vectors are permitted to be identical.[6] The critical regions of the problem can be calculated using the methods described in Section 2.4. We denote a general feasible parameter set Θ,

[6] As noted previously, the numerical computation of the parametric solution of MPED is not limited by the conditions described here.

and the associated finite set of critical regions as $\mathcal{R}_j(W_t, \chi_t, \sigma_t, \xi_t, \forall t \in \mathcal{T})$, $j = 0, \ldots, N_{CR}-1$, where there are N_{CR} critical regions. We have that

$$\bigcup_j \mathcal{R}_j = \Theta \tag{6.57}$$

We also define the *small storage set*, $\mathcal{R}_0$, which is the critical region containing the parameter values $(s = 0, d = \overline{d})$. This corresponds to the base case of MPED with no storage, either OAS or private. The notion of small storage refers to the fact that for some $s = \epsilon$, the set of active constraints from the case with $s = 0$. Thus the critical region would not change and the same set of analytic functions would describe the solution in both cases. This idea will come in useful when trying to make statements about participant savings with storage in Section 6.5.

6.4 Solving Bilevel Problems

Having analytic formulae for the LMP, $\lambda^\star$, and the generation optimizer, $g^\star$, contingent on being within a given critical region, we can now return to solve the private optimization problems described in (6.16). Using the analytical formulae, we will show how these non-convex problems can be reduced to a finite set of QPs, or a single tractable MIQP, using similar concepts to those presented in the previous Chapter. We will take the LSE's negative load payment maximization problem as an example to illustrate these ideas.

LSE - SOAS

The original problem is written as

$$\max_{s^p} \sum_{t=1}^{T} -\lambda_t^\star(s^p, \overline{d})^\top \overline{d}_t + \mu_t^\star(s^p, \overline{d})^\top s^p \tag{6.58a}$$

$$\text{s.t. } 0 \leq s^p \leq \overline{s} \tag{6.58b}$$

$$\lambda^\star,\ \mu^\star \in \text{EQBM}\left(MPED(s^p, \overline{d})\right) \tag{6.58c}$$

We assume that we have found the multiparametric solution of MPED over the finite parameter set Θ, where

$$\Theta = \{s \in \mathbb{R}^n, d \in \mathbb{R}^{n\times T} : 0 \leq s \leq \overline{s},\ d = \overline{d}\} \tag{6.59}$$

resulting in a finite set of critical regions, denoted $\mathcal{R}_j(W_t, \chi_t, \sigma_t, \xi_t, \forall t \in \mathcal{T})$, $j = 0, \ldots, N_{CR}-1$. We can now recast the above problem as follows

$$\max_{s^p} \sum_{t=1}^{T} -\lambda_t^\star(s^p, \overline{d})^\top \overline{d}_t - \lambda_t^\star(s^p, \overline{d})^\top \Delta_t s^p \tag{6.60a}$$

$$\text{s.t. } 0 \leq s^p \leq \overline{s} \tag{6.60b}$$

$$\lambda_t^\star = A_t(W_t)(\Delta_t s^p + \overline{d}_t) + \overline{\lambda}_t(W_t), \quad \text{if } (s^p, \overline{d}) \in \mathcal{R}_j, \forall j, \forall t \tag{6.60c}$$

The solution to this problem can be found by converting the set membership *if* condition to a binary constraint using the Big-M method, and solving the full MILP problem. Alternatively the optimal solution corresponds to the largest value over a finite set of convex LPs, defined as follows

$$LP^{\star}_{COAS} = \max_{j} \left(\max_{(s^p,\overline{d})\in\mathcal{R}_j} \sum_{t=1}^{T} f_t(s^p, \mathcal{R}_j) \right) \tag{6.61}$$

where

$$f_t(s^p, \mathcal{R}_j) = -(\Delta_t s^p + \overline{d}_t)^\top A_t(W_t)(\Delta_t s^p + \overline{d}_t) - \overline{\lambda}_t(W_t)^\top(\Delta_t s^p + \overline{d}_t) \tag{6.62}$$

The function $f_t(\cdot)$ is obtained by substituting the expression for $\lambda^\star_t$ in (6.60c) into the objective in (6.60a).

LSE - PSO

The original problem is written as

$$\max_{u^p} \sum_{t=1}^{T} -\lambda^\star_t(0, \overline{d} - u^p)^\top(\overline{d}_t - u^p_t) \tag{6.63a}$$

$$\text{s.t. } u^p \in \mathcal{U}(\overline{s}) \tag{6.63b}$$

$$\lambda^\star \in \text{EQBM}\left(MPED(0, \overline{d} - u^p)\right) \tag{6.63c}$$

Abusing notation, we assume that we have found the multiparametric solution of MPED over the finite parameter set Θ, where

$$\Theta = \{s \in \mathbb{R}^n, d \in \mathbb{R}^{n\times T} : s = 0,\ \overline{d} - \overline{s} \leq d \leq \overline{d} + \overline{s}\} \tag{6.64}$$

resulting in a finite set of critical regions, denoted $\mathcal{R}_j(W_t, \chi_t, \sigma_t, \xi_t, \forall t \in \mathcal{T})$, $j = 0, \ldots, N_{CR}-1$. We can recast it using the analytic solution for the LMPs as

$$\max_{u^p} \sum_{t=1}^{T} -\lambda^\star_t(0, \overline{d} - u^p)^\top(\overline{d}_t - u^p_t) \tag{6.65a}$$

$$\text{s.t. } u^p \in \mathcal{U}(\overline{s}) \tag{6.65b}$$

$$\lambda^\star_t = A_t(W_t)(\overline{d}_t - u^p_t) + \overline{\lambda}_t(W_t), \quad \text{if } (0, \overline{d} - u^p) \in \mathcal{R}_j, \forall j, \forall t \tag{6.65c}$$

Similarly to the COAS problem described above, the problem can be solved as an MIQP using the Big-M method to convert the set membership constraint using a binary variable. Alternatively the optimal solution corresponds to the largest value over a finite set of convex QPs, defined as follows

$$LP^{\star}_{PSO} = \max_{j} \left(\max_{(0,\overline{d}-u^p)\in\mathcal{R}_j} \sum_{t=1}^{T} f_t(u^p, \mathcal{R}_j) \right) \quad \text{s.t. } u^p \in \mathcal{U}(\overline{s}) \tag{6.66}$$

where

$$f_t(u^p, \mathcal{R}_j) = -(\overline{d}_t - u_t^p)^\top A_t(W_t)(\overline{d}_t - u_t^p) - \overline{\lambda}_t(W_t)^\top(\overline{d}_t - u_t^p) \tag{6.67}$$

We note the similarity between equations (6.62) and (6.67), where if one substitutes $u_t^p = -\Delta_t s$, we obtain the same objective function. This relates to the results in Theorem 10, where in the COAS case, the optimal storage control is restricted to be a feasible optimal primal solution of the ISO's optimization, whereas in the PSO case the private operator has a free choice of storage control within their feasible storage set. It can also be shown that the SOAS and PSO problems for merchant storage and storage-owning generation are also a finite set of convex QPs, and can be derived and solved in a similar manner.

A Note on Dimensionality and Tractability

In the worst case, the number of critical regions is equal to the number of possible combinations of active sets of the optimization problem. However, in many problems only a few active constraints sets generate full-dimensional critical regions inside the feasible parameter set. To keep the number of critical regions low and computable within a reasonable time, it is desirable to keep the dimension of the parameter space as small as possible. This can be achieved through judicious choice of which parameters to include, or 'parameterization of the parameters'.

For the previous examples the parameter sets are intentionally kept as small as possible. For SOAS, the parameter set is of dimension n, and restricted to the set $0 \leq s \leq \overline{s}$. Depending on the volume of storage capacity to be analyzed for the network, the number of resulting critical regions should be reasonable to compute. If further restriction is needed, the location of storage can be confined to fewer buses, reducing the dimensionality of the parameter set.

For PSO, at first glance the parameter set appears to be of dimension nT, however it is in fact only of dimension n. This is due to the fact that since $s = 0$, the individual problems of MPED are now decoupled in time, and represent different instances of the same single period economic dispatch problem. One could solve this problem parametrically with different values of nominal demand, however this may duplicate effort. Another alternative is to further 'parameterize the parameters' to obtain a single problem, assuming the system cost function is not time-varying. In the majority of power systems the demand vector is correlated across time, such that demand can be represented using a constant distribution vector, θ, and deviation vector, v.

$$d_t = \alpha_t \theta + v_t \tag{6.68}$$

where $\alpha_t > 0 \in \mathbb{R}$, $\underline{\alpha} \leq \alpha_t \leq \overline{\alpha}$, $\theta \in \mathbb{R}^n$, $\theta \geq 0$, $\|\theta\|_1 = 1$, and $v_t \in \mathbb{R}^n$, $-\overline{v} \leq v_t \leq \overline{v}$. This results in a single multiparametric program for the single period economic dispatch problem, with parameters α, v, and a parameter set of dimension $n + 1$, defined as

$$\Theta = \{d \in \mathbb{R}^n : d = \alpha\theta + v,\ \underline{\alpha} \leq \alpha \leq \overline{\alpha},\ -\overline{v} \leq v \leq \overline{v}\} \tag{6.69}$$

where for the example in 6.4, $\underline{\alpha} = \min_t\{\overline{d}_t\}$, $\overline{\alpha} = \max_t\{\overline{d}_t\}$, $\overline{v} = \overline{s}$.

As diccused in the previous chapter, relative to simply solving an MPEC, this parametric methodology has some advantages and disadvantages. It requires an intensive offline computation to calculate critical regions and associated functions, with a lightweight online computation to solve a relevant bilevel problem. This makes it useful for situations in which many instances of the problems need to be solved in quick succession, or in offline planning situations where computation time is less important. An advantage is that one automatically has the sensitivity functions dictating how the solution will change in response to deviations of the parameters. This is useful for real-time prediction and planning. A shortcoming of this approach is that is somewhat 'brittle', requiring complete recomputation of all parametric solutions if any major changes occur on the system that are not modelled parametrically. For example, the generation cost function, or the network topology. These variables can be modelled parametrically but will result in an increase in the dimension of the parameter set, incurring the scaling problems described previously. Depending on the intended use of the bilevel problems, operational or planning, a design tradeoff must be made between the set of variables that are modelled parametrically and the dimension and size of the parameter set.

6.5 Main Results

To analyze the payoffs and savings of participants, it is necessary to split the results into two sections:

1. Parameters (s, d) that remain within the small storage set $\mathcal{R}_0$.
2. Parameters (s, d) that lie in a different critical region than $\mathcal{R}_0$.

This distinction is necessary as it turns out that making general statements about participant savings is only possible in the first case.

Within the Small Storage Set $\mathcal{R}_0$

Savings Under OAS

We begin by considering savings under OAS, where for brevity we denote $\widehat{\pi}(s, \overline{d}, u^\star(s, \overline{d})) = \widehat{\pi}(s)$, and $s \in \mathcal{R}_0$ implies that $(s, \overline{d}) \in \mathcal{R}_0$. These results apply to both COAS and SOAS, where the saving under COAS is equal to $\widehat{\pi}(\overline{s})$, and the saving under SOAS is equal to $\max_{s^p}(\widehat{\pi}(s^p))$, where $s^p \in \mathcal{R}_0$, $0 \leq s^p \leq \overline{s}$.

Theorem 11. *For all $s \in \mathcal{R}_0$, the system savings and savings to participants under OAS are as follows*

$$\widehat{J}(s) = \sum_{k\in\mathcal{K}} \mu_k^\star(s,\overline{d})^\top s + \frac{1}{2} s^\top \Delta_k A_k(W_k)\Delta_k s \tag{6.70a}$$

$$\widehat{MP}(s) = \sum_{k\in\mathcal{K}} \mu_k^\star(s,\overline{d})^\top s \tag{6.70b}$$

$$\begin{aligned}\widehat{LP}(s) = \sum_{k\in\mathcal{K}} & 2\mu_k^\star(s,\overline{d})^\top s + s^\top \Delta_k A_k(W_k)\Delta_k s \\ & + s^\top \Delta_k \overline{\lambda}_k(W_k)\end{aligned} \tag{6.70c}$$

$$\begin{aligned}\widehat{GP}(s) = \sum_{k\in\mathcal{K}} & -\frac{1}{2} s^\top \Delta_k A_k(W_k)\Delta_k s \\ & + s^\top \Delta_k A_k(W_k)\left(\sum_{j\in\sigma(k)} Q_j^{-1} a_j\right)\end{aligned} \tag{6.70d}$$

$\widehat{J}(s)$ and $\widehat{MP}(s)$ are both nonnegative and concave in s, $\forall s \in \mathcal{R}_0$. In general, both $\widehat{LP}(s)$ and $\widehat{GP}(s)$ are indefinite, and can take both positive and negative values. We also have that $\widehat{J}(s)$ is a non-decreasing in s, $\forall s \in \mathcal{R}_0$, in particular that

$$\frac{\partial \widehat{J}(s)}{\partial s} = \sum_{k\in\mathcal{K}} \mu_k^\star(s,\overline{d}), \tag{6.71}$$

and that $\widehat{MP}(s)$ is initially non-decreasing in s, in particular that

$$\frac{\partial \widehat{MP}(0)}{\partial s} = \sum_{k\in\mathcal{K}} \mu_k^\star(0,\overline{d}), \tag{6.72}$$

We have the following theorem under stricter conditions.

Theorem 12. *If the generation cost is not time varying, such that $Q_t = Q$, $a_t = a$, there are no consecutive identical LMPs, such that $\mathcal{K} = \mathcal{T}$, and the line congestion W_t is identical in each period t, then*

$$\widehat{LP}(s) \geq 0, \qquad \forall s \in \mathcal{R}_0 \tag{6.73}$$

$$\widehat{LP}(s) = 2\widehat{J}(s), \qquad \forall s \in \mathcal{R}_0 \tag{6.74}$$

$$\widehat{LP}(s) \geq 2\widehat{MP}(s), \qquad \forall s \in \mathcal{R}_0 \tag{6.75}$$

$$\widehat{GP}(s) \leq 0, \qquad \forall s \in \mathcal{R}_0 \tag{6.76}$$

Additionally $\widehat{LP}(s)$ is a non-decreasing, concave function of s, and $\widehat{LP}(s)$ is a non-increasing, concave function of s, $\forall s \in \mathcal{R}_0$.

To digest these results, we first note that unsurprisingly both system savings and merchant storage savings[7] are nonnegative, and positive assuming a non-zero storage capacity and that there are at least two consecutive periods with an LMP difference. This could be deduced a priori, since a non-zero storage capacity enlarges the feasible set of MPED, leading to a nonnegative reduction in system cost, and the merchant storage savings are equal to the storage congestion surplus which we have seen is nonnegative.

A more counterintuitive result is that loads can lose money when their storage (or indeed any storage) is operated to maximize social welfare. A close examination of expression (6.70c), reveals that the only term which is not nonnegative is the final term, which relates to the constant term in the affine expression for the LMP, $\overline{\lambda}_k(W_k)$, and consequently the congestion pattern $\{W_k\}_{k\in\mathcal{K}}$. We will see in a later counterexample that this term can indeed be negative, resulting in negative load savings.

Similarly for the generation savings in (6.70d), we see that the only term which is not nonpositive is the final term, which relates to the coefficients of the generation cost function. We note that if $a_t = 0$, $\forall t \in \mathcal{T}$, then $\widehat{GP}(s) \leq 0$, $\forall s \in \mathcal{R}_0$.

Savings Under SOAS and PSO

Restricting ourselves to the small storage set $\mathcal{R}_0$, and recalling the bilevel problems of Section 6.4, we see that both SOAS and PSO correspond to a single convex QP over the small storage set. The objectives for both SOAS and PSO, are the payoff functions in (6.13), rederived as functions of s^p and u^p respectively, as in the bilevel subsection. To put the following discussion in context, the reader should refer to the load-owned storage payoff maximization problems formulated in Section 6.4.

We first note that while the problems are convex, they are not strongly convex since $A_k(W_k) \succeq 0$, so they will have an optimal solution but the solution may not be unique. Also, since the feasible set of both of these problems is the small storage set $\mathcal{R}_0$, which is some arbitary polytope, there are no general closed form solutions to these problems. However, the inequality in (6.17) always holds, and additionally the welfare savings under either SOAS or PSO are always less than or equal to the welfare savings under COAS. In other words,

$$\widehat{J}(s^p) \leq \widehat{J}(\overline{s}), \qquad \forall s^p, \overline{s} \in \mathcal{R}_0,\ 0 \leq s^p \leq \overline{s} \tag{6.77a}$$

$$\widehat{J}(u^p) \leq \widehat{J}(\overline{s}), \qquad \forall (0, \overline{d} - u^p), \overline{s} \in \mathcal{R}_0,\ u^p \in \mathcal{U}(\overline{s}) \tag{6.77b}$$

We will see from the counterexample in the following section, that neither the inequalities in (6.17) or (6.77) are tight in general. However, an interesting corollorary of Theorem 12 is that, under the conditions in Theorem 12, the COAS and optimal SOAS savings are identical for LSEs, that is $s^{p\star} = \overline{s}$. This can be seen from the fact that the system maximizes welfare savings under COAS using all available storage $\overline{s}$, and the load savings are a linear function

[7]Merchant storage savings are simply equal to the merchant storage payoff, since in the base case with no storage, their payoff is necessarily zero.

of the system savings and so will be maximized by the same use of storage, implying $s^{p\star} = \overline{s}$. Additionally if there is no line congestion in any period, then the optimal savings under PSO are identical to those under COAS and SOAS for LSEs, leading to the following Lemma.

Lemma 3. *If the conditions of Theorem 12 are satisfied, then*

$$\widehat{LP}^{\star}_{COAS} = \widehat{LP}^{\star}_{SOAS} = 2\widehat{J} \tag{6.78}$$

If additionally there is no line congestion in any period, then

$$\widehat{LP}^{\star}_{COAS} = \widehat{LP}^{\star}_{SOAS} = \widehat{LP}^{\star}_{PSO} = 2\widehat{J} \tag{6.79}$$

and

$$s^{p\star} = \overline{s} \qquad s^{p\star}, \overline{s} \in \mathcal{R}_0 \tag{6.80}$$

$$u^{p\star} = u^{\star}(\overline{s}, \overline{d}) \qquad (0, \overline{d} - u^{p\star}), \overline{s} \in \mathcal{R}_0 \tag{6.81}$$

where $s^{p\star}$ and $u^{p\star}$ are the optimal solutions of LP_{SOAS} and LP_{PSO} respectively. So in this special case, the load savings are identical under each regulatory regime and the private operation of storage is welfare maximizing. A general takeaway would be that in networks with very *little* congestion and *small* amounts of storage operated privately by LSEs, the storage is likely to be operated in a similar way to its ideal operation under OAS, and will result in welfare gains that are close to optimal. Additionally, from the final results of Theorem 11, we see that the merchant savings which are equal to the storage congestion surplus are initially increasing in s. Given that welfare savings are a non-decreasing function of s, this implies that the private operation of storage under SOAS will always be welfare maximizing up to some value of storage capacity.

Counterexample

This example demonstrates that even in a simple case, load savings can be negative, generation savings can be greater than load savings, and merchant storage can make a greater profit privately than its profit under OAS. We consider a two bus network, over two periods, with the following cost and network data.

$$Q_t = \begin{bmatrix} 1 & 0 \\ 0 & 10 \end{bmatrix}, \ \forall t \qquad a_t = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \ \forall t \tag{6.82a}$$

$$H = \begin{bmatrix} 0 & -1 \\ 0 & -1 \end{bmatrix}, \qquad c = \begin{bmatrix} 5 \\ 5 \end{bmatrix}, \tag{6.82b}$$

$$\overline{d} = \begin{bmatrix} 3 & 5 \\ 5 & 15 \end{bmatrix}, \qquad \overline{s} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{6.82c}$$

We first want to find the properties of the small storage set for this problem, so we solve $MPED(0, \overline{d})$ numerically using cvx in Matlab with the SDPT3 solver, and obtain the optimal solution, and the line and storage congestion patterns.

$$\lambda^\star(0,\overline{d}) = \begin{bmatrix} 7.27 & 10 \\ 7.27 & 100 \end{bmatrix}, \qquad g^\star(0,\overline{d}) = \begin{bmatrix} 7.27 & 10 \\ 0.73 & 10 \end{bmatrix}, \tag{6.83}$$

$$p^\star(0,\overline{d}) = \begin{bmatrix} 4.27 & 5 \\ -4.27 & -5 \end{bmatrix}, \qquad u^\star(0,\overline{d}) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \tag{6.84}$$

$$W_1 = \begin{bmatrix} \ \end{bmatrix}, \qquad W_2 = \begin{bmatrix} 1 & 0 \end{bmatrix}, \tag{6.85}$$

$$\chi = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \qquad \xi = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \tag{6.86}$$

$$\sigma = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{6.87}$$

We see that there is no line congestion in the first period, and line congestion in the direction of bus 1 to bus 2 in the second period. This solution results in the following social cost, which we refer to as the system payoff, and set of payoffs to participant collectives, assuming that there is no storage on the network, listed in Table 6.1.

Table 6.1: Payoffs in the absence of storage

	Payoff
System	579.1
Merchant Storage	0
Loads	-1608.2
Generation	579.1

We now solve $MPED(\overline{s}, \overline{d})$ to find the COAS solution.

$$\lambda^\star(0,\overline{d}) = \begin{bmatrix} 8.18 & 9 \\ 8.18 & 100 \end{bmatrix}, \qquad g^\star(0,\overline{d}) = \begin{bmatrix} 8.18 & 9 \\ 0.82 & 10 \end{bmatrix}, \tag{6.88}$$

$$p^\star(0,\overline{d}) = \begin{bmatrix} 4.18 & 5 \\ -4.18 & -5 \end{bmatrix}, \qquad u^\star(0,\overline{d}) = \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix}, \tag{6.89}$$

$$W_1 = \begin{bmatrix} \ \end{bmatrix}, \qquad W_2 = \begin{bmatrix} 1 & 0 \end{bmatrix}, \tag{6.90}$$

$$\chi = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \qquad \xi = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \tag{6.91}$$

$$\sigma = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{6.92}$$

We observe that the congestion states are identical for both problem instances, so $(\overline{s}, \overline{d}) \in \mathcal{R}_0$. Since we assume storage capacity at only one bus, this implies that the line segment $0 \leq s \leq \overline{s}$ is also contained in the small storage set.

We then solve the SOAS and PSO problems for each participant collective, resulting in the following optimal storage controls. In this example for each participant collective, although not in general, the optimal storage controls were the same under SOAS and PSO.

$$u^{\star}_{MP} = \begin{bmatrix} -0.71 & 0.71 \\ 0 & 0 \end{bmatrix} \tag{6.93}$$

$$u^{\star}_{LP} = \begin{bmatrix} -0.12 & 0.12 \\ 0 & 0 \end{bmatrix} \tag{6.94}$$

$$u^{\star}_{GP} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{6.95}$$

We see that all private storage controls differ from those obtained under COAS in (6.89). Each set of storage controls result in savings to participants, listed in Table 6.2.

Table 6.2: Participant savings when owning a capacity of storage, $\overline{s}$, under each regulatory regime.

	COAS	**SOAS**	**PSO**
Merchant Storage	0.82	0.97	0.97
Loads	-1.45	0.03	0.03
Generation	-0.95	0	0

We also obtain the welfare savings under each regime when storage is owned by each participant collective, listed in Table 6.3. To gain more insight we plot the private participant

Table 6.3: Welfare savings

	COAS	**SOAS**	**PSO**
Merchant Storage	1.77	1.46	1.46
Loads	1.77	0.31	0.31
Generation	1.77	0	0

savings as a function of the storage capacity at bus 1, s_1. We assume that the storage obeys 'water-filling' behaviour over the two periods, such that $-u_1 = u_2 = s_1$. The crosses on the figure correspond to the optimal savings under private operation (either SOAS or PSO here), listed in Table 6.2, and we see the savings made by each participant under COAS on the right of the figure when $s_1 = \overline{s}_1$.

The generators can never make any savings using any level of storage, so it is in their interest to leave the storage unused. The loads only make a small positive saving at a small level of storage capacity. Merchant storage makes a saving over the full range of storage capacity, but their optimal action is still to withold some capacity relative to the social optimum.

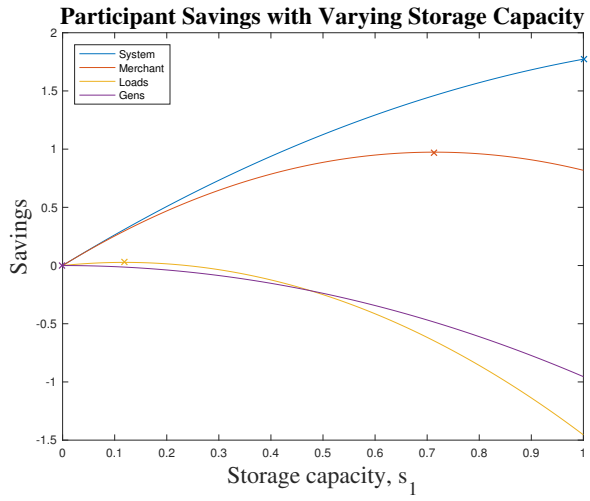


Figure 6.1: Participant savings with varying storage capacity.

For brevity we do not present it here, but a counterexample where generators make savings (additional profits) using storage, which reduces system welfare, has been found for the IEEE 14 bus case, completing the proof of Theorem 11.

Outside the Small Storage Set $\mathcal{R}_0$

The operation of storage, either by the ISO under OAS or by private participants under PSO, can result in a change in the set of active constraints from the base case. This means that the parameters now lie in a different critical region, and the system cost, optimizers, and dual variables are governed by a different set of analytic expressions. As an illustration, in Figure 6.2 we show the critical regions of MPED in s space for the previous example in (6.82), with $d = \overline{d}$ as a constant. Here the small storage set, $\mathcal{R}_0$, corresponds to critical region 2, and we see that $s = \overline{s}$ does indeed lie within this set. To summarize Table 6.4, regions 2, 5 each have the same line congestion pattern, in the direction of bus 1 to bus 2 in the second period. Regions 1, 3, 4, 6 each have the same line congestion pattern, in the direction of bus 1 to bus 2 in both periods. What distinguishes each of the regions with the same line congestion pattern is their storage congestion pattern. With this in mind, only regions 1, 2 are 'interesting', in that regions 3, 4, 5, 6 correspond to cases where the storage capacity is only partially used in the first period, implying that there is some excess storage capacity whose usage would not improve system welfare.[8]

Transmission congestion relief is often touted as a potential benefit of energy storage. However in this case, if additional storage capacity was placed at bus 2, such that the critical region changed from 2 to 1, this would result in an increase in line congestion. In this

[8]This does not necessarily mean that the excess capacity could not be used by a private participant to improve their payoff, however in this example it turns out not to be the case.

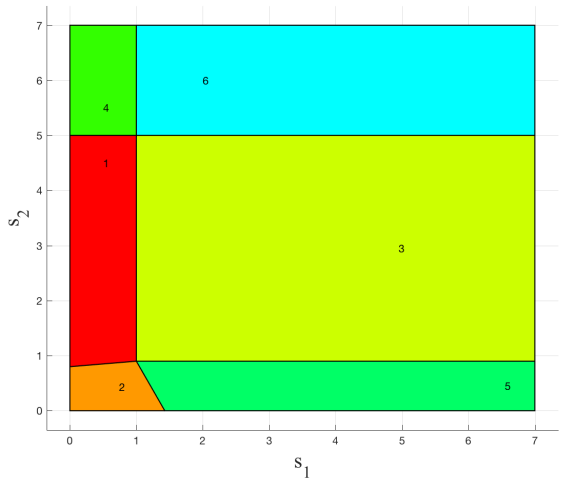


Figure 6.2: Critical regions of Two Bus Example

Table 6.4: Critical Regions

Region	W	χ	ξ
1	$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$
2	$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$
3	$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$
4	$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$
5	$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$
6	$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$

example there is no amount of storage ever able to eliminate line congestion on the network. In general, adding some amount of storage to a network, can both increase or decrease line congestion, depending on the nominal demand profile, network topology, generation cost function, and of course the location of the additional storage. A key insight is that storage operated to maximize social welfare can increase line congestion, so transmission congestion relief and welfare maximization are not necessarily compatible aims for storage development.

Changing critical region will clearly have an impact on participants' payoffs. If we assume that the ISO manages storage under OAS, a change in critical region can result in either savings or losses to a participant, depending on the properties of the new critical region,

other than merchant storage which always makes a nonnegative saving. This result echose that of Theorem 11, where even in the small storage set load and generation savings are indefinite

Theorem 13. *For $s \in \mathcal{R}$, where $\mathcal{R}$ is an arbitrary critical region of MPED, the system savings and savings to participants have the following properties*

$$\widehat{J}(s) \geq 0, \qquad s \in \mathcal{R} \tag{6.96a}$$

$$\widehat{MP}(s) \geq 0, \qquad s \in \mathcal{R} \tag{6.96b}$$

Both $\widehat{LP}(s)$ and $\widehat{GP}(s)$ are indefinite, and can take both positive and negative values.

Proof. It is clear that $\widehat{J}(s) \geq 0$, $\forall s$, since the feasible set of MPED is enlarged. $\widehat{MP}(s) \geq 0$, $\forall s$, since the merchant storage saving is equal to the merchant storage profit, which has been shown to be nonnegative $\forall s$.

For load and generation savings we see similar properties to Theorem 11, with additional complications. We examine the load savings under OAS for a storage capacity $s \notin \mathcal{R}_0$, assuming it lies in some other critical region $s \in \mathcal{R}$,

$$\widehat{LP}(s) = \sum_{t=1}^{T} -\lambda_t^\star(s,\overline{d})^\top \overline{d} + \mu_t^\star(s,\overline{d})^\top s - (-\lambda_t^\star(0,\overline{d})^\top \overline{d}) \tag{6.97a}$$

$$\begin{aligned} &= \sum_{t=1}^{T} \overline{d}_t^\top \left(A_{t,\mathcal{R}_0}(W_{t,\mathcal{R}_0}) - A_{t,\mathcal{R}}(W_{t,\mathcal{R}})\right) \overline{d}_t \\ &\quad + \left(\overline{\lambda}_{t,\mathcal{R}_0}(W_{t,\mathcal{R}_0}) - \overline{\lambda}_{t,\mathcal{R}}(W_{t,\mathcal{R}})\right)^\top \overline{d}_t \\ &\quad + \mu_t^\star(s,\overline{d})^\top s - \overline{d}_t A_{t,\mathcal{R}}(W_{t,\mathcal{R}}) \Delta_{t,\mathcal{R}} s \end{aligned} \tag{6.97b}$$

where we have introduced the additional subscripts $\mathcal{R}_0$, $\mathcal{R}$ to indicate which critical region each $A_t(W_t)$, $\overline{\lambda}_t(W_t)$, Δ_t come from. Given that $A_{t,\mathcal{R}}(W_{t,\mathcal{R}}) \succeq 0$ and $\overline{\lambda}_{t,\mathcal{R}}(W_{t,\mathcal{R}})$ are arbitrary in relation to $A_{t,\mathcal{R}_0}(W_{t,\mathcal{R}_0}) \succeq 0$ and $\overline{\lambda}_{t,\mathcal{R}_0}(W_{t,\mathcal{R}_0})$, the first two terms in (6.97b) are indefinite constants, meaning $\widehat{LP}(s)$ is indefinite in general. A similar result applies to the generation savings. □

In the private case, participants can control their storage and consequently control which critical region will apply to MPED. This allows them to ensure nonnegative savings from storage, satisfying the inequality in (6.19b). The private problem can be solved by the participant as described in Section 6.4. In general it can be beneficial to participants to either decrease or increase the line congestion on the network, as is true for the welfare maximizing ISO. An interesting case arises when participants actively congest lines for their own benefit when the ISO would not have done so (the converse can also occur). This could in general put the power network closer to its limits than it would otherwise need to be, and could also be welfare decreasing.

6.6 Conclusions

In answer to the title of this chapter, in general, for small installations of storage capacity, merchant storage owners and load-owned storage would be wise to stay, and allow the ISO to operate their storage capacity under a regime such as OAS. Generators would be very unwise to invest in storage as they lose money in almost all cases. This advice changes if there are large amounts of installed storage, if the network faces consistent high congestion, or if storage is consistently a price-maker. It is unclear, however whether a private storage operator would have enough information to solve the MPECs described in this chapter, or to consistently take advantage of network congestion to increase its profits, relative to OAS.

We have not addressed other use cases for storage in this chapter, including ancillary services and local load shaping and balancing. For grid services, the ISO could co-optimize the operation of storage across its markets, remaining within an OAS framework. However storage is often much more likely to be a price-maker in ancillary service markets as there is less competition, however, this may change as greater volumes of grid storage are installed. It is less clear how OAS can co-exist with private constraints and objectives, such as commercial load shaping. More work is needed in understanding how the many applications and full value stack of storage can be realized, and their integration with existing wholesale energy markets. Additionally In this chapter we have only addressed cases where a single collective owns storage, and not the more general case of heterogenous storage ownership. To go beyond work such as that in [133, 139] and use the models presented in this chapter would involve solving equilibrium problems of the kind described in Chapter 5. This presents an avenue for numerical study but analytical results will almost certainly not be obtainable.

Chapter 7

Load Flexibility for Congestion Relief

In the previous chapter we addressed the integration of energy storage into electricity markets using OAS, with a focus on energy arbitrage. Here we broaden the application to congestion relief, concentrating just on load flexibility. We will use similar ideas from multiparametric programming, as in the previous section, but here employ them more abstractly using the concept of critical regions without delving into the analytical functions associated with them. We demonstrate in this chapter the set of loads which can be met congestion free, given some amount of load flexibility. This chapter is the result of joint work with Prof. Enrique Baeyens, Prof. Kameshwar Poolla, and Prof. Pravin Varaiya, and was first published in [140]. The initial focus of the paper was *congestion-free dispatch*, *i.e.* can a particular load be met in such a way as to cause no congestion on the network. It is now the opinion of the author that congestion reduction should *not* be a goal in itself, only welfare maximization or profit maximization should be the goal of the ISO or market participant respectively. That is, it may well be in the interest of the ISO or participant to increase network congestion, or indeed reduce it, so as to maximize welfare or their own payoff. No congestion state is arbitrarily any better or worse than another, and should be a safe operating point of the system, assuming security constraints have been set appropriately. That said, the work presents interesting mathematical results as to the set of loads which can be met with a congestion free dispatch, and those that would save under such a dispatch. There is, however, a secondary application of flexibility to congestion relief, in mitigating constraints and component overloads, when the alternative would be curtailing or disconnecting a resource, or investing further in the transmission or distribution network. This is more common to systems without centralized dispatch, as in Europe, where participants self-schedule and balancing is left to a TSO. In the ISO model, constraints are implicitly respected in the a priori optimization of dispatch. This problem represents an interesting avenue of further research, but we do not pursue it here. The two Chapters following this describe mechanisms for transacting flexibility, which could easily be adapted for such a situation where flexibility is contracted to resolve constraints in the distribution network.

7.1 Introduction

Why pay rent? Congestion rents in transmission networks are a real economic expense to participants. We argue that load flexibility can mitigate congestion in electricity grids, whilst generating value for market participants.

Congestion occurs in a network when actual or scheduled flows of electricity over a line are constrained below desired levels. This results in out of merit order generation dispatch. There is firstly a physical capacity constraint due to thermal limits on each transmission line. Additionally line capacity is constrained to ensure continued service under contingency. The latter are usually characterised as stability or security constraints and are determined by system operators to assure an acceptable level of grid reliability. Low-level transmission congestion is very common, however "in more severe congestion conditions, transmission constraints can impair grid reliability by reducing the diversity of available electricity supplies and rendering the grid more vulnerable to unanticipated outages of major generators or transmission lines" [141].

Due to out of merit order generation dispatch, congestion causes prices to vary and typically increase across a network, usually resulting in higher costs for end consumers. For context, from 2008-2013, between 2 and 6% of PJM's total annual billing was attributable to congestion, representing an average annual cost of approximately $1bn [142]. There are several ways to mitigate congestion, for example: reducing electricity demand in the congested area through demand management programs, building additional generation capacity, or building additional transmission capacity. The latter options are highly expensive and decisions are made on very long timescales. The first option is typically much simpler, however little research exists with regards to load flexibility in mitigating congestion.

We propose the idea of a *congestion-free dispatch*, where the resulting network flows cause no transmission constraints to bind. While we refer to a dispatch, the necessary control could be undertaken by the ISO or private participants. The value proposition of a congestion-free dispatch is fourfold. First, it has the potential to generate savings for consumers by eliminating the congestion component of LMPs. However, the issue of savings is nuanced and a congestion free dispatch does not necessarily imply a reduction in total cost to consumers, as the energy component of the locational marginal prices (LMPs) may rise due to the need for dispatching more expensive generators. The effect of a congestion free dispatch on generator revenue must also be considered. Second, price risk would be greatly reduced, either reducing or eliminating altogether the need for financial transmission rights (FTRs), generating savings for parties engaged in bilateral or multilateral contracts [143]. Third, it is perhaps desirable that the grid be operated in a congestion free manner (or even in a reduced flow state) as it increases safety margins in the event of a contingency. Finally, a congestion free dispatch may avoid or delay the need for expensive upgrades to transmission system infrastructure.

In this chapter we investigate the potential of load flexibility in enabling a congestion free dispatch, and examine nuances of the value proposition that can subsequently be derived.

7.2 Problem Formulation

Load Flexibility

In this chapter we take a more specific definition of load flexibility as the ability to shift power consumption in time whilst enforcing a constant energy constraint. That is, any power consumption that is curtailed in one time period is recovered in some other time period, such that total energy consumption remains constant. In practice this represents traditional demand response, where load is shifted in time, typically to periods of lower demand or lower cost.[1]

We adopt the same power system and power flow model as in previous chapters. Rather than consider a discrete, multi-period problem, here we will consider the operation of the power system over some time duration of length T, and assume that the generation dispatch is optimized *continuously* by the ISO. We note the difference between the usage of T here as a length of time, and in previous chapters as the number of multi-period intervals. This is unrealistic as the ISO can only dispatch power in discrete time intervals, however this is a convenience for the presentation of later results, and we will see that all examples and results can be reformulated in a discrete time framework. We have an inelastic nominal demand $\overline{d}(t) \in \mathbb{R}^n$, and flexibility $v(t) \in \mathbb{R}^n$, such that apparent demand $d \in \mathbb{R}^{n \times T}$, is equal to $d = \overline{d} - v$. We will generally assume that nominal demand is constant over the duration T. We assume that there is some ramping constraint $\overline{v} \in \mathbb{R}^n$ on the flexibility, and enforce the zero total net energy constraint, such that the feasible set of flexibility $\mathcal{V}$, is written as

$$\mathcal{V} := \{v(t) : \ -\overline{v} \leq v(t) \leq \overline{v}, \ \int_0^T v(t) = 0, \ 0 \leq t \leq T\} \tag{7.1}$$

In this chapter we will be attempting to find the set of loads for which a congestion free dispatch can be achieved for any ramping capacity of load flexibility. As such, for convenience, we will generally consider $\overline{v} = \infty$, even though flexibility will clearly be bounded below by the nominal demand in any period *i.e.* flexibility cannot generate any power, only reduce or increase consumption. We will return to the issue of ramping constraints in Remark 6.

Economic Dispatch and Congestion

We adopt a single-period economic dispatch model with zero storage capacity described in (2.35), where in the notation of that model, $T = 1$. We assume that this 'single-period' optimization is performed continuously over $t \in [0, T]$. For convenience we now drop the t index, and assume that each quantity referred to below is the result of the continuous

[1] If suitable conditions are met on the ordering of the load curtailment and recovered consumption, then energy storage can also be considered as load flexibility under this definition. *i.e.* The storage must be charged before it can discharge. This ordering constraint can be neglected if we assume some initial state of charge on the storage.

economic dispatch optimization. We denote the generation cost function at each node as $C_i(\cdot)$, such that

$$C(g) = \sum_{i=1}^{n} C_i(g_i) \tag{7.2}$$

We recall the following concepts:

Locational Marginal Prices

The Lagrange multipliers λ_i are known as locational marginal prices (LMPs). If bus i is a net demander, λ_i equals the marginal benefit to consumers, and if bus i is a net supplier, λ_i equals the marginal cost of generation. Since generation costs are assumed to be increasing functions, it must follow that $\lambda_i > 0$. The LMP can be decomposed into three components: a pure energy term, whose value is the same at each node, a loss term, and a congestion term [36]. Since the DC formulation is lossless, the loss term of the LMP does not appear in our analysis. We recall that

$$\lambda = \gamma \mathbf{1} - H^\top \beta \tag{7.3}$$

Merchandising Surplus

The merchandising surplus is defined as

$$MS = -\lambda^\top p = \beta^\top c \tag{7.4}$$

This is the money that is left over after the system operator has paid all generators and collected all revenue from loads. It can be shown that this is always nonnegative for an economic dispatch [37].

Congestion Rent

Since $MS = \beta^\top c$ and β_{ij} is only non-zero when the line constraint on (i, j) is active, it can be concluded that only congested lines contribute revenue to the merchandising surplus[2]. The quantity $\beta^\top c$ is known as the congestion rent, and β_{ij} is known as the congestion price for the associated line flow constraint. The congestion price has a standard interpretation as the marginal value of increased thermal capacity.

Conclusions from Analysis of Economic Dispatch

It can be seen from (2.35c) that in the absence of congestion, $\beta_{ij} = 0 \; \forall \; (i, j)$, and thus all LMPs in the network will be equal, $\lambda = \gamma \mathbf{1}$. This global price will be set by the marginal

[2]This is not strictly true. The total sum of the merchandising surplus is determined solely by congested lines, however congestion causes prices to vary everywhere in the network. Thus there will be uncongested lines which contribute to the MS due to a positive price difference across them, equally balanced by 'negative' contributions from other uncongested lines with a negative price difference across them.

generation in the network.[3] Thus for a congestion free dispatch, generators are dispatched in merit order: least expensive first up to its capacity, then the second least expensive up to its capacity, and so on. For a given set of generators this merit order determines generation selection and cost for a given total load. This dispatch is equivalent to the dispatch for the hypothetical case where the network transmission capacities are unconstrained.

Additionally the absence of congestion also implies that $MS = 0$. Any nonnegative MS represents additional costs to consumers, so a congestion free dispatch should in general be beneficial to consumers. This begs the question then, under what circumstances is a congestion free dispatch possible?

7.3 Feasible Set of Loads for Congestion Free Dispatch

We wish to define the feasible set of loads for which a congestion free dispatch is possible. We first define the economic generation set $\mathcal{G}_E$, then the feasible set of nodal injections $\mathcal{P}$, and finally the congestion free load set $\mathcal{D}_F$. We consider a

Definition 6. *The generation vector $g \in \mathbb{R}^n$ is called feasible if $0 \leq g \leq \overline{g}$. The total power supplied by g is $\mathbf{1}^\top g$, and the total cost is $C(g) = \sum_{i=1}^n C_i(g_i)$. A generation vector $\widehat{g}$ is called a substitute for g if $\mathbf{1}^\top g = \mathbf{1}^\top \widehat{g}$, and $\widehat{g}$ is feasible.*

Definition 7. *The economic generation set is defined as*

$$G_E := \{g : g \text{ is feasible}, C(g) \leq C(\widehat{g}) \ \forall \text{ substitutes } \widehat{g}\}$$

Theorem 14. *We have the following:*

- *Assume linear generator cost functions of the form $C_i(g_i) = \pi_i g_i + \text{constant}$. Let the generators be ordered by increasing marginal cost a_i. Let $\overline{g}_0 = 0$ and $G_k = \sum_{i=0}^k \overline{g}_i$. Then, $g = \begin{bmatrix} g_1 & g_2 & \cdots & g_n \end{bmatrix}^T \in G_E$ if and only if*

$$g_k = \begin{cases} 0 & \text{if } \mathbf{1}^\top g \leq G_{k-1} \\ \mathbf{1}^\top g - G_{k-1} & \text{if } G_{k-1} \leq \mathbf{1}^\top g \leq G_k \\ G_k & \text{if } G_k \leq \mathbf{1}^\top g \end{cases} \tag{7.5}$$

- *Assume generation costs $C_i(\cdot)$ are convex and continuously differentiable, then*

$$G_E = \left\{ g : g \text{ is feasible}, \frac{\partial C_i}{\partial g_i} = \frac{\partial C_j}{\partial g_j}, \forall i, j \in \mathcal{N} \right\}$$

[3] In the case of linear generation costs, assuming no primal degeneracy, there will be a unique marginal generator. In the case of nonlinear generation costs, multiple generators may be marginal.

Remark 3. *The economic generation set is generally non-convex. It is a curve in* $\mathbb{R}^n$*. The construction of* G_E *in Theorem 14(a) is easily extended to piecewise linear cost functions. It suffices to consider each piecewise linear segment* k *of* $C_i(\cdot)$*, defined as* $C_{ik}(g_i) = \pi_{ik} g_i +$ *constant, as a separate generator, located at bus* i*. Then merit ordering is applied based on increasing* a_{ik}*, and* G_E *is constructed accordingly.*

Definition 8. *The feasible set of nodal injections is the feasible injection region* $\mathcal{P}$*, defined in* (2.14)*, and restated here*

$$\mathcal{P} := \{p : \mathbf{1}^\top p = 0,\ Hp \leq c\} \tag{7.6}$$

This set is constrained only by linear inequalities, and is a convex polytope in a subspace of dimension $(n-1)$. It should also be noted that for any $p \in \mathcal{P}$, it is also true that $-p \in \mathcal{P}$.

Definition 9. *The congestion free load set* $\mathcal{D}_F$ *is the set of all feasible loads* $l \in \mathbb{R}^n$ *for which congestion does not occur.*

Theorem 15. $\mathcal{D}_F$ *is characterised as the Minkowski sum of the economic generation set* $\mathcal{G}_E$ *and the feasible nodal injection set* $\mathcal{P}$*.*

$$\mathcal{D}_F = \mathcal{G}_E + \mathcal{P} := \{d : d = g + (-p), g \in \mathcal{G}_E, -p \in \mathcal{P}\}$$

The set $\mathcal{D}_F$ is generally non-convex, due to the nonconvexity of $\mathcal{G}_E$. It should also be noted that $\mathcal{D}_F$ is not restricted to purely positive loads. Momentarily negative loads can be realised through storage. If we preclude this possibility we must restrict $\mathcal{D}_F$ to the nonnegative orthant. For simplicity we will consider elements on the boundary of this set to be congestion free, *i.e.* $\mathcal{D}_F$ is a closed set.[4].

Remark 4. *The set of feasible loads achievable under both congested and uncongested dispatch is* $\mathcal{D} = \mathcal{G} + \mathcal{P}$*, where* $\mathcal{G} := \{g : 0 \leq g \leq G\}$ *is the set of feasible generation. This is due to the fact that a congested dispatch will result in generators being dispatched out of merit order and can require any dispatch in* $\mathcal{G}$*. The set of feasible loads under congestion is defined as* $\mathcal{D}_C = \mathcal{D}/\mathcal{D}_F$*. That is* $\mathcal{D} = \mathcal{D}_F \cup \mathcal{D}_C$*.*

7.4 Two Bus Example

We illustrate these feasible sets using a simple two bus example, shown in Fig. 7.1. We will then consider what economic benefits can be derived using flexibility to achieve a congestion free dispatch.

We have loads, d_1, d_2, generators g_1, g_2, and line capacity C. We assume linear generation costs, and that $\pi_1 < \pi_2$, where π_1, π_2 are the constant marginal costs of generation of g_1 and g_2 respectively. We also define $\overline{g}_1$, $\overline{g}_2$ as the upper generation limits of g_1 and g_2 respectively.

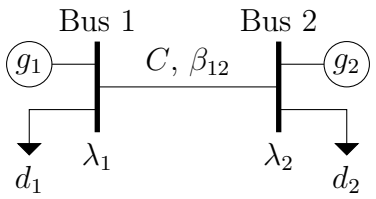
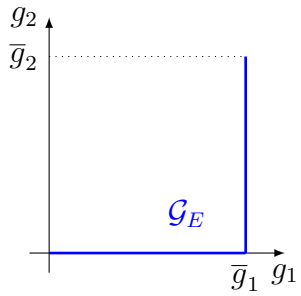


Figure 7.1: Two bus network



(a)

Figure 7.2: Economic generation set, $\mathcal{G}_E$.

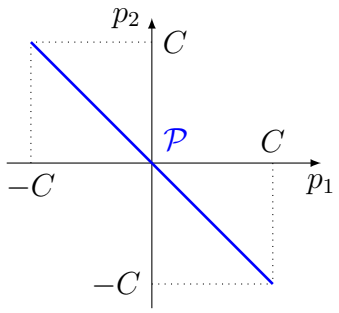


Figure 7.3: Feasible set of nodal injections, $\mathcal{P}$.

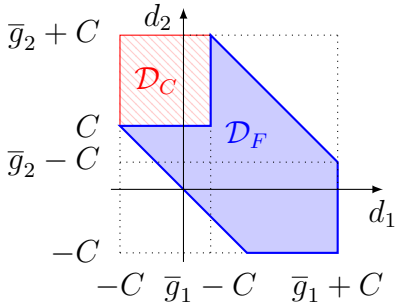


Figure 7.4: $\mathcal{D}_F = \mathcal{G}_E + \mathcal{P}$ in light blue fill, $\mathcal{D}_C$ in light red hatching.

For the purposes of analysis here we will assume that $C \leq \overline{g}_i$. We have LMPs λ_1 and λ_2 at bus 1 and 2 respectively, and a price of congestion β_{12} when the line is congested.

Feasible Sets

Using the analysis of the previous section we find the economic generation set $\mathcal{G}_E$ shown in Fig. 7.2, and the feasible set of nodal injections $\mathcal{P}$ shown in Fig. 7.3.

The feasible set of loads $\mathcal{D}_F$ is the Minkowski sum of these two sets, $\mathcal{G}_E$ and $\mathcal{P}$. This set is non-convex and is shown in light blue fill in Fig. 7.4. We also show the set of feasible loads that entail congestion $\mathcal{D}_C$ in light red hatching. For illustration, in this diagram $\frac{\overline{g}_1}{2} < C < \overline{g}_1$, and $\frac{\overline{g}_2}{2} < C < \overline{g}_2$, although this need not be the case in general.

We see that the non convexity of $\mathcal{D}_F$ arises from the non-convexity of $\mathcal{G}_E$. If $\mathcal{G}_E$ were convex, then $\mathcal{D}_F$ would also be convex, since convexity is preserved under Minkowski addition.

[4]Technically points on the boundary are primal degenerate, where the line is just congested but incurs no congestion rent

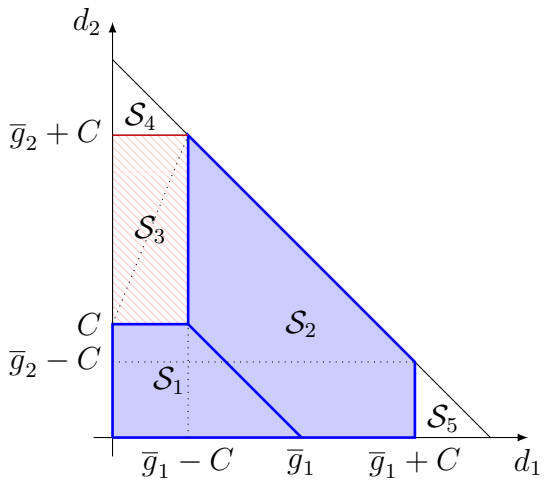


Figure 7.5: Nonnegative congestion free load set $\mathcal{D}_F^+ = \mathcal{D}_1 \cup \mathcal{D}_2$

The Economic Structure of $\mathcal{D}^+$

We now examine the economic effects of a congestion free dispatch enabled by flexibility. We restrict our analysis to nonnegative loads, and denote the nonnegative subset of $\mathcal{D}_F$ as $\mathcal{D}_F^+$. Similarly the nonnegative congested set of loads is denoted $\mathcal{D}_C^+$, and the total nonnegative feasible load set is denoted $\mathcal{D}^+$. This is shown in Fig. 7.5 for some arbitrary C, $\overline{g}_1$, $\overline{g}_2$.

We consider five sets, $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$, $\mathcal{S}_4$, and $\mathcal{S}_5$, where $\mathcal{D}_F^+ = \mathcal{S}_1 \cup \mathcal{S}_2$, $\mathcal{D}_C^+ = \mathcal{S}_3$, and $\mathcal{D}^+ = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3$, as shown in Fig. 7.5.

First, if the transmission line capacity was unconstrained, i.e. $C = \infty$, then $\mathcal{D}^+ = \bigcup_{i=1}^5 \mathcal{S}_i$. In this case $\mathcal{D}^+ = \{d : 0 \leq d_1 + d_2 \leq \overline{g}_1 + \overline{g}_2\}$. *i.e.* We cannot supply a total load greater than total generation capacity. In the case where $C < \overline{g}_i$, as illustrated, $\mathcal{S}_4$ and $\mathcal{S}_5$ are infeasible, since they lie outside the full nonnegative feasible load set $\mathcal{D}^+$.

$\mathcal{S}_1$ contains all load pairs (d_1, d_2) that can be served without the intervention of generator g_2 ($g_2 = 0$), and without congesting the transmission line. The generation structure is $(d_1 + d_2, 0)$. The LMPs are equal to the constant marginal cost of the first generator. That is $\lambda_1 = \lambda_2 = \pi_1$. Since the line is uncongested, the merchandising surplus is zero. Remembering the definition $MS = -\lambda^\top p = \lambda^\top (d - g)$, $MS = \pi_1(d_1 - g_1) + \pi_1 d_2 = 0$.

$\mathcal{S}_2$ contains all load pairs (d_1, d_2) that require the intervention of generator g_2 ($g_2 > 0$) to be served but do not result in congestion of the transmission line. The joint demand $d_1 + d_2$ exceeds the capacity of generator g_1. The generation structure is $(\overline{g}_1, d_1 + d_2 - \overline{g}_1)$. The LMPs are equal to the constant marginal cost of the second generator. That is $\lambda_1 = \lambda_2 = \pi_2$. Once again, since the line is uncongested, the merchandising surplus is zero. That is $MS = \pi_2(d_1 - \overline{g}_1) + \pi_2(d_2 - (d_1 + d_2 - \overline{g}_1)) = 0$.

$\mathcal{S}_3$ contains all load pairs (d_1, d_2) that are feasible but result in congestion of the transmission line. The generation structure is (g_1, g_2) where the generation levels are restricted by the binding transmission constraint. The LMPs are no longer equal as the two markets

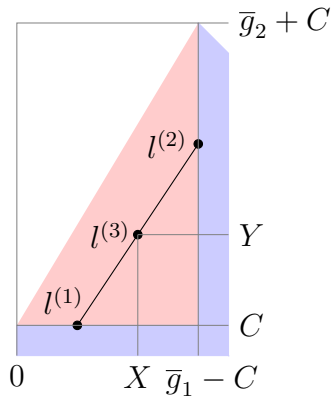


Figure 7.6: $d^{(3)} = (1-\alpha)d^{(1)} + \alpha d^{(2)}$

are effectively decoupled by the congested line, so $\lambda_1 = \pi_1$, $\lambda_2 = \pi_2$. The merchandising surplus is now non-zero. $MS = \pi_1(d_1 - g_1) + \pi_2(d_2 - g_2) = \beta_{12}C > 0$.

The astute reader will notice that $\mathcal{S}_1$, $\mathcal{S}_2$, and $\mathcal{S}_3$ are critical regions of the Economic dispatch problem, with load d as a parameter, as described in the previous chapters. Here in this simple case we have analytically characterized their structure, but for a more general problem the critical regions can be calculated recursively.

The Value of Flexibility

To illustrate the value of flexibility we will define three loads, shown in Fig. 7.6. The red triangular area in Fig. 7.6 is the set $\mathcal{S}_3 \cap \text{conv}(\mathcal{D}_F^+)$, shown by the upper dotted line in Fig. 7.5, where $\text{conv}(\mathcal{D}_F^+)$ denotes the convex hull of $\mathcal{D}_F^+$. We also

$$\begin{aligned} d^{(1)} &= \left(\frac{X - \alpha(\overline{g}_1 - C)}{(1-\alpha)}, C \right) \in \text{bd}(\mathcal{S}_1) \\ d^{(2)} &= \left(\overline{g}_1 - C, \frac{Y - (1-\alpha)C}{\alpha} \right) \in \text{bd}(\mathcal{S}_2) \\ d^{(3)} &= (X, Y) \in \mathcal{S}_3 \cap \text{conv}(\mathcal{L}_F^+) \\ &= (1-\alpha)d^{(1)} + \alpha d^{(2)} \end{aligned}$$

where $\alpha \in (0, 1)$, and $\text{bd}(\mathcal{S})$ denotes the boundary of the set $\mathcal{S}$. $d^{(3)}$ is an arbitrary point in the congested load set $\mathcal{D}_C^+$, with coordinates (X, Y). $d^{(3)}$ also lies on the line segment connecting $d^{(1)}$ and $d^{(2)}$. This suggests that if $d^{(3)}$ is constant over some time period, it can be met congestion free using flexibility. In other words suppose we need to serve a constant load $d^{(3)}$ over some time period T. Since $d^{(3)}$ cannot be served without congesting the network, suppose the loads admit flexibility and we dispatch $d^{(1)}$ over a time period $(1-\alpha)T$ and $d^{(2)}$ over a time period αT.

Scenario	$(1-\alpha)T$	αT	Congestion free
Non-flexible	$d^{(3)}$	$d^{(3)}$	No
Flexible	$d^{(1)}$	$d^{(2)}$	Yes

To demonstrate the economic effects of a flexible and congestion free dispatch, we consider each class of market participants as a collective and calculate the system cost (SC),[5] the generation revenue (GR), the load payment (LP), and the merchandising surplus (MS) resulting from each load pair. This is shown in Table 7.1. It should be noted that $MS = LP - GR$.

Table 7.1: Data for uncongested and congested load pairs

	$d^{(1)}$	$d^{(2)}$
SC	$\frac{\pi_1}{(1-\alpha)}(X - \alpha\overline{g}_1 + C)$	$\frac{1}{\alpha}(\pi_1\alpha\overline{g}_1 + \pi_2(Y - C))$
GR	$\frac{\pi_1}{(1-\alpha)}(X - \alpha\overline{g}_1 + C)$	$\frac{\pi_2}{\alpha}(\alpha\overline{g}_1 + Y - C)$
LP	$\frac{\pi_1}{(1-\alpha)}(X - \alpha\overline{g}_1 + C)$	$\frac{\pi_2}{\alpha}(\alpha\overline{g}_1 + Y - C)$
MS	0	0

	$d^{(3)}$, Non-flexible Scenario
$SC(n)$	$\pi_1 X + \pi_2 Y - (\pi_2 - \pi_1)C$
$GR(n)$	$\pi_1 X + \pi_2 Y - (\pi_2 - \pi_1)C$
$LP(n)$	$\pi_1 X + \pi_2 Y$
$MS(n)$	$(\pi_2 - \pi_1)C$

We now consider the two scenarios described above, denoting the non-flexible scenario (n), and the flexible scenario (f). Let us assume without loss of generality that $T = 1$, in which case the results for (n) are the same as for $d^{(3)}$ above. For (f) we dispatch $d^{(1)}$ over a time period $(1-\alpha)$ and $d^{(2)}$ over a time period α. The results for this scenario are shown in Table 7.2.

Table 7.2: Data for Flexible Scenario

	Flexible Scenario
$SC(f)$	$\pi_1 X + \pi_2 Y - (\pi_2 - \pi_1)C$
$GR(f)$	$\pi_1 X + \pi_2 Y + (\pi_2 - \pi_1)(\alpha\overline{g}_1 - C)$
$LP(f)$	$\pi_1 X + \pi_2 Y + (\pi_2 - \pi_1)(\alpha\overline{g}_1 - C)$
$MS(f)$	0

[5]This is the value of the objective function solved by the system operator *i.e.* the fuel cost of the generators.

Examining the differences between the two scenarios, we see that

$$\begin{aligned} SC(f) - SC(n) &= 0 \\ GR(f) - GR(n) &= (\pi_2 - \pi_1)\alpha\overline{g}_1 \\ LP(f) - LP(n) &= (\pi_2 - \pi_1)(\alpha\overline{g}_1 - C) \end{aligned}$$

Some Conclusions

1. The loads receive the same total energy in both cases, since $d^{(3)} = (1-\alpha)d^{(1)} + \alpha d^{(2)}$.
2. The system cost is identical in both cases.
3. Under flexibility, the generator collective receives a nonnegative profit $(\pi_2 - \pi_1)\alpha\overline{g}_1$.
4. Under flexibility, the load collective makes savings if $LP(f) < LP(n)$. This only occurs if $\alpha < C/\overline{g}_1$.

Referring to Fig. 7.6, it is clear that the choice of α is non-unique. Varying α will change the relative location of $d^{(1)}$ and $d^{(2)}$, however the line segment between them will always pass through $d^{(3)}$.[6] It is also clear that for any arbitrary load $d^{(3)} \in \mathcal{S}_3 \cap \text{conv}(\mathcal{D}_F^+)$ there exists a minimum and maximum allowable α, such that $d^{(1)} \in \text{bd}(\mathcal{S}_1)$ and $d^{(2)} \in \text{bd}(\mathcal{S}_2)$. This, combined with the condition that $\alpha < C/\overline{g}_1$ for the load collective to make savings, defines a set, shown in green in Fig. 7.7. Any load in this set can pick a valid α such that $d^{(1)} \in \text{bd}(\mathcal{S}_1),\ d^{(2)} \in \text{bd}(\mathcal{S}_2),\ \alpha < C/\overline{g}_1$, and the load collective makes savings under flexibility. Thus in general, the proportion of congested loads that can make savings using flexibility will be determined by a subset of the network parameters, in this case the line capacity C and the upper limit of the cheapest generator $\overline{g}_1$.

These economic transfers to loads and generators represent a redistribution of the original merchandising surplus. This can be seen from

$$\begin{aligned} MS(n) - MS(f) &= (LP(n) - GR(n)) - (LP(f) - GR(f)) \\ &= (LP(n) - LP(f)) + (GR(f) - GR(n)) \\ &= LS + GP \end{aligned}$$

where LS denotes load savings, and GP denotes generator profits. This redistribution of the merchandising surplus is shown in Fig. 7.8, with varying α. This also clearly illustrates the transition point at which the load collective begins losing money under flexibility. The generator collective in constrast, only ever makes a profit under flexibility.

We thus claim that in the general case, if the network topology and line capacity constraints result in savings for the load collective under a flexible uncongested dispatch (*i.e.* in the two bus case $\alpha < C/G_1$), then this dispatch does not disadvantage any class of market participants, and will result in benefits for all collective classes of market participants.

[6] One can imagine the points $d^{(1)}$ and $d^{(2)}$ sliding along the boundaries of sets $\mathcal{S}_1$ and $\mathcal{S}_2$ respectively.

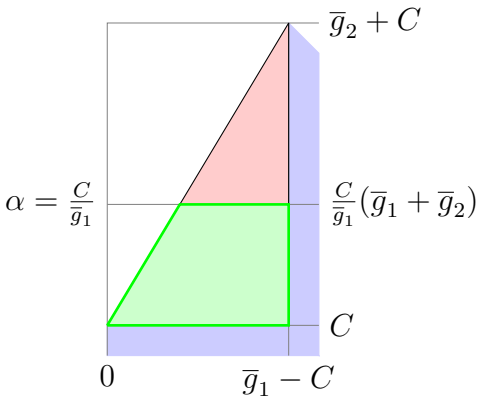


Figure 7.7: Set of congested loads which make savings under flexibility

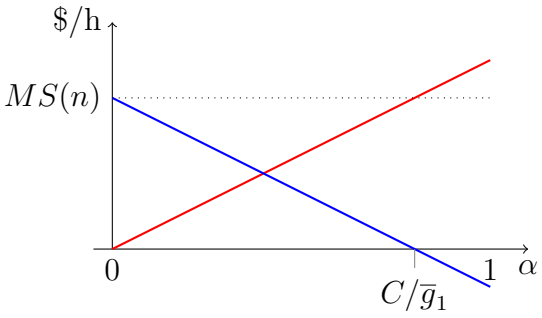


Figure 7.8: Load savings in blue, generator profits in red

7.5 General Network Results

We have shown that, under certain network conditions, flexibility has value by enabling congestion free dispatch. It now makes sense to ask what set of nominal loads can be met congestion free, using flexibility? The following results apply to any general network.

Constant Nominal Load

Theorem 16. *By allowing flexibility at each node, the feasible set of nominal loads for a congestion free dispatch $\mathcal{D}_F$ is enlarged to its convex hull, conv($\mathcal{D}_F$), as the number of time intervals of flexibility tends to infinity.*

To prove this theorem, we shall make use of the Shapley-Folkman Theorem, which states that the Minkowski sum of an increasing number of subsets of a metric space converges to its convex hull.

Theorem 17 (The Shapley-Folkman Theorem [144])**.** *For some set* $S \subset \mathbb{R}^n$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} S = conv(S) \tag{7.7}$$

Proof Consider a nominal load vector $\overline{d} \in \mathbb{R}^n$ that is constant over some time period T. Divide T into N equal time intervals t_k, where $t_k = T/N$, $k = 1, \ldots, N$.

t_1	t_2	$\cdots$	t_N
$\overline{d}$	$\overline{d}$	$\cdots$	$\overline{d}$

Now consider load flexibility in each time interval, denoted v_k, such that the load vector in time interval t_k is equal to $\overline{d} - v_k = d_k$. We now have

t_1	t_2	$\cdots$	t_N
d_1	d_2	$\cdots$	d_N

The net zero energy constraint requires that $\sum_{k=1}^{N} v_k = 0$. Thus,

$$\frac{1}{N} \sum_{k=1}^{N} d_k = \frac{1}{N} \sum_{k=1}^{N} (\overline{d} - v_k) = \overline{d} \tag{7.8}$$

Assume that $|v_k|$ is large enough such that $d_k \in \mathcal{D}_F,\ \forall k$. Thus $\overline{d} \in \frac{1}{N} \sum_{k=1}^{N} \mathcal{D}_F$, where the summation is a Minkowski addition.

By the Shapley-Folkman Theorem, if $\overline{d} \in \frac{1}{N} \sum_{k=1}^{N} \mathcal{D}_F$, as $N \to \infty$, $\overline{d} \in \text{conv}(\mathcal{D}_F)$. Consequently, the congestion free load set $\mathcal{D}_F$ is enlarged to the convex hull of $\mathcal{D}_F$ using flexibility.

Remark 5. *This general result can perhaps be more easily seen by considering Carathéodory's Theorem, which states that any point* $\overline{d}$ *in the convex hull of the set* $\mathcal{D}_F$ *can be written as the convex combination of at most* $l + 1$ *points in* $\mathcal{D}_F$ *where* $l = \dim(\mathcal{D}_F) = n$.

$$\begin{gathered} \overline{d} = \sum_{j=1}^{l+1} \alpha_j d_j,\ \overline{d} \in conv(\mathcal{D}_F),\ d_j \in \mathcal{D}_F, \\ \sum_{j=1}^{l+1} \alpha_j = 1,\ \alpha_j \geq 0,\ \forall j \end{gathered} \tag{7.9}$$

Intuitively the constants α_j *represent the fraction of time that must be spent at each load vector* d_j *to meet the nominal load vector* $\overline{d}$ *over some time period* T.

$\alpha_1 T$	$\alpha_2 T$	$\cdots$	$\alpha_{d+1} T$
d_1	d_2	$\cdots$	d_{l+1}

The α_j *are equal to either a finite or infinite sum of time intervals* t_k. *This correspondence between the two results is stated in the following theorem.*

Theorem 18. *If all α_j are rational numbers ($\alpha_j \in \mathbb{Q}$, $\forall j$), then any constant load vector $\overline{d}$ can be met in a finite number of time intervals of flexibility. If any α_j is an irrational number, then n can only be met by an infinite number of time intervals of flexibility.*

Proof If $\alpha_j \in \mathbb{Q}, \forall j$, then we can write $\alpha_j = \frac{m}{n}$, where $0 \leq m < \infty$, $0 < n < \infty$, $m, n \in \mathbb{Z}$. Knowing that $\alpha_j \leq 1$, necessarily $m \leq n$, and we can write $\alpha_j = \frac{m_j}{N}$, where $0 \leq m_j < N$, $0 < N < \infty$. N is the total number of time intervals t_k, and m_j is the number of time intervals t_k spent at point d_j. It follows that $\sum_{j=1}^{l+1} m_j = N$. To prove for irrational numbers, following a similar argument it suffices to see that there does not exist a finite N, such that α_j can be written as the quotient of a number of time intervals m_j and a total number of time intervals N.

Remark 6. *For Theorem 16 to hold in general, the minimum amount of flexibility v_i required at each node $i = 1, \ldots, n$ is*

$$|v_i| = \sup\{d(e_i^T v, e_i^T w) : v, w \in conv(\mathcal{D}_F)/\mathcal{D}_F\} \tag{7.10}$$

where $d(\cdot, \cdot)$ is the Euclidean distance function, and e_i is the standard basis vector with a 1 in the ith coordinate and 0 elsewhere.

Varying Nominal Load Profile

Consider a varying nominal load profile for a network over some time period T. Divide T into N equal time intervals t_k, where $t_k = T/N$, $k = 1, \ldots, N$. We assume a constant nominal load in each time interval t_k, denoted $\overline{d}_k \in \mathbb{R}^n$, $k = 1, \ldots, N$.[7]

$$\begin{array}{c|c|c|c} t_1 & t_2 & \cdots & t_N \\ \hline \overline{d}_1 & \overline{d}_2 & \cdots & \overline{d}_N \end{array} \tag{7.11}$$

There exists some average nominal load for the time period $\overline{d}$, where $\overline{d} = \frac{1}{N} \sum_{k=1}^{N} \overline{d}_k$.

Lemma 4. *If $\overline{d} \in \mathcal{L}_F$ then, by allowing flexibility at each node, the nominal load profile can always be met for the whole time period under a congestion free dispatch.*

Proof The proof follows by checking that $v_k = \overline{d} - \overline{d}_k$ is a flexibility sequence satisfying the total energy constraint $\sum_{k=1}^{N} v_k = 0$. The load in each time interval t_k is now $\overline{d}_k - v_k = \overline{d}$, and since $\overline{d} \in \mathcal{D}_F$, the load sequence $\overline{d}_k$ can be satisfied using a flexibility sequence v_k, congestion free.

Theorem 19. *If $\overline{d} \in conv(\mathcal{D}_F)$ then, by allowing flexibility at each node, the nominal load profile can always be met for the whole time period under a congestion free dispatch, as the number of time intervals of flexibility tends to infinity.*

[7] To represent a continuous load profile we would need $N \to \infty$, however it is not unreasonable to assume a constant load in a finite time period. For example most ISOs forecast constant loads in fixed time periods.

Proof This proof follows directly from Theorem 16 and Lemma 4. Using an initial flexibility sequence $v_k^{[1]}$, we transform the nominal load profile $\overline{d}_k$ into a constant load sequence, having value $\overline{d}$, $\forall k$. We now have the same case as in Theorem 16, where we have a constant nominal load vector at each time interval t_k. Thus the congestion free load set $\mathcal{D}_F$ can be enlarged to the convex hull of $\mathcal{D}_F$ using some flexibility sequence $v_k^{[2]}$, over an infinite number of time intervals. Thus if $\overline{d} \in \text{conv}(\mathcal{D}_F)$, then the nominal load sequence $\overline{d}_k$ can be met using the flexibility sequence $v_k^{[1]} + v_k^{[2]}$, congestion free.

7.6 Conclusions

We have proved that flexibility enables congestion free dispatch by enlarging the feasible set of congestion free loads. We have shown that flexibility has value since under certain network conditions a congestion free dispatch brings economic benefits to the market participants (loads and generators). If participant classes are considered as collectives, then no class of participants is economically disadvantaged. Any merchandising surplus that existed under a congested dispatch is fully recovered by market participants under a congestion free dispatch.

To summarise, the loads are effectively arbitraging on LMPs to redistribute the merchandising surplus in a beneficial way, using flexibility. This raises several questions. Firstly, would this be considered market manipulation? Indeed in these schemes market participants are actively influencing prices. However, we have shown that if the participant classes are considered as collectives then no market participant is economically disadvantaged, and the system cost is identical in both cases. Potential losers who are not active market participants are transmission line owners. Since there is no merchandising surplus under a congestion free dispatch, transmission lines no longer generate any revenue. A new mechanism to compensate transmission line owners must be designed.

There is also the question of incentives and coordination. Assuming this new paradigm, what are the incentives for loads and generators to act in an optimal manner? Is total coordination of loads needed to achieve a congestion free dispatch, or is only partial participation necessary? What are the effects of a congestion free dispatch on market power? Clearly the economics of flexibility are nuanced and more comprehensive study is required in this direction.

Chapter 8

Blockchains for Decentralized Optimization of Energy Resources in Microgrid Networks

In previous chapters we have addressed strategic behavior and the integration of flexible resources in transmission networks. In the following two chapters we turn to the distribution network, and specifically the question of market design for distributed energy resources (DERs). In this chapter we consider the emerging role of blockchain technology in the management and coordination of DERs, and, crucially, what a valid use case for the technology might be. The work also has implications beyond energy management, suggesting how distributed optimization could be secured using a blockchain. This chapter is the result of joint work with Dr. Eric Munsing, and Prof. Scott Moura, and was first published in [145]. It was acknowledged as the first published work to examine the use of blockchains to facilitate distributed optimization and control of DERs in [146].

8.1 Introduction

The energy production landscape is being reshaped by DERs — photovoltaic panels, electric vehicles, smart appliances, and battery storage systems, which provide low-voltage energy services and are often remotely controllable as part of the Internet of Things. When used intelligently, these DERs can reduce cost, improve reliability, and integrate renewable resources in the electric grid — features which have led regulators to introduce policies promoting their adoption [147, 148].

However, payments for DER services must be negotiated with electric utilities, monopolies who may be invested in preserving conventional generation systems. As a result, the deployment of DERs has often been met with animosity by utilities, which may bar the participation of DERs or seek monopoly rents in return for access to the distribution infrastructure [149, 150].

Local distribution markets for energy services have been proposed as a means of efficiently incentivizing and dispatching DERs, much as is done at the transmission scale [5, 151]. However, such a local distribution market would need to address both the monopoly incentive issues highlighted above, and also the abuses of market power observed in wholesale energy markets [152, 153]. These issues would be particularly pertinent in microgrid operation, which may not benefit from the scrutiny given to a larger utility [154].

Engineering literature has examined control schemes for providing energy services with DERs in distribution grids and microgrids [155], but has typically assumed operation is managed by a benevolent aggregator or cooperative utility. It is unclear how these proposed control systems would work in the face of incentive issues or regulatory shortcomings.

To address these issues, we leverage an emerging technology which has been developed to allow decentralized consensus between non-trusting agents: blockchains and smart contracts. Despite extensive use in financial applications for addressing trust issues [156], blockchains have seen limited deployment in the energy space [157] and have not been considered for coordinating DERs to manage network constraints [158].

We examine how a blockchain architecture can be used to distribute the aggregator's role across all devices on a microgrid network. This integrated architecture is demonstrated on a blockchain platform controlling a microgrid simulation, and demonstrates how to address incentive issues while respecting operational constraints.

The remainder of the chapter is structured as follows: Section 8.2 provides a brief overview of blockchains and smart contracts. Section 8.3 provides a survey of previous literature on dispatch of DERs in microgrids, decentralized optimization techniques, and blockchain use in energy applications. Section 8.4 presents the formulation of the optimal power flow problem with DERs and its ADMM equivalent, Section 8.5 describes the algorithm for utilizing a blockchain for securing the decentralized problem, Section 8.6 presents results from a simulation network, and Section 8.8 concludes.

8.2 Blockchains and smart contracts

Blockchains are an emerging technology for decentralized computation and data storage, secured by a combination of cryptographic signatures and a distributed consensus mechanism. Participants on the blockchain network are able to come to universal agreement on the system state σ^t at each time step t, even in the presence of cyberattacks, communication dropouts, and participants joining/departing the network. This is in stark contrast to conventional architectures where a central coordinator defines the state of the system, but may be subject to attack or malfeasance.

The general architecture of blockchains is described in [159] and illustrated in Fig. 8.1. Participants on the peer-to-peer network broadcast messages $M_i^t, i \in (1, \ldots, \mathrm{N_m})$. These messages contain commands which affect the state of the system (control actions, account withdrawals, etc), and the feasibility of each message can be checked by each node using a validation function $\pi(\sigma^{t-1}, M_i^t)$.

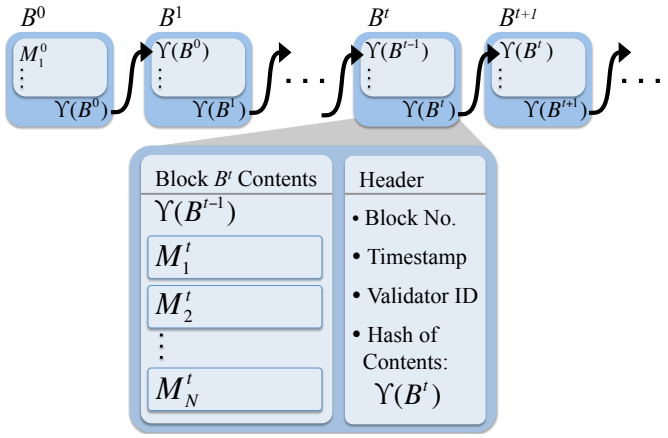


Figure 8.1: Symbolic representation of the data in a blockchain, showing blocks B^0 to B^{t+1} with detail of block B^t. Blocks are linked by their cryptographic hashes $\Upsilon(B^t)$, securing the contents from alteration and allowing transparent auditing of system history. Messages M_i^t contain information about changes to the system state, such as energy transfers or payments.

Participants listen to the network and collect a set of messages into the contents of the next block B^t. A block header H is formed which contains the timestamp, a concise *cryptographic hash* $\Upsilon(B^{t-1})$ of the contents of the previous block, and the results of a verification test that is computationally or economically difficult to forge. The new block is broadcast to the network, where its validity is checked and nodes reach consensus on the updated state of the system $\sigma^t = \Pi(\sigma^{t-1}, B^t)$. The utility of blockchains can be significantly expanded when the state transition function $\Pi(\cdot)$ can execute computer code embedded in the transmissions M_i. These *smart contracts* can be transparently inspected and audited, and are guaranteed to be faithfully executed on the network.

Recursively linking the contents of blocks, verifying new blocks with peer-to-peer consensus, and using cryptographic signatures to verify communication are the pillars of blockchain architecture. Together, they provide an immutable and robust representation of system state-without requiring the intervention of a trusted central authority. While this architecture introduces some computational overhead, it offers immutability, transparency, and verifiability which can make the system well suited for coordination between parties who do not trust each other. The reader is referred to [156, 158, 160] for additional details on the security, architecture, and applications of blockchains and smart contracts.

8.3 Prior Literature

This work draws on three bodies of research: control of distributed energy resources, the economics and regulation of microgrids, and research on blockchains and smart contracts. We provide a brief summary of relevant literature from each domain.

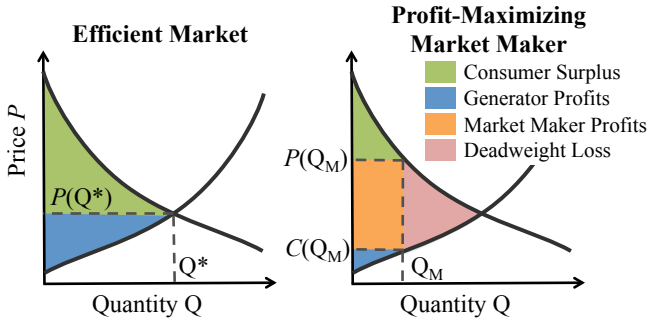


Figure 8.2: Comparison of an efficient market in which the optimal quantity $Q*$ is cleared at price $P(Q*)$, and a market operated by a monopoly who is able to charge separate prices for generation and consumption. In this model, the monopoly restricts output to Q_M, purchasing energy at $C(Q_M)$ and charging consumers $P(Q_M)$.

DER Control

Microgrids are electricity networks which can be controlled autonomously, and may operate in both grid-connected and self-sufficient modes [161]. Without the benefit of a large balancing territory, loads and generation must be coordinated carefully and the role of DERs becomes particularly important. Surveys of approaches to microgrid management can be found in [155, 162].

Conventionally, generation resources have been centrally controlled by a utility or system operator — but these centralized approaches do not scale well to large numbers of DERs, and recent research has focused on decentralized algorithms with low computational overhead. Decentralized algorithms have been explored for coordinating electric vehicles [163, 164], smart inverters [165], and for fleets of diverse DERs [166–168].

Constraints on network voltage and power power flows can become significant at high DER penetrations, and decentralized models for power flow in distribution systems have been explored in [169–171]. As the underlying AC optimal power flow problem (OPF) problem is nonconvex, each of these examines different assumptions or relaxations which grant computational tractability.

Microgrids and Monopoly Economics

In prior literature, DERs are compensated for providing energy services by an aggregator or a utility: a central authority who is trusted to act fairly in scheduling generators, satisfying loads, and rendering payments.

Like conventional electrical utilities, a microgrid operator faces a set of competing demands: minimizing consumer costs, investing in reliability and long-term capacity, and providing a return for shareholders [172]. Even without owning any assets, such a monopoly aggregator can have strong incentives to shift the market away from a cost-minimizing equilibrium and towards a profit-maximizing monopoly outcome, as shown in Figure 8.2. These

conflicts of interest are typically controlled through regulatory intervention, where auditors scrutinize market outcomes and regulate customer fees [173].

However, when regulatory efforts are expensive, a small degree of market inefficiency may be less burdensome than regulatory costs [174]. This can create distrust between the microgrid operator and producers/consumers, who cannot assess whether their bills reflect monopoly profits or justified costs [147, 154].

This trust issue is already visible in the integration of rooftop photovoltaic systems in distribution networks [149, 150], and can be expected to be a greater problem in microgrids if regulatory scrutiny cannot be efficiently implemented for small systems — for example, if regulation has a high fixed cost (such as for retaining auditors) [174].

Blockchains and Energy

Blockchain research is still a new field, with most existing work focused on security and scalability [156,158] and few applications for controlling physical devices [160,175]. Although blockchains rely on a distributed consensus mechanism to provide security, the parallels with decentralized consensus algorithms in engineering control and optimization research have not yet been explored.

While blockchains have been discussed for use in coordinating DERs in transactive energy markets [176, 177], these works have not considered physical constraints on DER operation – instead treating DERs as idealized financial assets [178–180]. In reality, any coordination system must consider the DER's own constraints as well as the constraints of the distribution network. Prior literature has not considered methods for addressing these constraints in blockchain applications.

Novel Contributions

With this background, blockchains and smart contracts hold unexplored potential for eliminating trust issues with microgrid operators, and as a natural platform for coordinating the decentralized optimization schemes described above. The following contributions extend prior literature:

- Distributed optimal power flow algorithm with batteries, shapable loads, and deferrable loads
- Recovery of distributed locational marginal prices from a decentralized OPF problem
- Use of a blockchain for coordinating devices with operational constraints
- Use of a blockchain to facilitate the aggregator step of a decentralized optimization algorithm

8.4 Problem Formulation

We consider a microgrid with a dispatchable central generator, uncontrolled plug loads, non-dispatchable renewable energy resources, shapable loads (e.g. electric vehicles), deferrable loads (e.g. appliances), and batteries. We consider a day-ahead scheduling problem, with the objective of minimizing cost of energy provision subject to the operational constraints of the DERs and of the distribution network.

Network and Power Flow Model

We adopt the Distflow model described in 2.2, using the SOCP relaxation. This gives the feasible injection region as

$$\mathcal{P} := \{P, Q, l, v : \underline{\mathrm{v}} \leq v \leq \overline{\mathrm{v}},\ -\hat{c} \leq P \leq \hat{c}\} \tag{8.1}$$

Controllable DERs

We consider a set of energy resources with complex injections/withdrawals s placed at nodes i throughout the microgrid network, denoted as follows:

s_i^g	Dispatchable generators	s_i^u	Uncontrollable loads
s_i^r	Renewable generators	s_i^d	Deferrable loads
s_i^b	Stationary batteries	s_i^s	Shapable loads

The net complex injection at a node i in period t is

$$s_i(t) = s_i^g(t) - s_i^l(t), \quad i = 0, \ldots, n \tag{8.2}$$

where

$$s_i^l(t) = s_i^u(t) + s_i^d(t) + s_i^s(t) - s_i^b(t) - s_i^r(t) \tag{8.3}$$

Dispatchable Generation

Dispatchable generators (e.g. microturbines, diesel generators, fuel cells) are considered to have quadratic increasing cost, which may be time-varying:

$$C_{i,t}(s_i^g(t)) = \alpha_{i,t} s_i^g(t)^2 + \beta_{i,t} s_i^g(t) + \gamma_{i,t} \tag{8.4}$$

We assume that each dispatchable generator has capacity limits, described as

$$\underline{\mathrm{s}_i^g} \leq s_i^g(t) \leq \overline{\mathrm{s}_i^g}], \quad i = 0, \ldots, n, \quad t = 1, \ldots, T \tag{8.5}$$

Power injection from renewable generators is considered to be deterministic and have no marginal cost $C_{i,t}(s_i^r(t)) = 0$.

Uncontrollable Loads

Power withdrawals due to uncontrollable loads (lights, plug loads) are considered deterministic, inflexible, and inelastic. We do not model thermostatically controlled loads or smart inverters, though those can be added to the formulation using the approaches in [181] and [171] respectively.

Stationary batteries are modeled as dispatchable loads which can be controlled to withdraw power ($s_i^b < 0$) or inject power ($s_i^b > 0$). We assume charging efficiency $\eta_{i,\text{in}}$, and discharging efficiency $\eta_{i,\text{out}}$. We assume that the battery should not undergo a net discharge of more than ε over the course of the dispatch period.

$$\forall t = 1 \dots T :$$
$$s_i^b(t) = d_i^b(t) - c_i^b(t) \tag{8.6a}$$
$$0 \leq c_i^b(t) \leq \overline{c}_i^b \tag{8.6b}$$
$$0 \leq d_i^b(t) \leq \overline{d}_i^b \tag{8.6c}$$
$$\underline{E}^b \leq E^b(t) \leq \overline{E}^b \tag{8.6d}$$
$$E_i^b(t) = E_i^b(t-1) + c_i^b(t)\Delta\text{t}\eta_{\text{i,in}} - d_i^b(t)\Delta\text{t}/\eta_{\text{i,out}} \tag{8.6e}$$
$$(1-\varepsilon)E_i^b(1) \leq E_i^b(T) \leq (1+\varepsilon)E_i^b(1) \tag{8.6f}$$

Shapable loads (e.g. electric vehicles with continuous charging levels, continuously variable fans) are modeled as having net energy demand E_i^s, and must be charged between times $\text{t}_{i,\text{startby}}$ and $\text{t}_{i,\text{endby}}$:

$$\underline{s}_i^s \leq s^s(t) \leq \overline{s}_i^s \quad \forall t = 1 \dots T \tag{8.7a}$$
$$\sum_{t=1}^{T} s^s(t) = E_i^s \tag{8.7b}$$
$$s_i^s(t) = 0 \quad \forall t = 1, \dots, \text{t}_{i,\text{startby}} \tag{8.7c}$$
$$s_i^s(t) = 0 \quad \forall t = \text{t}_{i,\text{endby}}, \dots, T \tag{8.7d}$$

Deferrable loads are considered to have some flexibility in their start time, but a defined load profile $l(\tau)\ \forall\ \tau = 1, \dots, L$ once started (e.g. appliances, manufacturing equipment). Following on the work in [167], we model the minimal starting time of the load as an arrival process $a(t)$, and the actual starting time as a departure process $d(t)$, where each of these variables takes the value 0 until the time of the request arrival/departure, at which point it takes the value 1. If the device can be started at most ζ time steps after the arrival request, we can formulate the constraints on our decision variable $d(t)$ as

$$\forall t = 1 \dots T :$$
$$0 \leq d_i(t-1) \leq d_i(t) \leq a_i(t) \tag{8.8a}$$
$$a_i(t-\zeta) \leq d(t) \tag{8.8b}$$
$$d_i(t) \in (0,1) \tag{8.8c}$$

Following [167] to formulate a matrix Φ which convolves the departure process $d(t)$ into a power consumption profile $s^d = \Phi d$, we can relax the binary constraint to allow scheduling to be expressed as a linear problem. We note from experience that this relaxation works well in practice, but offers no guarantees on exactness. Further work is required to assess the tightness of this relaxation.

Optimal Power Flow

We consider the problem of maximizing social welfare in the network over a day, which amounts to scheduling the controllable loads to minimize generation cost, while respecting network constraints. It is formulated as follows:

$$\min \sum_{t=1}^{T} \sum_{i=1}^{n} C_{i,t}(s_i^g(t)) \tag{8.9a}$$

$$\text{s.t. } (2.15a), (2.15b), (2.15c), (2.16), (8.1), (8.2), \quad t = 1, \ldots, T \tag{8.9b}$$

$$(8.5)_i, (8.6)_i, (8.7)_i, (8.8)_i, \quad i = 1, \ldots, n \tag{8.9c}$$

$$\text{over } s^g(t), s^d(t), s^s(t), s^b(t), \quad t = 1, \ldots, T$$

$$(P, Q, l, v)(t), \quad t = 1, \ldots, T$$

where constraints $(8.5)_i, (8.6)_i, (8.7)_i, (8.8)_i$, are specific to each node $i = 1, \ldots, n$, depending on the resources at that node.

In order to compensate the DER operators for their services and charge consumers for withdrawals, we want to compute nodal clearing prices, known as *distributed locational marginal prices* (DLMPs). The DLMP at a node represents the marginal cost to supply an additional unit of real power at that node. We denote the DLMP at node i as λ_i, and they can be found as the dual variables associated with the real power balance constraint (2.15a). As described in [29], the DLMP can be decomposed into contributions from energy, line losses, and voltage congestion.

Decomposition with ADMM

The Alternating Direction Method of Multipliers (ADMM) has gained popularity as a tool for decomposing difficult convex optimization problems into a set of simpler subproblems, coordinated through an aggregator step [182]. While convergence may be slow, the simplicity of the aggregator step and the guarantee of global optimality make the algorithm compelling for DER coordination. For examples of ADMM applications in various models of optimal dispatch problems, see [168, 170, 171, 183].

In the canonical ADMM problem, we consider a minimization problem with separable

objectives and constraints in vectors x and z:

$$\begin{aligned} \min_{x,z} \quad & f(x) + g(z) \\ \text{s. to:} \quad & x \in \mathcal{K}_x,\ z \in \mathcal{K}_z \\ & Ax + Bz = c \end{aligned}$$

We can form the augmented Lagrangian:

$$L_\rho(x, z, \xi) := f(x) + g(z) + \xi^\top (Ax + Bz - c) + \frac{\rho}{2}\|Ax + Bz - c\|^2$$

This then decomposes into the general form of ADMM:

$$x^{k+1} = \arg\min_{x \in \mathcal{K}_x} L_\rho(x, z^k, \xi^k) \tag{8.10a}$$

$$z^{k+1} = \arg\min_{z \in \mathcal{K}_z} L_\rho(x^{k+1}, z, \xi^k) \tag{8.10b}$$

$$\xi^{k+1} = \xi^k + \rho(Ax^{k+1} + Bz^{k+1} - c) \tag{8.10c}$$

When decomposing a problem into subproblems for solution with ADMM, it is useful to think of x and z in the above as *local* and *global* variables respectively. Local variables only pertain to their respective subproblems, whereas global variables couple subproblems together and must be agreed upon at the global optimum, reaching a distributed consensus among subproblems. An intuitive way to formulate this is to give each subproblem its own copy of any coupling variables, and then try and make these copies agree.

The economic dispatch problem of (8.9) can be reformulated in this way by forming an individual subproblem at each node, whose solutions are made to coincide at the global optimum through copied local coupling variables. Each subproblem has its own copy of the relevant global coupling variable, and consensus on their value is achieved among subproblems through the ADMM algorithm. The subproblem of node i takes the following form, where for clarity we have omitted the time index from each nodal variable.

$$\min \sum_{t=1}^{T} C_{i,t}(s_i^g(t)) \tag{8.11a}$$

$$\text{s.t. } p_i = P_i - \sum_{k \in \delta(i)} P_k + \mathrm{r}_i l_i, \quad t = 1, \ldots, T \tag{8.11b}$$

$$q_i = Q_i - \sum_{k \in \delta(i)} Q_k + \mathrm{x}_i l_i, \quad t = 1, \ldots, T \tag{8.11c}$$

$$v_i = v_{\pi(i)} + 2(\mathrm{r}_i P_i + \mathrm{x}_i Q_i) - (\mathrm{r}_i^2 + \mathrm{x}_i^2) l_i,\ t = 1, \ldots, T \tag{8.11d}$$

$$l_i \geq \frac{P_i^2 + Q_i^2}{v_i} \tag{8.11e}$$

$$(8.5)_i, (8.6)_i, (8.7)_i, (8.8)_i \tag{8.11f}$$

$$\text{over } s_i^g, s_i^b, s_i^d, s_i^s$$

$$(P_i, Q_i, l_i, v_i), (P_{\delta(i)}, Q_{\delta(i)}, v_{\pi(i)})$$

We first define a set of global variables $z := [P^\top, Q^\top, v^\top]^\top \in \mathbb{R}^{3n}$, a set of private local variables $x_i := [s_i^g, s_i^b, s_i^d, s_i^s, l_i]^\top$, and a set of coupling local variables:

$$\tilde{x}_i = [P_i^\top, Q_i^\top, v_i^\top, P_{\delta(i)}^\top, Q_{\delta(i)}^\top, v_{\pi(i)}^\top]^\top. \tag{8.12}$$

We see that each subproblem i is coupled to other subproblems through the coupling local variables $\tilde{x}_i$, each of which is a selection of the components of the global variable z. Using notation from [182], the mapping from local variable indices into the global variable index can be written as $g = \mathcal{G}(i,j)$, which means that local variable component $(\tilde{x}_i)_j$ corresponds to global variable component z_g. Achieving consensus between the local variables and the global variable means that

$$(\tilde{x}_i)_j = z_{\mathcal{G}(i,j)}, \forall i, j \tag{8.13}$$

We can equivalently define a selection matrix B_i, such that, $\tilde{z}_i = B_i z$, and at the optimum

$$\tilde{x}_i - B_i z = \tilde{x}_i - \tilde{z}_i = 0 \tag{8.14}$$

At each iteration k, each node i, receives $\tilde{z}_i^k$ from the central aggregator, and solves

$$\begin{aligned} \min\ & \sum_{t=1}^{T} C_{i,t}(s_{i,t}) + \xi_i^\top(\tilde{x}_i - \tilde{z}_i^k) + \frac{\rho}{2}\|\tilde{x}_i - \tilde{z}_i^k\|_2^2 \\ \text{s.t. } & (8.11) \\ \text{over } & x_i, \tilde{x}_i \end{aligned} \tag{8.15}$$

The node then sends its new $\tilde{x}_i^{k+1}$ to the central aggregator, who computes the following update for each individual global variable z_g^{k+1}

$$z_g^{k+1} := \frac{1}{k_g} \sum_{\mathcal{G}(i,j)=g} (\tilde{x}_i^{k+1})_j + \frac{(\xi_i^k)_j}{\rho} \tag{8.16}$$

where k_g is the number of local variable entries that correspond to global variable entry z_g. The update can be thought of as taking the average of all local copies of the global variable. The central aggregator then updates ξ_i as

$$\xi_i^{k+1} = \xi_i^k + \rho(\tilde{x}_i^{k+1} - \tilde{z}_i^{k+1}) \tag{8.17}$$

We define the stopping criteria using the following residuals

$$r_i^k = \tilde{x}_i^k - \tilde{z}_i^k,\ s^k = z^k - z^{k-1} \tag{8.18}$$

Defining $r^k := [r_1^k, \ldots, r_n^k]$, the algorithm is determined to have converged when the both the following conditions are met

$$\|r^k\|_2 \leq \epsilon_{pri},\ \|s^k\|_2 \leq \epsilon_{dual} \tag{8.19}$$

where ϵ_{pri}, ϵ_{dual} are suitably defined tolerances, and can be set using methods described in [182].

8.5 Blockchain and ADMM

We have formulated an optimal scheduling program for distributed energy resources through a decentralized algorithm. However, this only addresses part of the microgrid operation problem, and still has notable weaknesses:

- The aggregation step is not guaranteed against cyberattack or tampering by participants
- Individual DERs/consumers cannot verify that they are being paid/billed at fair prices
- Payments for actual generation/consumption will still be handled by a central utility

As an alternative, we propose to leverage the benefits of a blockchain architecture to create a fully peer-to-peer system which guarantees both operational feasibility and fair payments to all parties while taking full advantage of the decentralized structure of the problem.

repeat
  $\mathbf{P_i}$: *Private Optimization, compute locally*
    Gather private constraints
    Compute $\tilde{x}_i$ and send to smart contract S_1
  $\mathbf{S_1}$: *ADMM Aggregator, on blockchain*
    Update z
    if $\|r^k\|_2 \leq \epsilon_{pri},\ \|s^k\|_2 \leq \epsilon_{dual}$ **then**
      Compute final schedule and clearing prices
      Send schedule to S_2
    end
until $\|r^k\|_2 \leq \epsilon_{pri},\ \|s^k\|_2 \leq \epsilon_{dual}$

$\mathbf{M_i}$: *Each Smart Meter*
  Record energy consumption
  Send time-stamped & signed consumption to S_2
...time progresses
$\mathbf{S_2}$: *Billing contract, on blockchain*
  Compare schedule from S_1 with meter readings
  Compute penalties, payments, and charges
  Transfer payments between accounts

Algorithm 2: Computational elements in the microgrid control system. Function $\mathbf{P_i}$ is executed locally by each device participating in the market. The results are passed to the smart contract $\mathbf{S_1}$, which serves as publicly verifiable ADMM aggregation step. $\mathbf{P_i}$ and $\mathbf{S_1}$ iterate back and forth until ADMM converges, at which point the schedule is saved to the billing smart contract $\mathbf{S_2}$. Smart meters send trusted meter readings to $\mathbf{S_2}$, which computes payments and automatically transfers funds from consumers to generators.

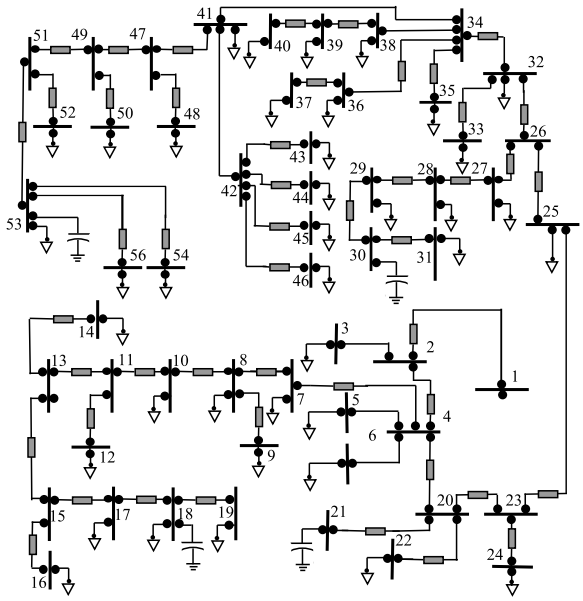


Figure 8.3: The 55-bus sample microgrid test feeder used in the simulation, with a microturbine placed at Bus 1 and DERs randomly distributed throughout the network.

As discussed in Section 8.2, blockchains provide a method for providing a transparent, trustless platform for data storage and computation. This makes a blockchain the perfect platform for conducting the aggregation step of ADMM, allowing all participants to audit the progress of the algorithm, the accuracy of the solution, and the veracity of their scheduled commitments. Further, ADMM is a natural fit for implementation on a blockchain, as it guarantees convergence yet has a *computationally cheap* aggregation step (minimizing the burden of verification).

Algorithm 2 provides an outline of the sequence of events in our proposed blockchain-based system. In it, we use the blockchain to (i) provide a fair computation of the ADMM aggregator update step from (8.16), (ii) store the resulting schedule, and (iii) compute payments and penalties for actual generation/consumption. Any participant can verify that the schedule maximizes social benefit while respecting network constraints, removing the possibility of monopolistic price manipulation.

This immutable record can also become the basis for reckoning payments if smart meters send consumption data to a billing contract S_2 which computes credits and debits for each node in the network and securely saves the updated account balance to the blockchain. Paired with a cryptocurrency as discussed in [156], this can form a complete payment system — removing the need for a utility or microgrid operator to handle scheduling and billing.

8.6 Implementation: Test network

We implement the proposed algorithm on a simulated SCE 55-bus test network shown in Fig. 8.3 with parameters described in [184]. Bus 1 is used as the reference bus, and is equipped with a large microturbine generator with quadratic cost function (this could also represent a connection to a utility grid). Each node has a deterministic load profile, created by adding a uniform random variable to the average uncontrollable load signal seen in Figure 8.4. We randomly place solar arrays at 60% of the buses, and assume a deterministic solar generation profile. We place deferrable loads at 70% of the buses, with earliest start times randomly distributed between hours 7:00-11:00; these represent appliances and industrial equipment. Shapable loads are also randomly placed at 70% of the buses, with net energy demand generated from a uniform random variable that is up to 10 times the peak power consumption of the uncontrollable loads; these are intended to represent electric vehicle loads. The time constraints are randomly generated such that the shapable loads begin self-scheduling as early as 10:00, and can continue to draw power as late as hour 24:00. Batteries are placed at each bus, with a power capacity of 50% of the peak controllable load at the bus, and with a 4 hour energy storage capacity.

We use a private Ethereum Homestead blockchain test network [185], and Python/CVXpy [186] to run the private optimization problems. Remote procedure calls through EthJsonRpc allow the Python scripts to communicate with the smart contracts.

Results

The ADMM algorithm converged in 204 iterations, using $\rho = 100$, $\epsilon_{pri} = 10^{-3}$, $\epsilon_{dual} = 0.1$, with each iteration taking at most 1.2s to compute. The optimal cost of the distributed solution was 0.4% larger than the welfare-maximizing centralized OPF solution.

The average power consumption across all nodes is shown in Fig. 8.4. The power consumption profiles of individual nodes primarily differ in the temporal constraints, size of shapable load, and presence or absence of solar. The deferrable and shapable loads self-schedule to coincide with solar generation, while the battery charges and discharges to smooth net load.

The impacts of network topology can be seen in the voltage of each bus, shown in Figure 8.5. Since there are no current flow constraints, at optimality the upper voltage limit at the generator bus (#1) becomes the binding constraint; the voltages at each of the other buses decrease with distance from the feeder due to line effects (the critical link between bus 4 and 20 can be clearly seen). General trends in voltage over the course of the day are visible, with a significant drop in hour 18 when the setting sun and peaking uncontrollable load leads to a spike in net load throughout the network. Upon closer inspection, the impacts of DER scheduling are also visible at some buses (e.g. 38, 48.49) as appliances and EVs switch on and off.

The distributed marginal prices are not shown here for brevity, but can be easily calculated from the net load supplied by the central generator (other resources are inframarginal).

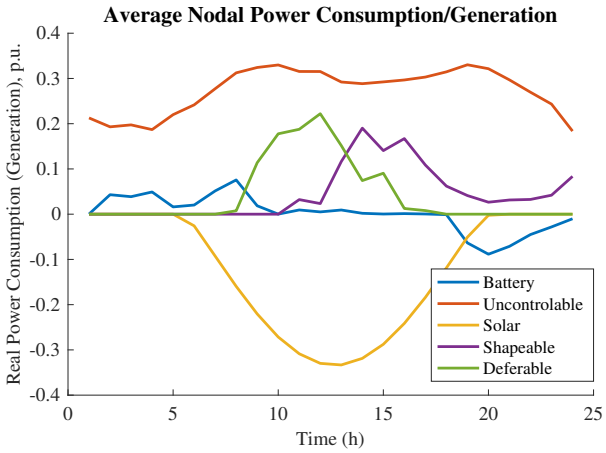


Figure 8.4: Schedule of commitments generated by the ADMM algorithm and stored to the smart contract. Positive values of power indicate real power consumption, and negative values indicate generation/injections.

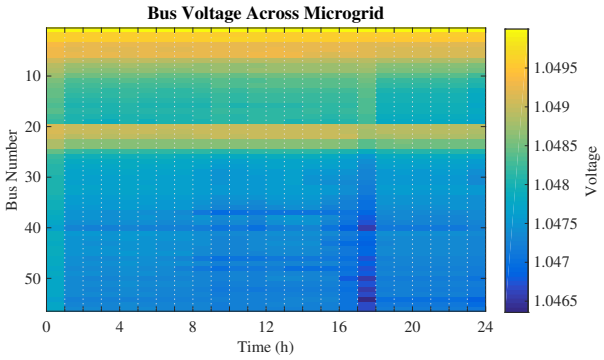


Figure 8.5: Voltage magnitude at each of the buses on the test network, for each hour in the simulation. Voltages vary based on local injections, and variations in time can be seen due to the impacts of local DER scheduling.

We found very little variation in DLMPs between buses (variance of <1% of hourly DLMP), reflecting a lack of binding line constraints on this small network.

8.7 Limitations

The communication overhead required for ADMM and the verification delay required for a blockchain may limit this approach to use for day-ahead scheduling, while secondary and tertiary control is served by conventional approaches. We envision a blockchain-based economic scheduling layer (described here) operating in tandem with a real-time control layer operating at much faster time scales to address real-time imbalances.

While here we envision each building's smart meter acting as a computational node in the blockchain network, alternate configurations may use devices, feeders, or substations as nodes. In each of these implementations, we assume that the network topology is fully known by all parties, and have not considered changes in line impedances (e.g. due to temperature changes) or in topology (e.g. due to outages).

This chapter relies on previous work for proofs of the security, transparency, and robustness of blockchain-based systems (see e.g. [159, 187]), and future work will explore the specific security concerns of the microgrid system outlined here.

8.8 Conclusions

We have shown how decentralized consensus techniques and blockchains can be used both to coordinate the scheduling of distributed energy resources on a microgrid, and to guarantee fair payments without requiring a utility or centralized microgrid aggregator. By using ADMM, we decompose our problem into a structure that naturally lends itself to a blockchain implementation, and show how blockchains and smart contracts can provide a natural solution for the trust, security, reliability, and immutability requirements of microgrid operation. We show the results are equivalent to a welfare-maximizing centralized dispatch with perfect insight into device constraints, yet avoid the risk of monopoly price manipulation and privacy concerns.

The proposed architecture can be improved with contributions from active areas of control research: addressing stochastic/uncertain data through model predictive control and robust optimization, examining resilience to network interruptions, utilizing fully distributed ADMM between nodes to reduce communication overhead, and developing fault detection algorithms to identify fraud and changes in system architecture.

While this work examines the integration of blockchain with distributed optimization of energy systems, we expect that many other applications are possible, both within the energy sector and in other engineering realms. Blockchain's distributed consensus mechanism has proven itself in the finance world by guaranteeing robust, trustless, and transparent execution; we highlight similar benefits for controlling physical devices. We see blockchains and smart contracts as a key technology that enables distributed optimization amongst non-trusting entities, at all scales of operation.

Chapter 9

APEX: The Automatic Power Exchange

In the previous chapter a blockchain-based implementation of ADMM was used to enable secure and trustless decentralized coordination among DERs in a distribution network. The coordination represents a market of a sort, but we assume that participants are self-scheduling against a utility price signal or local generation costs, rather than bidding in for their demand, excess generation or any potential flexibility. In this chapter we address the design of a marketplace for DERs and flexible resources that includes explicit products for different types of flexibility, and ensures that system constraints are respected. We also provide a variety of solution methods for matching orders and clearing the market in both a continuous and discrete manner. The implementation of the market is platform agnostic, and could be run using traditional cloud infrastructure or a blockchain based infrastructure.

This chapter is the result of joint work with Dr. Junjie Qin, Prof. Ram Rajagopal, Prof. Kameshwar Poolla, and Prof. Pravin Varaiya, and a selection of these results were first published in [138].

9.1 Introduction

Background

Wholesale electricity markets have long enabled efficient trading of bulk energy and services at the transmission scale. But there are many significant resources and assets connected to the distribution network that have not been fully monetized [188]. Novel distribution system markets that match the local intermittent supply with flexible demand can potentially greatly increase the utilization of these assets.

However, designing such markets is challenging for a number of reasons. First, many distributed energy resources are *variable* resulting in intermittent and uncertain power generation that introduces both *quantity risk* (e.g., not enough supply to meet demand) and

price risk (e.g., consumers may be charged highly volatile prices) into the market. Managing such risks require a sophisticated distribution system operator (DSO) solving stochastic dispatch programs, or *forward markets* in which the market participants can trade to hedge against uncertainty [189–191]. Second, although demand flexibility is ubiquitous, unlocking it usually requires upfront capital investments from users (e.g. for installing smart appliances and/or building energy management systems [192, 193]) that need to be justified by a clear expectation of (financial) benefits. Existing proposals around real time pricing could potentially provide such an expectation, but it may be blurred by difficulties in forecasting prices driven by both exogenous uncertainty (e.g. renewable generation) and endogenous uncertainty (e.g. other market participants' behaviors). Explicit *flexibility markets* that schedule flexible demand on behalf of the users could significantly reduce the burden on users and provide clear incentives for users to engage, reveal and trade their flexibility. Finally, as the market outcomes induce physical power flow on the distribution network, physical network constraints need to be managed to ensure the reliability of the distribution network.

Contributions

In this chapter, we propose a scalable market platform, referred to as *Automatic Power Exchange* (APEX), that enables monetization of these underutilized distribution system assets. The APEX platform allows distribution system participants to trade energy and services. It incorporates variable distributed energy resources through an *open-gate* forward market design. That is, for each delivery period, users can submit orders anytime inside of a trading time window, which if possible will be cleared as submitted. Effectively, this introduces a continuum of forward markets, where users can hedge against uncertainty through adjustment orders based on the most recent information. APEX also arranges an explicit flexibility market. Distribution system participants can submit the *availability* information of their flexible loads, and APEX will schedule these flexible loads on behalf of the participants. APEX will respect distribution network constraints on the flow of electricity either by directly managing the distribution network or by following a coordinated trading protocol [194, 195] operated by a third-party distribution system operator.

This chapter makes the following contributions:

(i) We propose a novel design for a distribution system market that addresses uncertainty from DERs using an open-gate forward market design and solicits demand flexibility by efficient in-market flexible demand scheduling, while managing distribution network constraints.

(ii) We study the non-convex problem for scheduling *non-preemptive* flexible loads and propose provably efficient algorithms to ensure the scalability of the APEX platform.

(iii) We analyze the properties of a natural marginal pricing mechanism in the APEX context, establishing that it is revenue adequate but may lead to inadmissible prices for flexible orders. We then suggest a simple alternative that is guaranteed to produce

admissible prices for users and adequate revenue for APEX when used together with the proposed scheduling algorithms.

(iv) We introduce a novel extension of the traditional continuous double auction to the case of coupled multiple periods and constraints, which enables continuous trading of orders in the context of APEX. The advantage of such an approach is that can be implemented purely using computational logic, and does not require any explicit optimization.

Literature

Forward markets have been implemented in many wholesale electricity markets. It is known that forward markets help manage uncertainty and incorporate generation technologies with different lead times [189]. When a sequence of forward markets are available, risk sensitive consumers (suppliers) may limit their risks by procuring from (offering to) multiple forward markets [196, 197]. Although many of these studies have focused on the wholesale market, empirical studies demonstrate that for smaller consumers having the option of participating in forward markets helps them to hedge their bill volatility [190]. Open-gate forward markets, compared to fixed-time forward markets such as day-ahead and hour-ahead markets, are not common for electricity. Yet, they are widely implemented in financial industries [198].

The utilization of flexible energy resources in distribution networks has been studied in a number of papers. Existing studies usually exploit the flexibility in restrictive settings where only one attribute of the flexible resources is allowed to vary. Such treatments lead to interesting control and pricing problems for electricity services that are differentiated according to that particular attribute [199–201]. It is usually assumed in these papers that these services are organized and provided by aggregators instead of a flexibility market that matches flexible resources with other resources. Furthermore, in our flexibility market, the flexible orders are allowed to simultaneously have many different attributes, thus bridging a gap between prior studies and practical implementations.

As a whole package, APEX provides a novel design for a distribution system electricity market with significant DER penetration. Existing alternative proposals for distribution system markets can be roughly categorized into *centralized* and *transactive*. Centralized designs seek to modify or augment existing utility companies' rate structures to align DERs' power consumption/production with wholesale price signals. Notable examples include real time pricing (RTP) and its variants (cf. [202–205] and references therein). The benefits of centralized design include tight management of the distribution network through utility companies and the fact that they are relatively easy to implement given the current institutional structure of retail electricity markets. Such designs, nevertheless, are inflexible as it is difficult to incorporate differentiated electricity services. In contrast, transactive designs rely on bilateral or multilateral transactions (or contracts) among individual participants (cf. [206, 207] and references therein). As the terms and conditions in the contracts can be tailored according to individual needs, it is very easy to incorporate various flexible resources in transactive market designs. However, these designs represent a large structural

departure from today's utility-managed distribution markets, requiring coordination to ensure reliability of the distribution network, [194] and may impose significant *search costs* on the participants. Compared to these two classes of distinct designs, we view APEX as a middle ground where flexible resources are incorporated through an explicit flexibility market with an expressive alphabet of standard commodities, and distribution network constraints are tightly managed (possibly by a coordinated trading protocol). Participants' search costs are also largely reduced in APEX.

Many of our technical results extend the growing literature on scheduling non-preemptive deferrable loads [208–210]. Most of these prior studies focus on an aggregator setting while we consider scheduling in a two-sided market. Furthermore, the fluid relaxation based scheduling algorithm that we propose is novel and well-suited for the region where there is a large number of flexible loads. As the scheduling problem is non-convex, the problem of pricing these flexible loads with distribution network constraints is challenging and understudied in the literature. Our results on understanding the properties of marginal pricing with sub-optimal scheduling algorithms may pave the way for future development on this topic.

Organization

The rest of the chapter is organized as follows. Section 9.2 describes the APEX market platform and states the order matching problem in APEX. Section 9.3 proposes efficient algorithms for the combinatorial optimization of scheduling non-preemptive flexible orders and establishes their performance guarantees. The associated pricing problem for APEX order matching is then considered in Section 9.3. Section 9.5 concludes the chapter.

9.2 APEX Platform

In a nutshell, APEX receives orders (Section 9.2) from users (Section 9.2), forms and maintains an orderbook (Section 9.2), and solves an order matching problem (Section 9.2) that fulfills standing orders in the orderbook by matching supply with demand respecting distribution network constraints (Section 9.2). The schematic of the trading process in APEX is demonstrated in Figure 9.1.

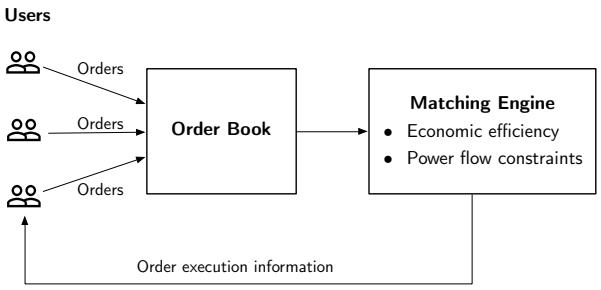


Figure 9.1: Trading process in APEX.

Trading in APEX happens within the following temporal structure. Time is slotted into time intervals of Δt length (e.g., Δt can be 5 minutes). Power delivery in each of these time intervals is traded. We thus work with a discrete time model, using $t \in \mathbb{Z}$ to denote each time period. At any time t[1], users can submit orders regarding power delivery in any future time intervals belonging to a *trading time window* $\mathcal{T}_t$ that includes T time intervals. The trading time window may start with the next time interval $t+1$ and ends with 24 hours after the next time interval, i.e., in this case T is 24 hours/Δt. As a result, orders regarding the power delivery in any time interval t' can be submitted at any t such that $t' \in \mathcal{T}_t$. This implements an open-gate forward market. Figure 9.2 gives an example of the open-gate forward markets and trading windows for two delivery intervals.

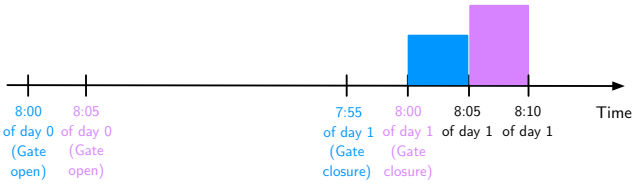


Figure 9.2: Example of APEX trading windows

We proceed to introduce individual components of the APEX platform.

Distribution network model

We adopt the *Simplified DistFlow* model, described in Section 2.2 and use notation from that Section. We additionally adopt the notation that vectors are written in lower case bold font, to differentiate from scalars. We reiterate the feasible injection region of the distribution network:

$$\mathcal{P} := \{\mathbf{p} \in \mathbb{R}^N : \mathbf{1}^\top \mathbf{p} = 0,\ \underline{\mathbf{v}} \leq \hat{\mathbf{v}} + R\mathbf{p} \leq \overline{\mathbf{v}},\ H\mathbf{p} \leq \boldsymbol{c}\},$$

As is common in practice, we assume $\underline{\mathrm{v}} \leq \hat{\mathbf{v}} \leq \overline{\mathrm{v}}$ and $\mathbf{0} \leq \mathbf{c}$, and therefore $\mathbf{0} \in \mathcal{P}$.

User model

We denote the set of users by $\mathcal{I} := \{1, \ldots, I\}$. Let the bus that user $i \in \mathcal{I}$ resides be denoted by n_i and the set of users located at bus n be denoted by $\mathcal{I}_n$. Each user may model an individual home, a commercial building, or an aggregation of many buildings coordinated by an aggregator or as a micro-grid. In this paper, APEX is agnostic to the level of aggregation inside of each user.

[1]In general this will in fact be within the time window $[t, t+\Delta t]$, where the window may be curtailed by some necessary computation time for the market ahead of delivery

Order formats

At any instance in time, referred to as $t = 0$, a participant located at bus n (i.e. $i \in \mathcal{I}_n$) can submit buy or sell orders for a trading window of time periods $t \in \mathcal{T} := \mathcal{T}_t = \{1, \ldots, T\}$. Considering typical supply-side and demand-side characteristics, we allow buy and sell orders in the formats specified as follows.

Definition 10 (Simple sell order)**.** *A simple sell order from participant $i \in \mathcal{I}_n$ is a tuple*

$$s = (n,\ t,\ \overline{q},\ \underline{\pi}),$$

where n is the bus index, $t \in \mathcal{T}$ is time of electricity delivery, $\overline{q} \in \mathbb{R}_+$ is the maximum amount of electricity to be sold and $\underline{\pi} \in \mathbb{R}$ is the minimum acceptable price of electricity for the sell order.

Symmetrically, we have simple buy order defined.

Definition 11 (Simple buy order)**.** *A simple buy order from participant $i \in \mathcal{I}_n$ is a tuple*

$$b = (n,\ t,\ \overline{q},\ \overline{\pi}),$$

where n is the bus index, $t \in \mathcal{T}$ is time of electricity delivery, $\overline{q} \in \mathbb{R}_+$ is the maximum amount of electricity to be bought and $\overline{\pi} \in \mathbb{R}$ is the maximum acceptable price of electricity for the buy order.

Simple orders may not be expressive enough to incorporate certain flexible loads such as *non-preemptive* deferrable and shapeable loads. Deferrable loads consume pre-defined load shapes but are indifferent to the time at which the loads are served as long as they are served in a certain time window. Shapeable loads consume a pre-defined quantity of energy over a period of time, but are indifferent, within pre-defined limits, to the rate at which this energy is delivered. For simplicity, for the remainder of the chapter we will focus on deferrable loads only. The results and definitions in this chapter are straightforwardly extended to include shapeable loads.

This motivates us to incorporate a flexible buy order as follows.

Definition 12 (Flexible buy order)**.** *A flexible buy order from participant $i \in \mathcal{I}_n$ is a tuple*

$$f = (n,\ t^{\mathrm{ES}},\ t^{\mathrm{LC}},\ \tau^{\mathrm{D}},\ \widehat{\mathbf{q}},\ \overline{\pi}),$$

where n is the bus index, $t^{\mathrm{ES}} \in \mathcal{T}$, $t^{\mathrm{LC}} \in \mathcal{T}$ and $\tau^{\mathrm{D}} \in \mathcal{T}$ denote the earliest starting time, latest completion time and duration of the flexible load, respectively, $\widehat{\mathbf{q}} \in \mathbb{R}_+^{\tau^{\mathrm{D}}}$ is the load shape to be consumed, and $\overline{\pi} \in \mathbb{R}$ is the maximum acceptable price of electricity[2] for the buy order.[3]

Figure 9.3 depicts the parameters used to define a flexible buy order.

[2]The maximum amount of payment associated with the order is $\overline{\pi}\mathbf{1}^\top\widehat{\mathbf{q}}$.

[3]For a flexible shapeable order, τ^{D} represents the duration over which power is delivered to the load, and rather than a load profile $\widehat{\mathbf{q}} \in \mathbb{R}_+^{\tau^{\mathrm{D}}}$ we would require a minimum and maximum power profile over the duration of the interval, denoted $\underline{\mathbf{q}} \in \mathbb{R}^{\tau^{\mathrm{D}}}$ and $\overline{\mathbf{q}} \in \mathbb{R}^{\tau^{\mathrm{D}}}$ respectively.

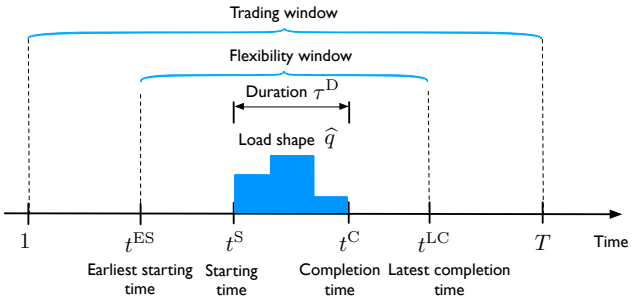


Figure 9.3: Parameters of a flexible buy order

Orderbook

In practice, buy and sell orders arrive continuously in time. There are three processes which can be used to manage the clearing of these orders. In the first process, whenever a new order arrives, APEX runs an efficient matching algorithm with all the standing orders and the newly arrived order, which determines the fulfillment of a subset of these orders. In the second process, the matching algorithm is run at short periodic intervals, incorporating all orders recieved since the matching algorithm was previously run. In the third process, the matching algorithm is run exactly once per market interval, clearing all orders in one go. Depending on the rate of arrival of orders, and the number of outstanding unfulfilled orders at any one time, each process may be more or less appropriate for a given instance of the market. Additionally the choice of process has an impact on the economic efficiency or global optimality of the set of matched orders. One can imagine clearing an order as soon as it arrives, however, had we waited another order may have arrived, which if matched with the original order would have generated a higher surplus or utility. This has relevance to the choice of utility function for the platform, discussed in (9.3), (9.4). It also illustrates the difficulty of assessing the economic efficiency of a multi-period open-gate market such as this.[4]

All unfulfilled orders, which cannot be matched due to (i) lack of supply for a buy order or lack of demand for a sell order, (ii) lack of a mutually acceptable price, and (iii) network constraints, remain standing and are recorded into an *orderbook*.

Definition 13 (Orderbook)**.** *The orderbook at time t is defined to be the triple $(\mathcal{B}, \mathcal{S}, \mathcal{F})$, where $\mathcal{B}$ is the collection of standing simple buy orders, $\mathcal{S}$ is the collection of standing simple sell orders, and $\mathcal{F}$ is the collection of standing flexible buy orders.*

In addition to the information supplied in the order formats described previously, an order may have an expiry time. If the order remains unfulfilled at this time it is removed from the orderbook. Additionally if an order remains unfulfilled after its specified delivery

[4] An appropriate benchmark may be solving the market with perfect foresight in terms of orders, and finding what the optimal matching would have been. However given the dynamics of the market it is unclear how meaningful this is, as the clearing of one order may impact whether a second order is made or not.

time has passed, it is also removed from the orderbook. This can be thought of as a default expiry time.

Furthermore, when making an order, there are a number of order types which may be specified. We describe these below.

Order Types

We define three types of orders:

Definition 14 (Limit Order)**.** *A* limit order *represents an order that may be partially fulfilled in the order matching process. These orders remain in orderbook with the desirable quantities ($\overline{q}_b$ or $\overline{q}_s$) updated by subtracting the fulfilled amounts.*

Definition 15 (Fill-or-Kill Order)**.** *A* fill-or-killl *order must be fulfilled in its entirety, or not at all.*

Definition 16 (Market Order)**.** *A* market order *is an order that is willing to be fulfilled at any price, i.e. the market price. It specifies no order price when the order is made. A market order can be thought of as a limit order with a maximum (minimum) acceptable price of the market cap (floor), since, assuming there is sufficient quantity on the other side of the market, a market order will always be fulfilled in its entirety.*

For simplicity, from hereon we assume that all orders are limit orders, which may be partially fufilled. Market orders are included as limit orders, using market price caps as the order prices. Including fill-or-kill orders requires more complex combinatorial logic which will be the subject of future work.

Order matching problem

To fulfill the orders in the orderbook $(\mathcal{B}, \mathcal{S}, \mathcal{F})$, the order matching process aims to determine an *admissible schedule* and an *admissible price* for each (partially) fulfilled order in the orderbook.

We denote the set of power profiles representing all admissible fulfillment of simple buy order b and simple sell order s by $\mathcal{Q}_b \subset \mathbb{R}^T$ and $\mathcal{Q}_s \subset \mathbb{R}^T$, respectively. Similarly, the set of power profiles representing all admissible fulfillment (including $q = 0$) of the flexible load f is denoted by $\mathcal{Q}_f \subset \mathbb{R}^T$.

Definition 17 (Admissible schedule)**.** *A power production or consumption schedule $\mathbf{q} \in \mathbb{R}^T$ over the trading window $\mathcal{T}$ is deemed admissible, if the following conditions hold.*

- *For simple sell order $s = (n,\ t,\ \overline{q},\ \underline{\pi})$:*

$$\mathbf{q} = \mathbf{q}_s \in \mathcal{Q}_s := \{q\mathbf{1}_t \in \mathbb{R}^T : 0 \leq q \leq \overline{q}\},$$

where $\mathbf{1}_t \in \mathbb{R}^T$ is the elementary vector with a 1 at t-th element and 0's elsewhere.

- *For simple buy order* $b = (n,\ t,\ \overline{q},\ \overline{\pi})$*:*

$$\mathbf{q} = \mathbf{q}_b \in \mathcal{Q}_b := \{q\mathbf{1}_t \in \mathbb{R}^T : 0 \le q \le \overline{q}\}.$$

- *For flexible buy order* $f = (n,\ t^{\mathrm{ES}},\ t^{\mathrm{LC}},\ \tau^{\mathrm{D}},\ \widehat{\mathbf{q}},\ \overline{\pi})$*:*

$$\mathbf{q} = \mathbf{q}_f \in \mathcal{Q}_f,$$

where $\mathcal{Q}_f$ *is defined as the set of power profiles* $\mathbf{q} \in \mathbb{R}^T$ *such that there exists a starting time* t^{S} *so* $(\mathbf{q}, t^{\mathrm{S}})$ *satisfies*

$$t^{\mathrm{S}} \in \{t^{\mathrm{ES}}, \dots, t^{\mathrm{LC}} - \tau^{\mathrm{D}} + 1\}, \tag{9.1a}$$

$$q(t) = \begin{cases} \widehat{q}(t - t^{\mathrm{S}} + 1), & \text{if } t^{\mathrm{S}} \le t < t^{\mathrm{S}} + \tau^{\mathrm{D}}, \\ 0, & \text{otherwise.} \end{cases} \tag{9.1b}$$

If the order is not to be scheduled, we denote $t^{\mathrm{S}} = 0$ *and so* $\mathbf{0} \in \mathcal{Q}_f$ *by definition.*

Definition 18 (Admissible price)**.** *For a (partially) scheduled order (i.e.,* $\mathbf{q} \neq 0$*), a clearing price* $\pi \in \mathbb{R}$ *is deemed admissible, if the following conditions hold.*

- *For simple sell order* $s = (n,\ t,\ \overline{q},\ \underline{\pi})$*:* $\pi \ge \underline{\pi}$*.*
- *For simple buy order* $b = (n,\ t,\ \overline{q},\ \overline{\pi})$*:* $\pi \le \overline{\pi}$*.*
- *For flexible buy order* $f = (n,\ t^{\mathrm{ES}},\ t^{\mathrm{LC}},\ \tau^{\mathrm{D}},\ \widehat{\mathbf{q}},\ \overline{\pi})$*:* $\pi \le \overline{\pi}$*.*

In the order matching problem, we try to identify admissible fulfillment of all orders $(\{\mathbf{q}_s\}, \{\mathbf{q}_b\}, \{\mathbf{q}_f\}) := (\{\mathbf{q}_s\}_{s\in\mathcal{S}}, \{\mathbf{q}_b\}_{b\in\mathcal{B}}, \{\mathbf{q}_f\}_{f\in\mathcal{F}})$ in a way that maximizes certain criteria designed by the operator of APEX, denoted by $U(\{\mathbf{q}_s\}, \{\mathbf{q}_b\}, \{\mathbf{q}_f\})$, while respecting the distribution network constraints[5]. This can be written as the following optimization problem

$$\max_{\{\mathbf{q}_s\},\{\mathbf{q}_b\},\{\mathbf{q}_f\}} \quad U\Big(\{\mathbf{q}_s\}, \{\mathbf{q}_b\}, \{\mathbf{q}_f\}\Big) \tag{9.2a}$$

$$\text{s.t.} \quad \mathbf{q}_s \in \mathcal{Q}_s, \quad s \in \mathcal{S}, \tag{9.2b}$$

$$\mathbf{q}_b \in \mathcal{Q}_b, \quad b \in \mathcal{B}, \tag{9.2c}$$

$$\mathbf{q}_f \in \mathcal{Q}_f, \quad f \in \mathcal{F}, \tag{9.2d}$$

$$\mathbf{p}_n = \sum_{s\in\mathcal{S}_n} \mathbf{q}_s - \sum_{b\in\mathcal{B}_n} \mathbf{q}_b - \sum_{f\in\mathcal{F}_n} \mathbf{q}_f, n \in \mathcal{N}, \tag{9.2e}$$

$$\mathbf{p}(t) \in \mathcal{P}_{\mathrm{D}}, \quad t \in \mathcal{T}. \tag{9.2f}$$

[5]In view of the equivalence between the setup where APEX directly manages distribution network constraints, and the setup where network constraints are managed by a minDSO and APEX communicates with the minDSO through a coordinated trading process [194, 195], we assume the first setup in this chapter.

where (9.2e) is the local power balance equation at each node n, with $\mathcal{S}_n$, $\mathcal{B}_n$ and $\mathcal{F}_n$ denoting the set of simple sell orders, simple buy order and flexible buy order submitted by users at bus n, respectively. We will refer to these collectively as nodal orderbooks.

APEX may optimize different criteria in the order matching problem depending on its real-world implementation (e.g., whether it is implemented by a for-profit platform company or by a regulated utility company). Here we list two possible objective function to optimize.

- Total surplus:

$$\begin{aligned} &U(\{\mathbf{q}_s\},\{\mathbf{q}_b\},\{\mathbf{q}_f\}) \\ &= \sum_{b\in\mathcal{B}} \overline{\pi}_b \mathbf{1}^\top \mathbf{q}_b + \sum_{f\in\mathcal{F}} \overline{\pi}_f \mathbf{1}^\top \mathbf{q}_f - \sum_{s\in\mathcal{S}} \underline{\pi}_s \mathbf{1}^\top \mathbf{q}_s. \end{aligned} \tag{9.3}$$

- Total volume:

$$U(\{\mathbf{q}_s\},\{\mathbf{q}_b\},\{\mathbf{q}_f\}) = \sum_{b\in\mathcal{B}} \mathbf{1}^\top \mathbf{q}_b + \sum_{f\in\mathcal{F}} \mathbf{1}^\top \mathbf{q}_f. \tag{9.4}$$

We note that with criteria (9.3) or (9.4), the order matching problem has a linear objective function. Meanwhile, constraints (9.2b), (9.2c), (9.2e) and (9.2f) are all linear inequality or equality constraints. However, (9.2) is challenging to solve due to non-convex constraint (9.2d). In fact, the combinatorial nature is evident if we return to the characterization of $\mathcal{Q}_f$ using the starting times (9.1).

While solving (9.2) gives an admissible schedule for each order in the orderbook, it does not provide admissible prices. Thus the second part of the order matching problem is to identify an admissible price for each fulfilled order. Given the non-convex nature of (9.2) and the fact that in some cases can only obtain approximate solutions of (9.2) in practice, the pricing problem for APEX is challenging. In particular, the natural application of the *marginal pricing* idea to this context requires a re-examination because its properties established for convex settings may no longer hold (see Section 9.3).

We now present two methodologies for solving (9.2), one using optimization, the other using logic. In Section 9.3 we describe the optimization approach, introducing two algorithms that solve (9.2) approximately, and in Section 9.3 we address pricing under these two algorithms. In Section 9.4 we describe the logic-based approach, and introduce an algorithm that solves the problem exactly in a continuous manner without optimization, and how pricing is addressed under such an algorithm. This latter approach more closely resembles the operation of a stock market rather than a centralized ISO clearning mechanism.

9.3 Optimization-Based Solution Methodology

Scheduling Algorithms

Fixing the schedule of flexible buy orders, the order matching problem is a linear program thus efficiently solvable. We therefore denote

$$
\begin{aligned}
&J(\{\mathbf{q}_f\}) := \\
&\quad \max_{\{\mathbf{q}_s\},\{\mathbf{q}_b\}} \quad U(\{\mathbf{q}_s\},\{\mathbf{q}_b\},\{\mathbf{q}_f\}) && \text{(9.5a)}\\
&\quad \text{s.t.} \quad \mathbf{q}_s \in \mathcal{Q}_s, \quad s \in \mathcal{S}, && \text{(9.5b)}\\
&\qquad\quad \mathbf{q}_b \in \mathcal{Q}_b, \quad b \in \mathcal{B}, && \text{(9.5c)}\\
&\qquad\quad \mathbf{p}_n = \sum_{s\in\mathcal{S}_n} \mathbf{q}_s - \sum_{b\in\mathcal{B}_n} \mathbf{q}_b - \sum_{f\in\mathcal{F}_n} \mathbf{q}_f, n \in \mathcal{N}, && \text{(9.5d)}\\
&\qquad\quad \mathbf{p}(t) \in \mathcal{P}_{\mathrm{D}}, \quad t \in \mathcal{T}. && \text{(9.5e)}
\end{aligned}
$$

and focus on the optimization for scheduling flexible orders:

$$
\begin{aligned}
&\max_{\{\mathbf{q}_f\}} \quad J(\{\mathbf{q}_f\}) && \text{(9.6a)}\\
&\text{s.t.} \quad \mathbf{q}_f \in \mathcal{Q}_f, \quad f \in \mathcal{F}. && \text{(9.6b)}
\end{aligned}
$$

We proceed to describe a way to solve this problem based on a greedy heuristic.

Greedy scheduling

We start by reformulating (9.6) to a set function maximization. Let $\mathbf{t}^{\mathrm{S}} \in \mathcal{T}^{|\mathcal{F}|}$ be the vector of starting times of all flexible buy orders which uniquely determines the schedule of all flexible buy orders $\{\mathbf{q}_f\}$. Denote the value of (9.6) with some fixed $\mathbf{t}^{\mathrm{S}}$ by $V(\mathbf{t}^{\mathrm{S}})$, i.e.,

$$
V(\mathbf{t}^{\mathrm{S}}) = \begin{cases} J(\{\mathbf{q}_f(t_f^{\mathrm{S}})\}), & \text{if } \mathbf{q}_f(t_f^{\mathrm{S}}) \in \mathcal{Q}_f, f \in \mathcal{F}, \\ -\infty, & \text{otherwise,} \end{cases}
$$

where $\mathbf{q}_f(t_f^{\mathrm{S}})$ is the power consumption profile induced by starting time t_f^{S}. Consider the pairs of flexible buy orders and their starting times in the set

$$
\Omega = \{(f, t_f^{\mathrm{S}}) : f \in \mathcal{F}, t_f^{\mathrm{S}} \in \mathcal{T} \cup \{0\}\}. \tag{9.7}
$$

Notice that any feasible scheduling can be represented by a subset of Ω; conversely, subsets of Ω that select no more than one starting time for each flexible buy order f can represent all feasible scheduling decisions. Define set-to-matrix mapping $\mathbb{I} : 2^{\Omega} \mapsto \mathbb{R}^{|\mathcal{F}|\times T}$

$$
[\mathbb{I}(X)]_{f,t} = \begin{cases} 1, & \text{if } (f,t) \in X, \\ 0, & \text{otherwise,} \end{cases} \tag{9.8}
$$

and normalized objective function

$$g(X) = V(\mathbb{I}(X)\boldsymbol{\delta}) - V(\mathbb{I}(\emptyset)\boldsymbol{\delta}), \tag{9.9}$$

where $\boldsymbol{\delta} \in \mathbb{R}^{T\times 1}$ is such that $\delta_t = t$, and the matrix vector product $\mathbb{I}(X)\boldsymbol{\delta}$ converts a subset $X \subset \Omega$ into the corresponding starting time vector t^{S}. With these definitions, problem (9.6) is equivalent to the following *subset selection problem*:

$$\max_{X\subset\Omega} \quad g(X), \tag{9.10a}$$

$$\text{s.t.} \quad \sum_{t=1}^{T} [\mathbb{I}(X)]_{f,t} \leq 1, \quad f \in \mathcal{F}, \tag{9.10b}$$

where the constraint ensures that X selects at most one starting time for each flexible buy order. The problem in general is NP hard as the number of subsets is $2^{|\Omega|}$.

The greedy approach for solving (9.10) amounts to scheduling flexible orders one-by-one according to the incremental benefit that scheduling a new order brings as measured by function $g(X)$. Algorithm 3 lists the steps for greedy scheduling. After this algorithm terminates, admissible schedules for flexible buy orders are obtained. We can then obtain admissible schedules for simple orders by solving (9.5) with the resulting $\{\mathbf{q}_f\}$ from the greedy algorithm.

$\mathcal{F}^{\mathrm{S}} = \emptyset$, $\mathcal{F}^{\mathrm{TBS}} = \mathcal{F}$;
$X \leftarrow \emptyset$;
while $\mathcal{F}^{\mathrm{TBS}} \neq \emptyset$ **do**
  $\mathcal{C} \leftarrow \{(f, t^{\mathrm{S}}) : f \in \mathcal{F}^{\mathrm{TBS}}, t^{\mathrm{S}} \in \{t_f^{\mathrm{ES}}, \ldots, t_f^{\mathrm{LC}} + 1 - \tau_f^{\mathrm{D}}\}\}$;
  $(\widehat{f}, \widehat{t^{\mathrm{S}}}) \leftarrow \arg\max_{(f,t^{\mathrm{S}})\in\mathcal{C}} g(X \cup \{(f, t^{\mathrm{S}})\}) - g(X)$;
  if $g(X \cup \{(\widehat{f}, \widehat{t^{\mathrm{S}}})\}) > g(X)$ **then**
    $X \leftarrow X \cup \{(\widehat{f}, \widehat{t^{\mathrm{S}}})\}$;
    $\mathcal{F}^{\mathrm{S}} \leftarrow \mathcal{F}^{\mathrm{S}} \cup \{\widehat{f}\}$, $\mathcal{F}^{\mathrm{TBS}} \leftarrow \mathcal{F}^{\mathrm{TBS}} \backslash \{\widehat{f}\}$
  else
    break;
  end
end

Algorithm 3: Greedy scheduling

Fluid relaxation

Although the complexity of greedy scheduling is polynomial in the number of flexible orders, it becomes relatively slow when there are a large number of orders because it needs to loop over the remaining orders and their feasible starting times in each step. In this section, we consider an alternative scheduling algorithm based on relaxing the non-convex constraint

$\mathbf{q}_f \in \mathcal{Q}_f$. In particular, it first relaxes the requirement that each load shape needs to follow the exact load shape $\widehat{\mathbf{q}}_f$ submitted by the user, solves a convex optimization to determine the schedule $\widetilde{\mathbf{q}}_f$, and then "projects" the schedule $\widetilde{\mathbf{q}}_f$ to a feasible schedule $\Pi_{\mathcal{Q}_f}(\widetilde{\mathbf{q}}_f) \in \mathcal{Q}_f$. As the key step in this algorithm is to remove the load shape requirement by allowing any profile to be scheduled in the time window $\{t_f^{\mathrm{ES}}, \ldots, t_f^{\mathrm{LC}}\}$, we refer to this algorithm as *fluid relaxation*. Similar ideas have been used to develop approximation algorithms for job shop scheduling problems (cf. [211]).

In fluid relaxation, we replace the constraint $\mathbf{q}_f \in \mathcal{Q}_f$ by $\mathbf{q}_f \in \widetilde{\mathcal{Q}}_f$, with $\widetilde{\mathcal{Q}}_f$ defined as the set of power consumption profiles $\mathbf{q}_f \in \mathbb{R}^T$ satisfying the following constraints

$$\mathbf{1}^\top \mathbf{q}_f \le \mathbf{1}^\top \widehat{\mathbf{q}}_f, \tag{9.11a}$$

$$\mathrm{TV}(\mathbf{q}_f) \le \mathrm{TV}(\widehat{\mathbf{q}}_f), \tag{9.11b}$$

$$q_f(t) \ge 0, \quad t \in \{t_f^{\mathrm{ES}}, \ldots, t_f^{\mathrm{LC}}\}, \tag{9.11c}$$

$$q_f(t) = 0, \quad t \notin \{t_f^{\mathrm{ES}}, \ldots, t_f^{\mathrm{LC}}\}, \tag{9.11d}$$

where the total variation of a vector $\mathbf{x} \in \mathbb{R}^d$ is defined as

$$\mathrm{TV}(\mathbf{x}) = \sum_{t=0}^{d+1} |x(d+1) - x(t)|, \quad x(0) = x(d+1) := 0.$$

In the definition of $\widetilde{\mathcal{Q}}_f$, constraint (9.11a) requires that the total energy of the scheduled consumption profile is no larger than the total energy of the submitted consumption profile; constraints (9.11c) and (9.11d) impose the non-negativity requirement for the consumption profile and restrict the profile can only have positive consumption in time periods specified by the time window submitted by the user; constraint (9.11b) controls the flexibility of the scheduled profile, by limiting the total variation of the scheduled profile with that of the submitted profile. It is easy to see that $\mathcal{Q}_f \subset \widetilde{\mathcal{Q}}_f$.

As $\widetilde{\mathcal{Q}}_f$ is a convex polytope for each f, the resulting convex relaxation for scheduling flexible orders is

$$\max_{\{\mathbf{q}_f\}} \quad J(\{\mathbf{q}_f\}) \tag{9.12a}$$

$$\text{s.t.} \quad \mathbf{q}_f \in \widetilde{\mathcal{Q}}_f, \quad f \in \mathcal{F}, \tag{9.12b}$$

whose solution is denoted by $\{\widetilde{\mathbf{q}}_f\}$.

The solution of the convex relaxation may be infeasible with respect to the original constraint $\mathbf{q}_f \in \mathcal{Q}_f$. One possibility is that for some f, constraint (9.11a) may not hold with equality at the solution $\widetilde{\mathbf{q}}_f$. In this case, we simply round down such that $\mathbf{q}_f(t) = 0$ for all t. If constraint (9.11a) holds with equality at the solution and so the energy requirement of the flexible load is satisfied, we identify a feasible starting time t_f^{S} for each $f \in \mathcal{F}$ by finding the time window with τ_f^{D} periods that contains the maximum total power consumption according to $\widetilde{\mathbf{q}}_f$, i.e.,

$$t_f^{\mathrm{S}} = \operatorname*{arg\,max}_{t_f^{\mathrm{S}} \in \{t_f^{\mathrm{ES}}, \ldots, t_f^{\mathrm{LC}} - \tau_f^{\mathrm{D}} + 1\}} \sum_{t=t_f^{\mathrm{S}}}^{t_f^{\mathrm{S}} + \tau_f^{\mathrm{D}} - 1} \widetilde{q}_f(t). \tag{9.13}$$

The "projected" power consumption schedule is thus the power consumption profile induced by this starting time, denoted by $\mathbf{q}_f(t_f^{\mathrm{S}})$:

$$\Pi_{\mathcal{Q}_f}\left(\widetilde{\mathbf{q}}_f\right) := \begin{cases} \mathbf{q}_f(t_f^{\mathrm{S}}), & \text{if } \mathbf{1}^\top \widetilde{\mathbf{q}}_f = \mathbf{1}^\top \widehat{\mathbf{q}}_f, \\ \mathbf{0}, & \text{otherwise.} \end{cases} \tag{9.14}$$

with t_f^{S} defined in (9.13).

Performance guarantees

In this section, we provide a stylized analysis of the proposed scheduling algorithms for solving the non-convex (combinatorial) optimization (9.6) with the following two assumptions.

A1 The network can be represented as a single-bus network.

A2 The maximization criteria is the total surplus (9.3).

Assumption **A**1 allows us to focus on the combinatorial nature of the problem rather than the network constraints. We stress, however, both of the proposed algorithms result in admissible schedules that respect distribution network constraints for general networks. Analyzing the performance of approximation algorithms with network constraints represents a major challenge and is left for future work. Assumption **A**2 is introduced without loss of generality. In fact, the total volume objective function (9.4) can be thought of as a special case of the total surplus objective, with prices suitably modified.

We start with the performance guarantee for the greedy algorithm[6]. As is often the case, the greedy algorithm is $(1-1/e)$-optimal if the underlying problem is *submodular*. For our problem of scheduling flexible orders, by extending the analysis in [209], we can show submodularity indeed holds and therefore we have the following performance guarantee for the greedy algorithm. Proofs are omitted and may be found in [214].

Lemma 5 (Greedy performance)**.** *Under Assumptions* **A***1 and* **A***2, the greedy scheduling algorithm is* $(1-1/e)$*-optimal:*

$$\frac{J(\{\mathbf{q}_f^{\mathrm{g}}\})}{J(\{\mathbf{q}_f^{\star})\}} \geq 1 - \frac{1}{e}, \tag{9.15}$$

where $\{\mathbf{q}_f^{\mathrm{g}}\}$ *is the greedy schedule and* $\{\mathbf{q}_f^{\star}\}$ *is the optimal schedule.*

[6] A slight modification (i.e., using a specific initialization of $\mathcal{F}^{\mathrm{S}}$ instead of $\mathcal{F}^{\mathrm{S}} = \emptyset$) of the greedy algorithm presented in (3) may be needed due to the knapsack constraint (9.10b). The modified greedy algorithm is still polynomial time but is much slower due to the time-consuming initialization step. We stated the simple greedy algorithm in Algorithm 3 because we have been observing similar performances with and without the modification. See [212] and [213] for more discussions.

One important feature of the performance bound (9.15) is that it is independent of the problem instance. It is thus referred to as an *a priori* bound as it can be stated before seeing the actual problem data.

For the fluid heuristic, we can derive general *a posteriori* bounds based on the following observation:

Proposition 1. *Let* $\{\widetilde{\mathbf{q}}_f\}$ *and* $\{\Pi_{\mathcal{Q}_f}(\widetilde{\mathbf{q}}_f)\}$ *be the solutions of fluid relaxation* (9.12) *and its projected solution, respectively. Then*

$$J(\{\Pi_{\mathcal{Q}_f}(\widetilde{\mathbf{q}}_f)\}) \leq J(\{\mathbf{q}_f^\star\}) \leq J(\{\widetilde{\mathbf{q}}_f\}). \tag{9.16}$$

This result bounds the unknown quantity $J(\{\mathbf{q}_f^\star\})$ by quantities that are computable in the fluid relaxation steps. This bound is general in that it holds without assuming **A**1 and **A**2. As such, given any problem instance and having computed $J(\{\Pi_{\mathcal{Q}_f}(\widetilde{\mathbf{q}}_f)\})$ and $J(\{\widetilde{\mathbf{q}}_f\})$, we can gauge the (sub-)optimality of the solution by comparing these two quantities. If they are close, we can assert *a posteriori* that the fluid algorithm has produced a $J(\{\Pi_{\mathcal{Q}_f}(\widetilde{\mathbf{q}}_f)\})/J(\{\widetilde{\mathbf{q}}_f\})$ optimal solution.

We can in fact show that the fluid heuristic is asymptotically optimal, with the following additional assumption:

A3 For all $f \in \mathcal{F}$, $\widehat{q}_f(t) \equiv \bar{q}_f$ does not change with t, $t = 1, \ldots, \tau_f^{\mathrm{D}}$. Furthermore, $\bar{q}^{\min} \leq \bar{q}_f \leq \bar{q}^{\max}$ for all $f \in \mathcal{F}$, where $0 < \bar{q}^{\min} < \bar{q}^{\max} < \infty$.

A4 The supply is sufficient as for each period $t \in \mathcal{T}$ there is a supply order with unbounded $\bar{q}$ and a large acceptable price $\underline{\pi}_{\mathrm{u}} \geq \max_{s\in\mathcal{S}} \underline{\pi}_s$. Furthermore, among $|\mathcal{F}|$ flexible orders, there are at least $\alpha|\mathcal{F}|$ flexible orders such that $\overline{\pi}_f \geq \underline{\pi}_{\mathrm{u}}$, where $\alpha \in (0, 1)$.

Assumption **A**3 replaces possibly time-varying load shapes by rectangles. This certainly limits the practicality of our next result. Removing it is possible as any time-varying load shapes can be approximated, with any desired level of accuracy, by rectangles. Assumption **A**4 focuses our analysis on the case where there is enough supply for each future delivery interval. This can be the case when a utility company (or load serving entity) participates in the platform and offers to sell sufficient amount of energy at a relatively high price $\underline{\pi}_{\mathrm{u}}$. Under these assumptions, we can establish the (weak) asymptotic optimality of the proposed fluid heuristic, in the following sense.

Theorem 20 (Asymptotic optimality of fluid heuristic)**.** *Under Assumptions* **A***1-*__A__*4, consider a set of flexible orders with increasing size* $|\mathcal{F}| \to \infty$*. For each* $\mathcal{F}$*, there exists an optimal schedule* $\{\widetilde{\mathbf{q}}_f\}$ *for the fluid relaxation* (9.12) *such that*

$$\frac{J(\{\Pi_{\mathcal{Q}_f}(\widetilde{\mathbf{q}}_f)\})}{J(\{\mathbf{q}_f^\star\})} \to 1, \quad \textit{as} \quad |\mathcal{F}| \to \infty, \tag{9.17}$$

where $\{\mathbf{q}_f^\star\}$ *is an optimal schedule for the original problem* (9.6)*.*

The key intuition behind this theorem is that when the number of flexible orders becomes large, individual load shapes no longer play a significant role in the quality of the solution because (with a single-bus network) it is the aggregate load shape of all the flexible loads that matters in the optimization (9.6).

Pricing

Algorithms presented in Section 9.3 only provides admissible schedules for orders in the orderbook, with clearing prices to be determined. Denote the collection of prices for orders by $\{\pi_s\}, \{\pi_b\}, \{\pi_f\}$. Basic requirements for these prices include (i) admissible, as defined in Definition 18, and (ii) *revenue adequate* so that the *merchandising surplus* of the platform, denoted by MS, is non-negative:

$$\text{MS} = \sum_{b\in\mathcal{B}} \pi_b \mathbf{1}^\top \mathbf{q}_b + \sum_{f\in\mathcal{F}} \pi_f \mathbf{1}^\top \mathbf{q}_f - \sum_{s\in\mathcal{S}} \pi_s \mathbf{1}^\top \mathbf{q}_s \geq 0. \tag{9.18}$$

Without flexible orders, it can be shown that both requirements above can be met with a generalization of the simple idea of *marginal pricing*. In the rest of this section, we first state this generalization and then exam its properties when used with the scheduling algorithms proposed in Section 9.3.

Marginal pricing

Given any schedule of the flexible orders $\{\mathbf{q}_f\}$, we consider the optimization for scheduling the remaining simple orders (9.5). Denote the optimal dual variable associated with constraint (9.5d) of the linear program by $\boldsymbol{\lambda}_n \in \mathbb{R}^T$, $n \in \mathcal{N}$. This is a collection of $N \times T$ prices, one for each *(bus, future delivery time)* pair. Thus these prices may be referred to as *temporal and locational marginal prices* [215], which are functions of the schedules of flexible orders in our setting.

In particular, for orders with non-zero cleared quantities, we define a *marginal pricing rule* as

$$\pi_s = \lambda_n(t), \quad s = (n,\ t,\ \overline{q},\ \underline{\pi}), \tag{9.19a}$$

$$\pi_b = \lambda_n(t), \quad b = (n,\ t,\ \overline{q},\ \overline{\pi}), \tag{9.19b}$$

$$\pi_f = \frac{\boldsymbol{\lambda}_n^\top \mathbf{q}_f}{\mathbf{1}^\top \mathbf{q}_f}, \quad f = (n,\ t^{\text{ES}},\ t^{\text{LC}},\ \tau^{\text{D}},\ \widehat{\mathbf{q}},\ \overline{\pi}). \tag{9.19c}$$

Thus under the marginal pricing rule, simple orders are paid or charged the locational marginal price for the future delivery time interval. As flexible orders usually span multiple delivery time intervals, they are charged the average locational marginal prices weighted by the amount of power they consume in different time intervals.

Properties of marginal pricing

We analyze properties of the marginal pricing rule based on Assumption **A**2, otherwise the dual variables of optimization (9.5) may not have a clear economic meaning.

Using the KKT condition of the linear program (9.5) and strong duality, we can establish the following property of the marginal pricing rule, given any schedules for the flexible orders $\{\mathbf{q}_f\}$:

Lemma 6 (Properties of marginal pricing)**.** *Under Assumption* **A***2, the marginal pricing rule is revenue adequate and leads to admissible prices for simple buy orders and simple sell orders.*

As the schedules of flexible orders $\{\mathbf{q}_f\}$ are not decision variables of the optimization (9.5), little can be said about whether the marginal prices will be admissible for flexible orders without considering the actual algorithms used to determine these schedules. Considering the algorithms proposed in Section 9.3, we have the following negative results established using counter examples:

Lemma 7 (Inadmissibility for flexible orders)**.** *Under Assumption* **A***2, the marginal pricing rule with greedy or fluid schedule is not guaranteed to produce admissible prices for flexible buy orders.*

The observation above stems from the fundamental difficulties in non-convex pricing problems and the fact that the proposed algorithms use more than the marginal information to determine the schedules of flexible orders. To be precise, in each step of the greedy algorithm, it determines whether to schedule a flexible order based on the *total benefit* about which it brings measured by the change of function value $J(\{\mathbf{q}_f\})$ with and without the newly scheduled order. Charging it with the marginal cost, which corresponds to the cost of producing the last $\epsilon > 0$ amount of power, may result in a price higher than $\overline{\pi}_f$. For the fluid algorithm, the projection step does not use price information and may result in inadmissible prices.

While marginal pricing has many desirable properties as studied in the transmission market literature (cf. [216] and references therein), modifying it to extend some of these properties to the non-convex setting here require much additional work. Instead, in the next subsection, we provide a simple mechanism that will ensure price admissibility and budget adequacy.

Name-your-own-price mechanism

In the *name-your-own-price* mechanism, alternatively a *pay-as-bid* mechanism, we simply pay or charge users based on the prices they submit with their order

$$\pi_s = \underline{\pi}_s, \quad \pi_b = \overline{\pi}_b, \quad \pi_f = \overline{\pi}_f, \tag{9.20}$$

for all (partially) fulfilled orders. By definition, this pricing mechanism produces admissible prices for all orders. We can also show that revenue adequacy is achieved if we use greedy scheduling algorithm with the total surplus objective:

Lemma 8. *Under Assumption* **A***2, the name-your-own-price mechanism is revenue adequate with greedy scheduling.*

It is an easy consequence of Theorem 20 that if we use the fluid scheduling algorithm we are guaranteed revenue adequacy in an asymptotic sense.

Corollary 3. *Under the same assumptions of Theorem 20 and with the name-your-own-price mechanism, consider a set of flexible orders with increasing size* $|\mathcal{F}| \to \infty$*. For each* $\mathcal{F}$*, there exists an optimal schedule* $\{\tilde{\mathbf{q}}_f\}$ *for the fluid relaxation* (9.12) *such that the merchandizing surplus induced by the resulting schedule* $\{\Pi_{\mathcal{Q}}(\tilde{\mathbf{q}}_f)\}$ *satisfies*

$$\lim_{|\mathcal{F}|\to\infty} \mathrm{MS} > 0.$$

In summary, both proposed scheduling algorithms result in admissible and revenue-adequate prices if the name-your-own-price mechanism is used.

9.4 Logic-Based Solution Methodology

The Constrained, Coupled, Continuous Double Auction (C^{3}DA)

We introduce an alternate matching algorithm, *The Constrained, Coupled, Continuous Double Auction* (C^{3}DA), which in contrast to the *greedy* and *fluid* algorithms requires no optimization, and more closely resembles the operation of a traditional stock market. The lack of optimization means the algorithm can be much faster and has the potential for greater scalability.

In a classic continuous double auction (CDA), buyers and sellers continuously submit bids (an offer to buy at price π_b) and asks (an offer to sell at price π_a) respectively for homogeneous, single-attribute goods, and the market clears whenever a bid exceeds an ask (i.e. when $\pi_b \geq \pi_a$). Bids and asks are typically listed on a bulletin board, and the auction runs throughout a 'trading day' with a deadline, after which no more orders are accepted. Various order matching protocols exist to tiebreak among orders made at the same price, such as FiFo (First in, First out) and Pro-Rata. CDAs are generally unconstrained, in that there are no external limits on which trades may be made. They are also generally uncoupled, both in time and space. In time, each trading day is considered to be independent, and one may not make bids in one trading day for a subsequent trading day.[7] In space, the price among geographically separate commodity markets is rarely explicity coupled, perhaps

[7]In commodity markets, there is typically both a spot and futures market, with the latter partially fulfilling this coupling role, allowing parties to transact for delivery of the commodity at a future date.

with the exception of nodal electricity markets. There may of course be an implicit price coupling, reflecting the cost of transport of the commodity between different markets. There is no central optimization in a CDA, or any kind of welfare maximization.

The C^{3}DA consists of a set of nodal CDAs connected by a constrained transport network, employing an open-gate forward market structure. Participants can make orders at any instance in time for a future market $t \in \mathcal{T}$, and the matching algorithm is triggered whenever a new order arrives. To determine which nodes can trade with which other nodes, there is a computational step that is triggered every time a new bid clears, to determine the amount of power that can be safely sent between nodes. Flexible bids are represented in the simple orderbooks by the *lowest priced equivalent simple order* that would *clear the flexible bid if it were met*.

Additional Orderbooks

We refer to the orderbooks described in Definition 13, $\mathcal{S}$, $\mathcal{B}$, $\mathcal{F}$ as the *Global Orderbook*. To recap, the nodal orderbooks, $\mathcal{S}_n$, $\mathcal{B}_n$, $\mathcal{F}_n$, defined in Section 9.2, detail all offers made at a give node. For the C^{3}DA it is necessary to introduce the concept of the *Local Orderbook*. We will first provide an informal definition, and then a more technical definition after some preliminaries. The local orderbook at node n, contains all orders made at nodes which can make *feasible trades* with node n. For example, in an unconstrained network, every local orderbook at each node would correspond to the global orderbook, since every node can make trades with every other node. We denote the local orderbooks as $\widehat{\mathcal{S}}_n$, $\widehat{\mathcal{B}}_n$, $\widehat{\mathcal{F}}_n$, and we have the following relation between orderbooks

$$\mathcal{B}_n \subseteq \widehat{\mathcal{B}}_n \subseteq \mathcal{B} \tag{9.21a}$$

$$\mathcal{S}_n \subseteq \widehat{\mathcal{S}}_n \subseteq \mathcal{S} \tag{9.21b}$$

$$\mathcal{F}_n \subseteq \widehat{\mathcal{F}}_n \subseteq \mathcal{F} \tag{9.21c}$$

The local orderbook is a function of the network constraints and previously cleared orders whose delivery falls within $\mathcal{T}$. It must be recalculated whenever a new set of orders is cleared. We denote the resulting net injection from previously cleared orders as $\mathbf{p}^c(t) \in \mathbb{R}^N$, $t \in \mathcal{T}$. We say that two nodes m, n can make a feasible trade in the direction $m \to n$ at time t, if $\exists\ \alpha_{m,n} > 0$, such that

$$\alpha_{m,n}\mathbf{e}_{m,n} + \mathbf{p}^c(t) \in \mathcal{P}_D \tag{9.22}$$

where $\alpha_{m,n} \in \mathbb{R}$, and $\mathbf{e}_{m,n}$ denotes the unit injection vector between m, n, defined to have 1 in the mth position, -1 in the nth position, and 0 elsewhere. We can in fact calculate the maximum amount of power that can be transferred between nodes m, n, denoted $\overline{\alpha}_{m,n}$. We assume the linear constraint set is defined in the $\mathcal{H}$ representation, such that $\mathcal{P}_D = \{p : Ap \leq b\}$[8]. We can then see that

$$(9.22) \Rightarrow \alpha_{m,n}A\mathbf{e}_{m,n} \leq b - A\mathbf{p}^c(t) \tag{9.23}$$

[8]Equality constraints of the form $a^\top p = b$, are included in the $\mathcal{H}$ representation through $a^\top p \leq b$, $a^\top p \geq b$

and that

$$\overline{\alpha}_{m,n} = \min_k \left(\max \left(\frac{b_k - A_k \mathbf{p}^c(t)}{A_k \mathbf{e}_{m,n}}, 0 \right) \right) \tag{9.24}$$

where A_k is the kth row of A, and b_k is the kth element of b. We are effectively attempting to find the first constraint that would bind if we move in the direction $\mathbf{e}_{m,n}$ from $\mathbf{p}^c(t)$. Since the power balance constraint is trivially satisfied by $\mathbf{e}_{m,n}$, it can be shown that $2N(N-1)(2N-1)$ operations are required in this case to compute the feasible power transfer between each pair of nodes on the network. Additionally it can be shown that for each pair m, n, only constraints k where $A_k \mathbf{e}_{m,n} > 0$ need be considered in (9.24), since these will always bind before constraints where $A_k \mathbf{e}_{m,n} \leq 0$. This calculation can be performed offline, reducing the online computational burden. Assuming that vectors $A_k^\top$ are uniformly distributed in space, this would reduce the number of computations required in expectation by a factor of two.

We can now formally define the local orderbook.

Definition 19 (Local Orderbook)**.** *The local orderbook at node n at time t is defined to be the triple $(\widehat{\mathcal{B}}_n, \widehat{\mathcal{S}}_n, \widehat{\mathcal{F}}_n)$, where $\widehat{\mathcal{B}}_n$ is the collection of standing simple buy orders, $\widehat{\mathcal{S}}_n$ is the collection of standing simple sell orders, and $\widehat{\mathcal{F}}_n$ is the collection of standing flexible buy orders, each of which correspond to power supply/demand delivered at nodes which can make feasible trades with node n.*

Formally,

$$\widehat{\mathcal{B}}_n = \bigcup_m \mathcal{B}_m, \qquad \forall \overline{\alpha}_{m,n} > 0 \tag{9.25}$$

$$\widehat{\mathcal{S}}_n = \bigcup_m \mathcal{S}_m, \qquad \forall \overline{\alpha}_{n,m} > 0 \tag{9.26}$$

$$\widehat{\mathcal{F}}_n = \bigcup_m \mathcal{F}_m, \qquad \forall \overline{\alpha}_{m,n} > 0 \tag{9.27}$$

Treatment of Simple and Flexible Bids

It is now useful to define the outstanding local bid and ask.

Definition 20 (Local Outstanding Bid)**.** *The local outstanding bid at node n at time t, denoted $b_n^\star$ is the highest priced buy order in the local buy orderbook*[9]*:*

$$b_n^\star = \arg\max_{b \in \widehat{\mathcal{B}}_n} \ \overline{\pi}_b. \tag{9.28}$$

Definition 21 (Local Outstanding Ask)**.** *The local outstanding ask at node n at time t, denoted $s_n^\star$ is the lowest priced sell order in the local sell orderbook:*

$$s_n^\star = \arg\min_{s \in \widehat{\mathcal{S}}_n} \ \underline{\pi}_s \tag{9.29}$$

[9]Tiebreaking between orders with same price can be achieved using the FIFO protocol

The outstanding local bid and ask are the prices that are communicated to market participants as 'market prices' in real time. *i.e.* the price a buyer (seller) would have to bid (ask) at to clear their order.

Each time a new simple buy (sell) order arrives from node n, it is immediately matched against the local outstanding ask (bid) at node n, and if $\overline{\pi} \geq \underline{\pi}$, then the order is cleared. If the order can only be partially cleared, it is repeatedly matched with other orders in order of increasing (decreasing) price in the local orderbook, until it is completely cleared, or there is no further matching order. Any uncleared quantity remaining on the order is added to the standing orderbook.

This framework is straightforward enough for simple orders, but it is less obvious how to integrate flexible orders. At node n, there will be a sequence of local orderbooks $\widehat{\mathcal{S}}_n(t)$, $\forall t \in \mathcal{T}$. To match a flexible order f we check each candidate start time for the deferrable bid and find its average price over the duration of the demand, matching it to the lowest priced simple sell orders in the local orderbook in each period. If the lowest among these average prices is less than the bid price of the deferrable load, then the order is cleared. Otherwise the order is added to the collection of standing flexible orders.

In the event that the order is not cleared, we want to reflect the flexible order in the standing orderbook, such that it can impact the outstanding local bid and ask. One approach is to find the implied simple, single-period order that *would* clear the flexible bid if matched. From the above candidate start times, we take the start time with the lowest average price. For each period k of the demand duration, we find the quantity weighted price that would have cleared the deferrable load, which we denote $\widetilde{\pi}_k$. For each time period in the duration of the flexible order $k = 1, \ldots, \tau^{\mathrm{D}}$, we have

$$\overline{\pi} \sum_{l=1}^{\tau^{\mathrm{D}}} q_l = \widetilde{\pi}_k q_k + \sum_{l \neq k}^{\tau^{\mathrm{D}}} \pi_l q_l \tag{9.30}$$

$$\widetilde{\pi}_k = \frac{\overline{\pi} \sum_{l=1}^{\tau^{\mathrm{D}}} q_l - \sum_{l \neq t}^{\tau^{\mathrm{D}}} \pi_l q_l}{q_k} \tag{9.31}$$

$$\widetilde{\pi}_k = \overline{\pi} + \frac{\sum_{l \neq t}^{\tau^{\mathrm{D}}} (\overline{\pi} - \pi_l) q_l}{q_k} \tag{9.32}$$

where π_l is the average price for quantity q_l to be cleared in period l from the outstanding local orderbook. The deferrable load makes the implied bid

$$b = (n, t, q_t, \widetilde{\pi}_t), \quad t = 1, \ldots, \tau^{\mathrm{D}} \tag{9.33}$$

The purpose of this implied bid is to provide an appropriate signal to sellers to show, in the same way as for simple orders, what ask price would be needed to match with the deferrable load order. The benefit of this approach is that we need only check if the implied bid clears to check the clearing of the entire deferrable load. The implied bids for flexible orders should be recalculated every time a new order clears. Alternatively they could be recalculated periodically. It remains to be seen what the most computationally efficient way to handle this might be.

Pricing

In general we adopt the κ-pricing protocol. In the case of simple orders the clearing price π_c, both paid by the buyer and received by the seller, takes the following form.

$$\pi_c = \underline{\pi} + \kappa(\overline{\pi} - \underline{\lambda}) \tag{9.34}$$

$$= \kappa\overline{\pi} + (1-\kappa)\underline{\pi} \tag{9.35}$$

where $\kappa \in [0, 1]$ is a constant determining the surplus allocation. $\kappa = 0.5$ is a common choice where this protocol is adopted.

For the case of multi-period orders, we adopt the following protocol. The maximum payment that the buyer is willing to make is equal to $\overline{\pi}\mathbf{1}^\top\widehat{\mathbf{q}}$, and the required minimum payment to matched sellers is equal to $\sum_s \underline{\pi}_s q_s$, where s is the index of matched sell orders. A revenue adequate pricing scheme might take the following form. The buyer pays

$$\pi_c = \kappa\overline{\pi} + (1-\kappa)\frac{\sum_s \underline{\pi}_s q_s}{\mathbf{1}^\top\widehat{\mathbf{q}}} \tag{9.36}$$

It can be shown that the surplus (in \$) to sellers is equal to

$$\kappa(\overline{\pi}\mathbf{1}^\top\widehat{\mathbf{q}} - \sum_s \underline{\pi}_s q_s) \geq 0 \tag{9.37}$$

Each seller receives a quantity weighted allocation of this surplus, which represents a constant price adder to all sellers' bid prices. The price paid to each seller s is denoted $\pi_{c,s}$ and can be calculated as

$$\pi_{c,s} = \underline{\pi}_s + \frac{1}{q_s}\frac{q_s}{\mathbf{1}^\top\widehat{\mathbf{q}}}\kappa\left(\overline{\pi}\mathbf{1}^\top\widehat{\mathbf{q}} - \sum_s \underline{\pi}_s q_s\right) \tag{9.38}$$

$$= \underline{\pi}_s + \kappa\left(\overline{\pi} - \frac{\sum_s \underline{\pi}_s q_s}{\mathbf{1}^\top\widehat{\mathbf{q}}}\right) \tag{9.39}$$

This pricing scheme is revenue adequate since $\pi_c \leq \overline{\pi}$ and $\pi_{c,s} \geq \underline{\pi}_s$, $\forall s$.

9.5 Conclusion

In this chapter, we propose APEX, a market platform that enables monetization of underutilized distribution system assets. It features an open-gate forward market design and an explicit flexibility market. The forward markets help to incorporate variable distributed energy resources and reduce risks of market participants. In the flexibility market, resources submit their flexibility information with a simple yet expressive order format and APEX schedule these resources on behalf of the users efficiently. All functionalities of APEX are executed while ensuring the reliability of the distribution network, either by directly managing the distribution system constraints, or by communicating with a minDSO through a coordinated trading protocol.

For the proposed market platform, we study the non-convex problem of scheduling non-preemptive flexible resources and propose two optimization-based polynomial time algorithms that have finite or asymptotic performance guarantees. We analyze the properties of marginal pricing together with the proposed algorithms and suggest a simple alternative that leads to admissible prices for all users and guarantees adequate revenue for the platform. We also propose a logic-based market clearing methodology, termed the C^3DA, which extends the classic notion of CDAs to the case of coupled time-periods and constrained transport networks. We demonstrate how feasible trades are identified and propose a revenue adequate pricing scheme.

Appendix A

A.1

Analytic derivation of the solution of MPED.

Proof. For situations where $\Delta_{\sigma_t} = \mathbf{I}$, the solution can be derived analytically. This corresponds to the case where $\lambda_{t+1} = \lambda_t$, arising when the LMPs are the same across periods when there is no line congestion, or very special cases when there is line congestion. This can in general extend over a number of consecutive time periods, which we will denote $\widetilde{T}$, such that $\lambda_t = \lambda_{t+1} = \cdots = \lambda_{t+\widetilde{T}}$. We also define $\sigma(t) := \{t_1, \ldots, t_{\widetilde{T}}\}$ as the set of consecutive time periods for which $\Delta_{\sigma_j} = \mathbf{I}$, $j \in \{t_1, \ldots, t_{\widetilde{T}-1}\}$, and $\Delta_{\sigma_{t_{\widetilde{T}}}} = 0$. It is always true that $t \in \sigma(t)$. As an example, say that $\Delta_{\sigma_1} = \mathbf{I}$, $\Delta_{\sigma_2} = 0$, implying that $\lambda_2 = \lambda_1$. Then $\sigma(1) = \sigma(2) = \{1, 2\}$.

This allows the problem to decouple in time across segments where $\lambda_t^\star$ is distinct, giving the following form

$$\max_{\lambda_t, \gamma, \widetilde{\beta}} \ \psi_t(\lambda, \gamma, \widetilde{\beta}) \tag{A.1a}$$

$$\text{s.t. } \lambda_t = \gamma_j \mathbf{1} - H_t^\top \widetilde{\beta}_j, \qquad j \in \sigma(t) \tag{A.1b}$$

where

$$\psi_t(\lambda_t, \gamma, \widetilde{\beta}) = \sum_{j \in \sigma(t)} -\frac{1}{2}(\lambda_t - a_j)^\top Q_j^{-1}(\lambda_t - a_j) + d_j^\top \lambda_t - c_j^\top \widetilde{\beta}_j - s^\top(\Delta_{\chi_j} - \Delta_{\chi_{j-1}})\lambda_t \tag{A.2}$$

Given that the columns of $H_t^\top$ are linearly independent of $\mathbf{1}$, it can be shown that γ_j, $\widetilde{\beta}_j$ are identical for all $j \in \sigma(t)$, simplifying the analytical solution considerably. The proof then closely follows the proof detailed in [34]. The results in this case are very similar to the original, but with some subtle differences relating to $Q_{\sigma(t)}$. □

A.2

The matrix $A_t(W_t)$ is positive semi-definite. That is $A_t(W_t) \succeq 0$. Additionally, $M_t \succeq 0$, $K_t \succ 0$.

Proof. We first show that $M_t \succeq 0$. We observe that M_t has the form of a Schur complement.

$$M_t = Q_t^{-1} - Q_t^{-1}\mathbf{1}(\mathbf{1}^\top Q_t^{-1}\mathbf{1})^{-1}\mathbf{1}^\top Q_t^{-1} \tag{A.3}$$

We form the associated symmetric block matrix, denoted X_t, where A, B, C represent general notation for its blocks

$$X_t = \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix} = \begin{bmatrix} Q_t^{-1} & Q_t^{-1}\mathbf{1} \\ \mathbf{1}^\top Q_t^{-1} & \mathbf{1}^\top Q_t^{-1}\mathbf{1} \end{bmatrix} \tag{A.4}$$

We see that $M_t = A - BC^{-1}B^\top = X_t/C$, that is the Schur complement of C in X_t. We have that if $C \succ 0$, then $X_t \succeq 0 \Leftrightarrow X_t/C \succeq 0$. Since $C = \mathbf{1}^\top Q_t^{-1}\mathbf{1} \succ 0$, due to the fact that $Q_t^{-1} \succ 0$, to show that $M_t \succeq 0$, it suffices to show that $X_t \succeq 0$. This can be seen by observing that X_t can be decomposed as

$$X_t = \begin{bmatrix} \mathbf{I} \\ \mathbf{1}^\top \end{bmatrix} Q_t^{-1} \begin{bmatrix} \mathbf{I} & \mathbf{1} \end{bmatrix} \tag{A.5}$$

$$= N^\top Q_t^{-1} N \tag{A.6}$$

$$\succeq 0 \tag{A.7}$$

where $N = [\mathbf{I}\ \mathbf{1}]$. This follows from a result in [217], which states that: for $Q \in \mathbb{S}^{n\times n}$, and $N \in \mathbb{C}^{n\times m}$,

Remark 7. *Suppose $Q \succeq 0$, then $N^*QN \succeq 0$.*

Remark 8. *Suppose $Q \succ 0$, then $N^*QN \succ 0$ if and only if rank$(N) = m$*

In our case above we observe that the condition in Remark 7 is satisfied, but that $N \in \mathbb{R}^{n\times n+1}$, and rank$(N) = n$, so the condition in Remark 8 is not satisfied. Thus $X_t \succeq 0$.

We can additionally show that M_t has exactly one eigenvalue at 0, associated with the eigenvector $\mathbf{1}$. This is proved in Appendix A.3.

We now show that $K_t \succ 0$. We begin by writing out the full expansion of K_t, observing that it also has the form of a Schur complement.

$$K_t = H_t M_t H_t^\top \tag{A.8}$$

$$= H_t Q_t^{-1} H_t^\top - H_t Q_t^{-1}\mathbf{1}(\mathbf{1}^\top Q_t^{-1}\mathbf{1})^{-1}\mathbf{1}^\top Q_t^{-1} H_t^\top \tag{A.9}$$

Again we form the associated symmetric block matrix, denoted Z_t. We abuse notation and use A, B, C to refer to the blocks of Z_t.

$$Z_t = \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix} = \begin{bmatrix} H_t Q_t^{-1} H_t^\top & H_t Q_t^{-1}\mathbf{1} \\ \mathbf{1}^\top Q_t^{-1} H_t^\top & \mathbf{1}^\top Q_t^{-1}\mathbf{1} \end{bmatrix} \tag{A.10}$$

We note that $K_t = Z_t/C$, and that $Z_t \succ 0 \Leftrightarrow C \succ 0, Z_t/C \succ 0$. We show that $Z_t \succ 0$, by decomposing it as

$$Z_t = \begin{bmatrix} H_t \\ \mathbf{1}^\top \end{bmatrix} Q_t^{-1} \begin{bmatrix} H_t^\top & \mathbf{1} \end{bmatrix} \tag{A.11}$$

$$= F_t^\top Q_t^{-1} F_t \tag{A.12}$$

$$\succ 0 \tag{A.13}$$

where $F_t = [H_t^\top \ \mathbf{1}]$. We see that Z_t satisfies the stricter condition in Remark 8. We observe that $F_t \in \mathbb{R}^{n \times m_t + 1}$, H_t is assumed to have full row rank by Flow LICQ, and $\mathbf{1}$ is linearly independent of the columns of $H_t^\top$, thus $\text{rank}(F_t) = m_t + 1$, and $Z_t \succ 0$. This implies that $Z_t/C = K_t \succ 0$.

It follows that $R_t \succeq 0$ from Remark 7. It can then be seen that $A_t(W_t) \succeq 0$ by Weyl's inequality, or alternately forming a Schur complement construction similar to those above.

The proof of $P_t \preceq 0$ follows from a similar Schur complement construction as that used to prove $M_t \succeq 0$. □

A.3

$M_t \succeq 0$, with exactly one eigenvalue at 0, with corresponding eigenvector $\mathbf{1}$.

Proof. For clarity we drop the subscript t in this proof. We begin by considering the matrix $(-M)$. We denote the vector of eigenvalues of $(-M)$ as φ, and denote the vector of eigenvalues (the diagonal entries) of $(-Q^{-1})$ as z. We see that $(-M)$ can be rewritten as a rank-one modification of a diagonal matrix

$$-M = -Q^{-1} + \frac{Q^{-1}\mathbf{1}\mathbf{1}^\top Q^{-1}}{\mathbf{1}^\top Q^{-1}\mathbf{1}} \tag{A.14}$$

$$= -Q^{-1} + \rho z z^\top \tag{A.15}$$

Where $\rho = \frac{1}{\mathbf{1}^\top Q^{-1}\mathbf{1}} = -\frac{1}{\mathbf{1}^\top z}$ is defined as in (6.41). The eigenvalues of such a diagonal-plus-rank-one matrix are the zeros of the secular function [218]

$$f(\varphi) = 1 + \rho z^\top (-Q^{-1} - \varphi I)^{-1} z \tag{A.16}$$

We show that 0 is an eigenvalue of $-M$. Noting that $z^\top Q = -\mathbf{1}^\top$

$$f(0) = 1 + \rho z^\top (-Q^{-1})^{-1} z \tag{A.17}$$

$$= 1 - \rho z^\top Q z \tag{A.18}$$

$$= 1 - \frac{\mathbf{1}^\top z}{\mathbf{1}^\top z} \tag{A.19}$$

$$= 0 \tag{A.20}$$

Following the proof in [218], the eigenvalues of $-M$ and $-Q^{-1}$ follow the strict interlacing property,

$$\varphi_1 > z_1 > \varphi_2 > z_2 > \cdots > \varphi_n > z_n \tag{A.21}$$

where the elements of φ and z are decreasingly ordered, $\varphi_1 > \varphi_2 > \cdots > \varphi_n$, $z_1 > z_2 > \cdots > z_n$. Since $z_i < 0, \forall i = 1, \ldots, n$, we must have that $\varphi_1 = 0$, which is the largest eigenvalue of $(-M)$, and that $\varphi_i \leq 0, \forall i = 1, \ldots, n$. Since M is symmetric and its eigenvalues are just $-\varphi$ then M must be positive semi-definite, with exactly one eigenvalue at 0.

To show that the corresponding eigenvector of the 0 eigenvalue is $\mathbf{1}$, it suffices to observe that $M_t\mathbf{1} = 0$. □

A.4

Additional remarks and identities for the analytic solution of MPED.

Remark 9. *It has been observed, but remains to be proved that rank$(A_t(W_t)) = m_t + 1$. It is clear that rank$(K_t) = m_t$, and it can be shown that rank$(R_t) = m_t$, using rank identities of full column and row rank matrix products. It is still unclear how to prove rank$(A_t(W_t)) = m_t + 1$. Weyl's inequality could be used to show that the eigenvalues don't interfere with each other, i.e. that the $(m_t + 1)$th eigenvalue is > 0.*

Additionally we observe that $A_t(W_t) = Q_t P_t Q_t + Q_t$, and that $Q_t^{-1} A_t(W_t) = P_t Q_t + \mathbf{I}$, where the second term is the affine gain matrix for the generation optimizers. It has been observed that $\lambda(P_t Q_t + \mathbf{I}) = \{1, \ldots, 1, 0, \ldots, 0\}$, where the unit eigenvalue has multiplicity $m_t + 1$, and rank$(P_t Q_t + \mathbf{I}) = m_t + 1$. Clearly, rank$(P_t Q_t + \mathbf{I}) =$ rank$(A_t(W_t))$, by Sylvester's inequality, but it is interesting that the matrix has unit eigenvalues. We also note that $\lambda(P_t Q_t) = \{-1, \ldots, -1, 0, \ldots, 0\}$, where the negative unit eigenvalue has multiplicity m_t.

The following identites are stated:

$$A_t(W_t)Q_t^{-1}A_t(W_t) = A_t(W_t), \tag{A.22}$$
$$B_t^{\top}(W_t)Q_t^{-1}B_t(W_t) = K_t^{-1}, \tag{A.23}$$
$$A_t(W_t)Q_t^{-1}B_t(W_t) = B_t(W_t), \tag{A.24}$$
$$A_t(W_t)Q_t^{-1}\overline{\lambda}_t(W_t) = \overline{\lambda}_t(W_t) \tag{A.25}$$
$$A_t(W_t)M_tH_t^{\top}K_t^{-1} = B_t(W_t) \tag{A.26}$$
$$M_tQ_tM_t = M_t, \tag{A.27}$$
$$R_tM_tR_t = R_t, \tag{A.28}$$
$$Q_tP_t(W_t)Q_t + Q_t = A_t(W_t), \tag{A.29}$$
$$Q_t^{-1}A_t(W_t)Q_t^{-1} - Q_t^{-1} = P_t(W_t), \tag{A.30}$$
$$R_tP_t(W_t)R_t = 0, \tag{A.31}$$
$$Q_tM_tR_tM_tQ_tM_t = A_t(W_t)M_t, \tag{A.32}$$
$$Q_tM_tR_tM_t = A_t(W_t)M_t, \tag{A.33}$$
$$H_tP_t(W_t) = 0, \tag{A.34}$$
$$P_t(W_t)A_t(W_t)Q_t^{-1} = 0, \tag{A.35}$$
$$P_t(W_t)B_t(W_t) = 0, \tag{A.36}$$
$$P_t(W_t)\overline{\lambda}_t(W_t) = 0, \tag{A.37}$$
$$H_tM_tR_t = H_t \tag{A.38}$$

Bibliography

[1] IPCC, "Climate change 2014: Mitigation of climate change," tech. rep., 2014.

[2] CAISO, "What the duck curve tells us about managing a green grid," tech. rep., 2013.

[3] IPCC, "Global warming of 1.5°c," tech. rep., 2018.

[4] IEA, "World energy outlook 2017," tech. rep., 2017.

[5] Council GridWise Architecture, "GridWise Transactive Energy Framework Version 1.0," pp. 1–43, 2013.

[6] "Inside the cunning, unprecedented hack of ukraine's power grid." `https://www.wired.com/2016/03/inside-cunning-unprecedented-hack-ukraines-power-grid/`. Accessed: 2018-10-11.

[7] D. Chen, D. Irwin, P. Shenoy, and J. Albrecht, "Combined heat and privacy: Preventing occupancy detection from smart meters," in *2014 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp. 208–215, March 2014.

[8] "What is a self consumption community." `https://www.younergy.ch/en/self-consumption/`. Accessed: 2018-09-05.

[9] I. N. England, "Iso new england manual for market operations," tech. rep., 2018.

[10] D. Bertsimas, E. Litvinov, X. A. Sun, J. Zhao, and T. Zheng, "Adaptive Robust Optimization for the Security Constrained Unit Commitment Problem," *IEEE Transactions on Power Systems*, vol. 28, pp. 1–8, 2013.

[11] A. Lorca and X. A. Sun, "Adaptive Robust Optimization With Dynamic Uncertainty Sets for Multi-Period Economic Dispatch," vol. 30, no. 4, pp. 1702–1713, 2015.

[12] K. Hreinsson, M. Vrakopoulou, and G. Andersson, "Electrical Power and Energy Systems Stochastic security constrained unit commitment and non-spinning reserve allocation with performance guarantees," *International Journal of Electrical Power and Energy Systems*, vol. 72, pp. 1–7, 2015.

[13] P. P. Varaiya, F. F. Wu, and J. W. Bialek, "Smart Operation of Smart Grid: Risk-Limiting Dispatch," *Proceedings of the IEEE*, vol. 99, no. 1, pp. 40–57, 2011.

[14] Energy and Environmental Economics (E3), "Investigating a Higher Renewables Portfolio Standard in California," Tech. Rep. January, Energy and Environmental Economics, Inc., San Francisco, CA, 2014.

[15] "First solar proves that pv plants can rival frequency response services from natural gas peakers." `https://www.greentechmedia.com/articles/read/pv-plants-can-rival-frequency-response-services-from-natural-gas-peakers#gs.U9yvxQk`. Accessed: 2018-10-12.

[16] G. Research, "U.s. energy storage monitor q3 2018," tech. rep., 2018.

[17] S. E. P. Alliance, "2018 utility energy storage market snapshot," tech. rep., 2018.

[18] "Pg and e proposes world's biggest batteries to replace south bay gas plants." `https://www.greentechmedia.com/articles/read/pge-proposes-worlds-biggest-batteries-to-replace-south-bay-gas-plants`. Accessed: 2018-10-16.

[19] J. T. Hughes, A. D. Dominguez-Garcia, and K. Poolla, "Identification of virtual battery models for flexible loads," *IEEE Transactions on Power Systems*, vol. 31, pp. 4660–4669, Nov 2016.

[20] D. Pudjianto, D. Pudjianto, C. Ramsay, C. Ramsay, G. Strbac, and G. Strbac, "Virtual power plant and system integration of distributed energy resources," *Renewable Power Generation, IET*, vol. 1, no. 1, pp. 10–16, 2007.

[21] "A snapshot of the us gigawatt-scale non-wires alternatives market." `https://www.greentechmedia.com/articles/read/gtm-research-non-wires-alternatives-market`. Accessed: 2018-10-16.

[22] F. Dorfler, J. W. Simpson-Porco, and F. Bullo, "Breaking the hierarchy: Distributed control and economic optimality in microgrids," *IEEE Transactions on Control of Network Systems*, vol. 3, pp. 241–253, Sept 2016.

[23] "Ohmconnect." `https://www.ohmconnect.com/`. Accessed: 2018-10-18.

[24] "Pg and e to plug enphase smart inverters and solarcity storage systems into its grid control platform." `https://www.greentechmedia.com/articles/read/pge-to-plug-enphase-smart-inverters-solarcity-storage-systems-into-new-derm`. Accessed: 2018-10-16.

[25] S. Bose, D. F. Gayme, U. Topcu, and K. M. Chandy, "Optimal Placement of Energy Storage in the Grid," *IEEE Conference on Decision and Control*, pp. 5605–5612, 2012.

[26] A. Von Meier, *Electric power systems: a conceptual introduction.* John Wiley & Sons, 2006.

[27] M. E. Baran and F. F. Wu, "Optimal capacitor placement on radial distribution systems.," *IEEE Transactions on Power Delivery*, vol. 4, no. 1, pp. 725–734, 1989.

[28] M. E. Baran and F. F. Wu, "Optimal Sizing of Capacitors Placed on a Radial Distribution System," *IEEE Transactions on Power Delivery*, vol. 4, no. 1, pp. 735–743, 1989.

[29] N. Li, "A Market Mechanism for Electric Distribution Networks," *Conference on Decision and Control (CDC)*, no. Cdc, pp. 2276–2282, 2015.

[30] M. Farivar and S. Low, "Branch Flow Model: Relaxations and Convexification #x2014;Part I," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2554–2564, 2013.

[31] S. H. Low, "Convex Relaxation of Optimal Power Flow, Part II: Exactness," vol. 2014, no. June, pp. 1–51, 2014.

[32] M. Cain, R. O'Neill, and A. Castillo, "History of Optimal Power Flow and Formulations," *Federal Energy Regulatory Commission. . . .*, no. December, pp. 1–36, 2012.

[33] D. Munoz-Alvarez and E. Bitar, "Financial Storage Rights in Electric Power Networks," p. 11, 2014.

[34] J. Qin, I. Yang, and R. Rajagopal, "Submodularity of energy storage placement in power networks," *2016 IEEE 55th Conference on Decision and Control (CDC)*, no. Cdc, pp. 686–693, 2016.

[35] F. Schweppe and T.-R. B. R. Caramanis, M.C, *Spot Pricing of Electricity.* Boston, MA: Kluwer, 1988.

[36] T. Orfanogianni and G. Gross, "A general formulation for LMP evaluation," *IEEE Transactions on Power Systems*, vol. 22, no. 3, pp. 1163–1173, 2007.

[37] F. Wu, P. Varaiya, P. Spiller, and S. Oren, "Folk theorems on transmission access: Proofs and counterexamples," *Journal of Regulatory Economics*, vol. 10, pp. 5–23, Jul 1996.

[38] F. Borrelli, A. Bemporad, and M. Morari, *Predictive control for linear and hybrid systems.* Cambridge University Press, 2017.

[39] J. Mather, E. Bitar, and K. Poolla, "Virtual Bidding: Equilibrium, Learning, and the Wisdom of Crowds," *IFAC-PapersOnLine*, vol. 50, no. 1, 2017.

[40] W. W. Hogan, "Virtual bidding and electricity market design," *Electricity Journal*, vol. 29, no. 5, pp. 33–47, 2016.

[41] M. Celebi, A. Hajos, and P. Q. Hanser, "Virtual bidding: The good, the bad and the ugly," *Electricity Journal*, vol. 23, no. 5, pp. 16–25, 2010.

[42] A. G. Isemonger, "The Benefits and Risks of Virtual Bidding in Multi-Settlement Markets," *Electricity Journal*, vol. 19, no. 9, pp. 26–36, 2006.

[43] S. D. Ledgerwood and J. P. Pfeifenberger, "Using virtual bids to manipulate the value of financial transmission rights," *Electricity Journal*, vol. 26, no. 9, pp. 9–25, 2013.

[44] J. E. Parsons, C. Colbert, J. Larrieu, T. Martin, and E. Mastrangelo, "Financial Arbitrage and Efficient Dispatch in Wholesale Electricity Markets Erin Mastrangelo Financial Arbitrage and Efficient Dispatch in Wholesale Electricity Markets," no. February, 2015.

[45] C. Saravia, "Speculative Trading and Market Performance : The Effect of Arbitrageurs on Efficiency and Market Power in the New York Electricity Market," no. November, 2003.

[46] S. Borenstein, J. Bushnell, C. R. Knittel, and C. Wolfram, "Inefficiencies and Market Power in Financial Arbitrage: A Study of California's Electricity Markets," *The Journal of Industrial Economics*, vol. 56, no. 2, pp. 347–378, 2008.

[47] R. Li, A. J. Svoboda, and S. S. Oren, "Efficiency impact of convergence bidding in the california electricity market," *Journal of Regulatory Economics*, vol. 48, no. 3, pp. 245–284, 2015.

[48] A. Jha and F. Wolak, "Testing for Market Efficiency with Transactions Costs : An Application to Convergence Bidding in Wholesale Electricity Markets," 2015.

[49] D. Munoz-Alvarez, E. Bitar, L. Tong, and J. Wang, "Piecewise Affine Dispatch Policies for Economic Dispatch under Uncertainty," pp. 4–8, 2014.

[50] W. Tang, R. Rajagopal, K. Poolla, and P. Varaiya, "Model and Data Analysis of Two-Settlement Electricity Market with Virtual Bidding," *Proceedings of the IEEE Conference on Decision and Control*, 2016.

[51] R. Baldick, R. Grant, and E. Kahn, "Theory and application of linear supply function equilibrium in electricity markets," *Journal of Regulatory Economics*, vol. 25, no. 2, pp. 143–167, 2004.

[52] F. Galton, "Vox Populi," *Nature*, vol. 75, pp. 450–451, 1907.

[53] J. Surowiecki, *The Wisdom of Crowds: Why the Many are Smarter Than the Few and How Collective Wisdom Shapes Business Economies Societies and Nations*. New York Doubleday, 2004.

[54] D. Fudenberg and D. Levine, "Learning and Evolution: Where Do We Stand? Learning in games," *European Economic Review*, vol. 42, pp. 631–639, 1998.

[55] E. Hopkins, "A Note on Best Response Dynamics," *Games and Economic Behavior*, vol. 29, no. 1-2, pp. 138–150, 1999.

[56] J. Mather and E. Munsing, "Robust cournot-bertrand equilibria on power networks," in *American Control Conference (ACC), 2017*, pp. 2747–2754, IEEE, 2017.

[57] S. Borenstein, J. Bushnell, and F. Wolak, "Measuring Market Inefficiencies in California's Restructured Wholesale Electricity Market," *Power*, 2000.

[58] A. Hortaçsu and S. L. Puller, "Understanding strategic bidding in multi-unit auctions: A case study of the Texas electricity spot market," *RAND Journal of Economics*, vol. 39, no. 1, pp. 86–114, 2008.

[59] A. Kannan, U. V. Shanbhag, and H. M. Kim, "Addressing supply-side risk in uncertain power markets: stochastic Nash models, scalable algorithms and error analysis," *Optimization Methods and Software*, vol. 28, no. 5, pp. 1095–1138, 2013.

[60] A. Kannan, U. V. Shanbhag, and H. M. Kim, "Strategic behavior in power markets under uncertainty," *Energy Systems*, vol. 2, no. 2, pp. 115–141, 2011.

[61] D. Zhang, H. Xu, and Y. Wu, "A two stage stochastic equilibrium model for electricity markets with two way contracts," *Mathematical Methods of Operations Research*, vol. 71, no. 1, pp. 1–45, 2010.

[62] S. a. Gabriel, A. J. Conejo, D. Fuller, B. Hobbs, and C. Ruiz, *Complementarity Modeling in Energy Markets*. Spring International Series in Operations Research & Management Science, 2010.

[63] W. Tang and R. Jain, "Game-theoretic analysis of the nodal pricing mechanism for electricity markets," *Proceedings of the IEEE Conference on Decision and Control*, pp. 562–567, 2013.

[64] C. J. Day, B. F. Hobbs, and J. S. Pang, "Oligopolistic competition in power networks: A conjectured supply function approach," *IEEE Transactions on Power Systems*, vol. 17, no. 3, pp. 597–607, 2002.

[65] M. Ventosa, Á. Ì. Baíllo, A. Ramos, and M. Rivier, "Electricity market modeling trends," *Energy Policy*, vol. 33, no. 7, pp. 897–913, 2005.

[66] B. F. Hobbs, C. Metzler, and J. S. Pang, "Strategic gaming analysis for electric power networks: An MPEC approach," *IEEE Transactions on Power Systems*, vol. 15, no. 2, pp. 638–645, 2000.

[67] J. Barquín and M. Vázquez, "Cournot equilibrium in power networks," *Working Paper, Submitted to IEEE Transactions on Power Systems*, no. April, pp. 1–8, 2005.

[68] K. Neuhoff, J. Barquin, M. G. Boots, A. Ehrenmann, B. F. Hobbs, F. A. M. Rijkers, and M. Vazquez, "Network-constrained Cournot models of liberalized electricity markets: the devil is in the details," *Energy Economics*, vol. 27, no. 3, pp. 495–525, 2005.

[69] K. H. Lee, "Strategy equilibrium in Stackelberg model with transmission congestion in electricity market," *Journal of Electrical Engineering and Technology*, vol. 9, no. 1, pp. 90–97, 2014.

[70] J. Yao, S. S. Oren, and B. F. Hobbs, "Hybrid Bertrand-Cournot Models of Electricity Markets with Multiple Strategic Subnetworks and Common Knowledge Constraints," *Restructured Electric Power Systems: Analysis of Electricity Markets with Equilibrium Models*, pp. 167–192, 2010.

[71] O. L. Mangasarian and H. Stone, "Two-Person Nonzero-Sum Games and Quadratic Programming," *Journal of Mathematical Analysis and Applications*, vol. 9, no. 3, pp. 348–355, 1964.

[72] F. Wen and a. K. David, "Optimal bidding strategies for competitive generators and large consumers," *International Journal of Electrical Power and Energy System*, vol. 23, no. 1, pp. 37–43, 2001.

[73] J. Yao, S. S. Oren, and I. Adler, "Two-settlement electricity markets with price caps and Cournot generation firms," *European Journal of Operational Research*, vol. 181, no. 3, pp. 1279–1296, 2007.

[74] R. H. Kwon and D. Frances, "Optimization-Based Bidding in Day-Ahead Electricity Auction Markets: A Review of Models for Power Producers," in *Handbook of Networks in Power Systems I*, 2012.

[75] A. J. Conejo, M. Carrion, and J. M. Morales, *Decision Making Under Uncertainty in Electricity Markets*. New York, NY: Springer Science+Business Media, 2010.

[76] R. Rajagopal, E. Bitar, P. Varaiya, and F. Wu, "Risk-limiting dispatch for integrating renewable power," *International Journal of Electrical Power and Energy Systems*, vol. 44, no. 1, pp. 615–628, 2013.

[77] L. Wang, M. Mazumdar, M. D. Bailey, and J. Valenzuela, "Oligopoly models for market price of electricity under demand uncertainty and unit reliability," *European Journal of Operational Research*, vol. 181, no. 3, pp. 1309–1321, 2007.

[78] M. Dicorato, G. Forte, M. Trovato, and E. Caruso, "Risk-Constrained Profit Maximization in Day-Ahead Electricity Market," *IEEE Transactions on Power Systems on Power Systems*, vol. 24, no. 3, pp. 1107–1114, 2009.

[79] A. Baillo, M. Ventosa, M. Rivier, and A. Ramos, "Optimal offering strategies for generation companies operating in electricity spot markets," *IEEE Transactions on Power Systems*, vol. 19, no. 2, pp. 745–753, 2004.

[80] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust Optimization*. Princeton University Press, 2009.

[81] D. Bertsimas and M. Sim, "The Price of robustness," *Operations Research*, vol. 52, pp. 35–53, 2004.

[82] D. Bertsimas, D. B. D. Brown, and C. Caramanis, "Theory and Applications of Robust Optimization," *Operations Research*, p. 50, 2010.

[83] M. Aghassi and D. Bertsimas, "Robust game theory," *Mathematical Programming*, vol. 107, no. 1-2, pp. 231–273, 2006.

[84] R. Nishimura, S. Hayashi, and M. Fukushima, "Robust nash equilibria in N-person noncooperative games: Uniqueness and reformulation," *Pacific Journal of Optimization*, vol. 5, no. 2, pp. 237–259, 2009.

[85] A. Street, F. Oliveira, and J. M. Arroyo, "Contingency-Constrained Unit Commitment With Security Criterion: A Robust Optimization Approach," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1581–1590, 2011.

[86] R. Jiang, J. Wang, and Y. Guan, "Robust unit commitment with wind power and pumped storage hydro," *IEEE Transactions on Power Systems*, vol. 27, no. 2, pp. 800–810, 2012.

[87] T. Summers, J. Warrington, M. Morari, and J. Lygeros, "Stochastic optimal power flow based on convex approximations of chance constraints," *Proceedings - 2014 Power Systems Computation Conference, PSCC 2014*, vol. 72, pp. 116–125, 2014.

[88] L. Baringo and A. J. Conejo, "Offering strategy via robust optimization," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1418–1425, 2011.

[89] H. Haghighat, "Strategic offering under uncertainty in power markets," *International Journal of Electrical Power and Energy Systems*, vol. 63, pp. 1070–1077, 2014.

[90] C. Metzler, B. F. Hobbs, and J.-S. Pang, "Nash-Cournot Equilibria in Power Markets on a Linearized DC Network with Arbitrage: Formulations and Properties," *Networks and Spatial Economics*, vol. 3, pp. 123–150, 2003.

[91] B. F. Hobbs, E. Bartholomew, Y. Chen, G. Drayton, and W. Lise, "Improved transmission representations in oligopolistic market models," *2006 IEEE PES Power Systems Conference and Exposition*, pp. 81–86, 2006.

[92] J. Yao, I. Adler, and S. S. Oren, "Modeling and Computing Two-Settlement Oligopolistic Equilibrium in a Congested Electricity Network," *Operations Research*, vol. 56, no. 1, pp. 34–47, 2008.

[93] Y. Xie and U. V. Shanbhag, "On robust solutions to uncertain monotone linear complementarity problems (LCPs) and their variants," *53rd IEEE Conference on Decision and Control*, pp. 2834–2839, 2014.

[94] Y. Xie and U. V. Shanbhag, "On robust solutions to uncertain monotone linear complementarity problems and their variants," *Siam Journal of Optimization*, 2017.

[95] B. F. Hobbs, "Linear complementarity models of nash-Cournot competition in bilateral and POOLCO power markets," *IEEE Transactions on Power Systems*, vol. 16, no. 2, pp. 194–202, 2001.

[96] W. F. Samuelson and S. G. Marks, *Managerial Economics*. John Wiley & Sons, Inc., 7 ed., November 2011.

[97] H. Von Stackelberg, *Market structure and equilibrium*. Springer Science & Business Media, 2010.

[98] D. Pozo, E. Sauma, and J. Contreras, "Basic theoretical foundations and insights on bilevel models and their applications to power systems," *Annals of Operations Research*, no. March, pp. 1–32, 2017.

[99] B. Bahmani-Firouzi, S. Sharifinia, R. Azizipanah-Abarghooee, and T. Niknam, "Scenario-Based Optimal Bidding Strategies of GENCOs in the Incomplete Information Electricity Market Using a New Improved Prey-Predator Optimization Algorithm," *IEEE Systems Journal*, vol. 9, no. 4, pp. 1485–1495, 2015.

[100] V. P. Gountis and A. G. Bakirtzis, "Bidding Strategies for Electricity Producers in a Competitive Electricity Marketplace," *IEEE Transactions on Power Systems*, vol. 19, no. 1, pp. 356–365, 2004.

[101] N. K. Yadav, M. Kumar, D. Sharma, G. Bhargava, and A. Bala, "Profit calculation of strategic producer," no. Iccccm, 2016.

[102] Y. Ma, C. Jiang, Z. Hou, and C. Wang, "The Formulation of the Optimal Strategies for the Electricity Producers Based on the Particle Swarm Optimization Algorithm," *IEEE Transactions on Power Systems*, vol. 21, no. 4, pp. 1663–1671, 2006.

[103] A. K. Jain, S. C. Srivastava, S. N. Singh, and L. Srivastava, "Strategy Under Transmission Congestion," vol. 9, no. 1, pp. 141–151, 2015.

[104] T. Niknam, S. Sharifinia, and R. Azizipanah-Abarghooee, "A new enhanced bat-inspired algorithm for finding linear supply function equilibrium of GENCOs in the competitive electricity market," *Energy Conversion and Management*, vol. 76, pp. 1015–1028, 2013.

[105] M. Prabavathi and R. Gnanadass, "Energy bidding strategies for restructured electricity market," *International Journal of Electrical Power and Energy Systems*, vol. 64, pp. 956–966, 2015.

[106] P. D. Klemperer and M. A. Meyer, "Supply Function Equilibria in Oligopoly under Uncertainty," 1989.

[107] R. J. Green and D. M. Newbery, "Competition in the British Electricity Spot Market Author (s): Richard J . Green and David M . Newbery Published by : The University of Chicago Press Stable URL : http://www.jstor.org/stable/2138629 Accessed : 06-05-2016 12 : 42 UTC Your use of the JSTO," *Journal of Political Economy*, vol. 100, no. 5, pp. 929–953, 2016.

[108] R. Baldick and W. Hogan, "Capacity Constrained Supply Function Equilibrium Models of Electricity Markets : Stability , Non-decreasing constraints , and Function Space Iterations," *Power Working Paper*, 2001.

[109] F. Gao, G. B. Sheble, K. W. Hedman, and C.-N. Yu, "Optimal bidding strategy for gencos based on parametric linear programming considering incomplete information," *International Journal of Electrical Power & Energy Systems*, vol. 66, pp. 272–279, 2015.

[110] M. Ghamkhari, A. Sadeghi-Mobarakeh, and H. Mohsenian-Rad, "Strategic bidding for producers in nodal electricity markets: A convex relaxation approach," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 2324–2336, 2017.

[111] J.-S. Pang and M. Fukushima, "Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games," *Computational Management Science*, vol. 2, no. 1, pp. 21–56, 2005.

[112] M. Hu and M. Fukushima, "Multi-Leader-Follower Games: Models, Methods and Applications," *Journal of the Operations Research Society of Japan*, vol. 58, no. 1, pp. 1–23, 2015.

[113] X. Hu and D. Ralph, "Using EPECs to Model Bilevel Games in Restructured Electricity Markets with Locational Prices," *Operations Research*, vol. 55, no. 5, pp. 809–827, 2007.

[114] D. Pozo and J. Contreras, "Finding multiple Nash equilibria in pool-based markets: A stochastic EPEC approach," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1744–1752, 2011.

[115] J. Weber and T. Overbye, "A two-level optimization problem for analysis of market bidding strategies," *199 IEEE Power Engineering Society Summer Meeting. Conference Proceedings (Cat. No.99CH36364)*, vol. 2, pp. 682–687, 1999.

[116] I. Taheri, M. Rashidinejad, A. Badri, and A. Rahimi-Kian, "Analytical Approach in Computing Nash Equilibrium for Oligopolistic Competition of Transmission-Constrained GENCOs," *IEEE Systems Journal*, vol. 9, no. 4, pp. 1452–1462, 2015.

[117] S. A. Gabriel, A. J. Conejo, J. D. Fuller, B. F. Hobbs, and C. Ruiz, "Complementarity modeling in energy markets, ser. international series in operations research & management science," 2012.

[118] N. P. Faísca, P. M. Saraiva, B. Rustem, and E. N. Pistikopoulos, "A multi-parametric programming approach for multilevel hierarchical and decentralised optimisation problems," *Computational Management Science*, vol. 6, no. 4, pp. 377–397, 2009.

[119] S. Borenstein, J. Bushnell, and S. Stoft, "The competitive effects of transmission capacity in a deregulated electricity industry," tech. rep., National Bureau of Economic Research, 1997.

[120] C. Dutang, "Existence theorems for generalized Nash equilibrium problems: an analysis of assumptions," *Journal of Nonlinear Analysis and Optimization. Theory and Applications*, vol. 4, no. 2, pp. 115–126, 2013.

[121] H. Nikaidô, K. Isoda, *et al.*, "Note on non-cooperative convex games," *Pacific Journal of Mathematics*, vol. 5, no. Suppl. 1, pp. 807–815, 1955.

[122] T. Ichiishi and M. Quinzii, "Decentralization for the core of a production economy with increasing return," *International Economic Review*, pp. 397–412, 1983.

[123] P. Dasgupta and E. S. Maskin, "The Existence of Equilibrium in Discontinuous Economic Games, I: Theory," 1986.

[124] H. Dixon, "The Existence of Mixed-Strategy Equilibria in a Price-Setting Oligopoly with Convex Costs," *Economics Letters*, vol. 16, pp. 205–212, 1984.

[125] X. He, E. Delarue, W. D'haeseleer, and J.-M. Glachant, "A novel business model for aggregating the values of electricity storage," *Energy Policy*, vol. 39, no. 3, pp. 1575–1585, 2011.

[126] J. A. Taylor, "Financial Storage Rights," *IEEE Transactions on Power Systems*, vol. 30, no. 2, pp. 1–9, 2014.

[127] D. Muñoz-Álvarez and E. Bitar, "Financial storage rights in electric power networks," *Journal of Regulatory Economics*, vol. 52, no. 1, 2017.

[128] R. Sioshansi, "Using Storage-Capacity Rights to Overcome the Cost-Recovery Hurdle for Energy Storage," *IEEE Transactions on Power Systems*, vol. PP, no. 99, pp. 1–13, 2016.

[129] S. Bose, D. W. H. Cai, S. Low, and A. Wierman, "The role of a market maker in networked cournot competition," in *53rd IEEE Conference on Decision and Control*, pp. 4479–4484, IEEE, dec 2014.

[130] C. Thrampoulidis, S. Bose, and B. Hassibi, "Optimal Placement of Distributed Energy Storage in Power Networks," *IEEE Transactions on Automatic Control*, vol. 9286, no. c, pp. 1–1, 2015.

[131] H. Mohsenian-Rad, "Coordinated Price-Maker Operation of Large Energy Storage Units in Nodal Energy Markets," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 786–797, 2016.

[132] R. Sioshansi, P. Denholm, T. Jenkin, and J. Weiss, "Estimating the value of electricity storage in PJM: Arbitrage and some welfare effects," *Energy Economics*, vol. 31, no. 2, pp. 269–277, 2009.

[133] R. Sioshansi, "Welfare impacts of electricity storage and the implications of ownership structure," *Energy Journal*, vol. 31, no. 2, pp. 173–198, 2010.

[134] N. Gast, J.-Y. Le Boudec, A. Proutière, and D.-C. Tomozei, "Impact of storage on the efficiency and prices in real-time electricity markets," *Proceedings of the the fourth international conference on Future energy systems - e-Energy '13*, p. 15, 2013.

[135] R. Sioshansi, "When energy storage reduces social welfare," *Energy Economics*, vol. 41, pp. 106–116, 2014.

[136] I. Taheri, H. A. Abyaneh, S. H. Hosseinian, and A. Bakhshai, "Effects of LAs ' Strategic Bidding on Wholesale Electricity Market Equilibrium," no. September, pp. 1–12, 2017.

[137] K. Hartwig and I. Kockar, "Impact of Strategic Behavior and Ownership of Energy Storage on Provision of Flexibility," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 2, pp. 744–754, 2016.

[138] J. Qin, J. Mather, J.-Y. Joo, R. Rajagopal, K. Poolla, and P. Varaiya, "Automatic power exchange for distributed energy resource networks: Scheduling and pricing," *Decision and Control (CDC)*, 2018.

[139] A. Shahmohammadi, R. Sioshansi, and A. J. Conejo, "Market Equilibria and Interactions Between Strategic Generation , Wind, and Storage," pp. 1–8, 2016.

[140] J. Mather, E. Baeyens, K. Poolla, and P. Varaiya, "The real value of load flexibility-congestion free dispatch," in *Proceedings of the American Control Conference*, vol. 2015-July, 2015.

[141] U. D. of Energy, "National electric transmission congestion study," tech. rep.

[142] M. Analytics, "2013 state of the market report for pjm," tech. rep., 2014.

[143] J. Rosellon and T. Kristiansen, *Financial Transmission Rights*. Springer.

[144] R. Starr, "Quasi-equilibria in markets with non-convex preferences," *Econometrica: Journal of the Econometric Society*, pp. 25–38, 1969.

[145] E. Munsing, J. Mather, and S. Moura, "Blockchains for decentralized optimization of energy resources in microgrid networks," in *1st Annual IEEE Conference on Control Technology and Applications, CCTA 2017*, vol. 2017-Janua, 2017.

[146] T. Chen, Q. Alsafasfeh, H. Pourbabak, and W. Su, "The next-generation U.S. retail electricity market with customers and prosumers-A bibliographical survey," *Energies*, vol. 11, no. 1, 2018.

[147] New York State Energy Planning Board, "The Energy to Lead: New York State Energy Plan," tech. rep., New York State Energy Planning Board, 2015.

[148] North Carolina Clean Energy Technology Center, "Net Metering Policy Map," *DSIRE*, no. July, 2016.

[149] F. Flores-espino, "Compensation for Distributed Solar : A Survey of Options to Preserve Stakeholder Value," Tech. Rep. NREL/TP-6A20-62371, National Renewable Energy Laboratory, 2015.

[150] G. Meyers, "Black Hole Forms On Solar Net Metering In Nevada," jan 2016.

[151] S. Chen and C.-c. Liu, "From Demand Response to Transactive Energy: the State-of-the-Art," *J. Mod. Power Syst. Clean Energy*, 2016.

[152] S. Borenstein, J. Bushnell, and F. Wolak, "Measuring Market Ineffieciencies in California's Restructured Wholesale Electricity Market," *Csem Wp*, vol. 102, no. June, pp. 1–58, 2002.

[153] F. Wolak, "Lessons from the California Electricity Crisis," in *Electricity deregulation: choices and challenges*, ch. 4, 2005.

[154] California Public Utilities Commission, "Microgrids : A Regulatory Perspective," tech. rep., 2014.

[155] L. I. Minchala-Avila, L. E. Garza-Castañón, A. Vargas-Martínez, and Y. Zhang, "A review of optimal control techniques applied to the energy management and control of microgrids," *Procedia Computer Science*, vol. 52, no. 1, pp. 780–787, 2015.

[156] A. Killeen, *Handbook of Digital Currency*. 2015.

[157] Deutsche Energie-Agentur GmbH (dena), "Blockchain in the energy transition: A survey among decision-makers in the German energy industry," tech. rep., 2016.

[158] J. Bonneau, A. Miller, J. Clark, A. Narayanan, J. A. Kroll, and E. W. Felten, "SoK: Research perspectives and challenges for bitcoin and cryptocurrencies," *Proceedings - IEEE Symposium on Security and Privacy*, vol. 2015-July, pp. 104–121, 2015.

[159] G. Wood, "Ethereum: a secure decentralised generalised transaction ledger," *Ethereum Project Yellow Paper*, pp. 1–32, 2014.

[160] K. Christidis and M. Devetsikiotis, "Blockchains and Smart Contracts for the Internet of Things," *IEEE Access*, vol. 4, pp. 2292–2303, 2016.

[161] D. T. Ton and M. A. Smith, "The U.S. Department of Energy's Microgrid Initiative," *The electricity Journal*, vol. 25, no. 8, pp. 84–94, 2012.

[162] A. Ahmad Khan, M. Naeem, M. Iqbal, S. Qaisar, and A. Anpalagan, "A compendium of optimization objectives, constraints, tools and algorithms for energy management in microgrids," *Renewable and Sustainable Energy Reviews*, vol. 58, pp. 1664–1683, 2016.

[163] C. Le Floch, F. Belletti, and S. Moura, "Optimal Charging of Electric Vehicles for Load Shaping : a Dual Splitting Framework with Explicit Convergence Bounds," *IEEE Transactions on Transportation Electrification*, vol. 7782, pp. 1–9, 2016.

[164] C. Le Floch, F. Belletti, S. Saxena, A. M. Bayen, and S. Moura, "Distributed optimal charging of electric vehicles for demand response and load shaping," *Proceedings of the IEEE Conference on Decision and Control*, vol. 2016-Febru, no. Cdc, pp. 6570–6576, 2016.

[165] O. Sondermeijer, R. Dobbe, D. Arnold, and C. Tomlin, "Regression-based Inverter Control for Decentralized Optimal Power Flow and Voltage Regulation," in *Power & Energy Society General Meeting. IEEE*, 2016.

[166] S. Mhanna, A. C. Chapman, and G. Verbic, "A Fast Distributed Algorithm for Large-Scale Demand Response Aggregation," *IEEE Transactions on Smart Grid*, vol. 7, no. 4, pp. 2094–2107, 2016.

[167] S. C. Tsai, Y. H. Tseng, and T. H. Chang, "Communication-efficient distributed demand response: A randomized ADMM approach," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–11, 2015.

[168] Y. Wang, L. Wu, and S. Wang, "A Fully-Decentralized Consensus-Based ADMM Approach for DC-OPF With Demand Response," *IEEE Transactions on Smart Grid*, pp. 1–11, 2016.

[169] E. Dall'Anese, H. Zhu, and G. B. Giannakis, "Distributed Optimal Power Flow for Smart Microgrids," *IEEE Transactions on Smart Grid*, vol. 4, no. 3, pp. 1464–1475, 2013.

[170] Q. Peng and S. H. Low, "Distributed Algorithm for Optimal Power Flow on an Unbalanced Radial Network," *53rd IEEE Conference on Decision and Control*, no. Cdc, pp. 0–7, 2015.

[171] P. Sulc, S. Backhaus, and M. Chertkov, "Optimal Distributed Control of Reactive Power Via the Alternating Direction Method of Multipliers," *IEEE Transactions on Energy Conversion*, vol. 29, no. 4, pp. 968–977, 2014.

[172] S. Borenstein and J. Bushnell, "The U.S. Electricity Industry after 20 Years of Restructuring," *Annual Review of Economics*, vol. 7.1, pp. 437–463, 2015.

[173] R. Hirsh, *Power Loss: The Origins of Deregulation and Restructuring in the American Electric Utility System*. Cambridge, MA: MIT Press, 2000.

[174] J. D. Hertog, "Review of economic theories of regulation," *Tjalling C. Koopmans Institute Discussion Paper Series*, vol. 10, no. 18, pp. 1–59, 2010.

[175] M. Conoscenti, A. Vetro, and J. C. De Martin, "Blockchain for the Internet of Things: a Systematic Literature Review," no. November, 2016.

[176] J. Mattila, T. Seppala, C. Naucler, R. Stahl, M. Tikkanen, A. Badenlid, and J. Seppala, "Industrial Blockchain Platforms : An Exercise in Use Case Development in the Energy Industry," tech. rep., ETLA, 2016.

[177] J. Horta, D. Kofman, and D. Menga, "Novel paradigms for advanced distribution grid energy management," 2016.

[178] M. Mihaylov, I. Razo-Zapata, R. Radulescu, S. Jurado, N. Avellana, and A. Nowé, "Smart grid demonstration platform for renewable energy exchange," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9662, pp. 277–280, 2016.

[179] L. P. Johnson, A. Isam, N. Gogerty, and J. Zitoli, "Connecting the Blockchain to the Sun to Save the Planet," dec 2015.

[180] Microgrid Media, ""It's Like the Early Days of the Internet," Blockchain-based Microgrid Tests P2P Energy Trading in Brooklyn," mar 2016.

[181] E. M. Burger and S. J. Moura, "Generation following with thermostatically controlled loads via alternating direction method of multipliers sharing algorithm," *Electric Power Systems Research*, vol. 146, pp. 141–160, 2017.

[182] S. Boyd and L. Vandenberghe, *Convex Optimization*, vol. 25. Cambridge University Press, 2004.

[183] P. Scott and S. Thiébaux, "Dynamic Optimal Power Flow in Microgrids using the Alternating Direction Method of Multipliers," *CoRR*, pp. 1–8, 2014.

[184] Q. Peng and S. Low, "Optimal Branch Exchange for Distribution System Reconfiguration," *Arxiv Preprint*, p. 12, 2013.

[185] Ethereum Foundation, "Ethereum Frontier Guide," 2016.

[186] S. Diamond and S. Boyd, "{CVXPY}: A {P}ython-Embedded Modeling Language for Convex Optimization," *Journal of Machine Learning Research*, vol. 17, no. 83, pp. 1–5, 2016.

[187] S. Barber, X. Boyen, E. Shi, and E. Uzun, "Bitter to Better - How to Make Bitcoin a Better Currency,"

[188] North American Electric Reliability Corporation, "Distributed energy resources: Connection modeling and reliability considerations," tech. rep., 2017.

[189] P. P. Varaiya, F. F. Wu, and J. W. Bialek, "Smart operation of smart grid: Risk-limiting dispatch," *Proceedings of the IEEE*, vol. 99, pp. 40–57, Jan 2011.

[190] S. Borenstein, "Customer risk from real-time retail electricity pricing: Bill volatility and hedgability," tech. rep., National Bureau of Economic Research, 2006.

[191] K. Alshehri, S. Bose, and T. Başar, "Cash-settled options for wholesale electricity markets," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 13605–13611, 2017.

[192] A. Radovanović, A. Ramesh, R. Koningstein, D. Fork, W. Weber, S. Kim, J. Schmalzried, J. Sastry, M. Dikovsky, K. Bozhkov, *et al.*, "Powernet for distributed energy resource networks," in *Power and Energy Society General Meeting (PESGM), 2016*, pp. 1–5, IEEE, 2016.

[193] A. I. Dounis and C. Caraiscos, "Advanced control systems engineering for energy and comfort management in a building environment - a review," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 6-7, pp. 1246–1261, 2009.

[194] F. Wu and P. Varaiya, "Coordinated multilateral trades for electric power networks: theory and implementation," *International Journal of Electrical Power & Energy Systems*, vol. 21, no. 2, pp. 75 – 102, 1999.

[195] J. Qin, R. Rajagopal, and P. P. Varaiya, "Flexible market for smart grid: Coordinated trading of contingent contracts," *IEEE Transactions on Control of Network Systems*, vol. PP, no. 99, pp. 1–1, 2017.

[196] R. Rajagopal, E. Bitar, F. F. Wu, and P. Varaiya, "Risk-Limiting Dispatch for Integrating Renewable Power," *International Journal of Electrical Power and Energy Systems, to appear*, 2012.

[197] J. Qin and R. Rajagopal, "Price of uncertainty in multistage stochastic power dispatch," in *53rd IEEE Conference on Decision and Control*, pp. 4065–4070, Dec 2014.

[198] Grinblatt and Titman, *Financial markets and corporate strategy*. McGraw Hill, 2004.

[199] C.-W. Tan and P. Varaiya, "Interruptible electric power service contracts," *Journal of Economic Dynamics and Control*, vol. 17, no. 3, pp. 495–517, 1993.

[200] E. Bitar and S. Low, "Deadline differentiated pricing of deferrable electric power service," in *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, pp. 4991–4997, Dec 2012.

[201] A. Nayyar, M. Negrete-Pincetic, K. Poolla, and P. Varaiya, "Duration-differentiated energy services with a continuum of loads," *IEEE Transactions on Control of Network Systems*, vol. 3, no. 2, pp. 182–191, 2016.

[202] S. Borenstein, "Wealth transfers among large customers from implementing real-time retail electricity pricing," *The Energy Journal*, pp. 131–149, 2007.

[203] S. Borenstein, "The long-run efficiency of real-time electricity pricing," *The Energy Journal*, pp. 93–116, 2005.

[204] A. W. Berger and F. C. Schweppe, "Real time pricing to assist in load frequency control," *IEEE Transactions on Power Systems*, vol. 4, no. 3, pp. 920–926, 1989.

[205] P. L. Joskow and C. D. Wolfram, "Dynamic pricing of electricity," *American Economic Review*, vol. 102, pp. 381–85, May 2012.

[206] D. J. Hammerstrom, R. Ambrosio, T. A. Carlon, *et al.*, "Pacific Northwest GridWise Testbed Demonstration Projects; Part I. Olympic Peninsula Project," tech. rep., PNNL, 2008.

[207] E. G. Cazalet, "Automated transactive energy (TEMIX)," in *Grid-Interop Forum*, 2011.

[208] A. Subramanian, M. J. Garcia, D. S. Callaway, K. Poolla, and P. Varaiya, "Real-time scheduling of distributed resources," *IEEE Transactions on Smart Grid*, vol. 4, no. 4, pp. 2122–2130, 2013.

[209] G. O'Brien and R. Rajagopal, "Scheduling non-preemptive deferrable loads," *IEEE Transactions on Power Systems*, vol. 31, pp. 835–845, March 2016.

[210] A. Gupta, R. Jain, and R. Rajagopal, "Scheduling, pricing, and efficiency of non-preemptive flexible loads under direct load control," in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 1008–1015, Sept 2015.

[211] D. Bertsimas, D. Gamarnik, and J. Sethuraman, "From fluid relaxations to practical algorithms for high-multiplicity job-shop scheduling: The holding cost objective," *Operations Research*, vol. 51, no. 5, pp. 798–813, 2003.

[212] J. Qin, I. Yang, and R. Rajagopal, "Submodularity of energy storage placement in power networks," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 686–693, Dec 2016.

[213] M. Sviridenko, "A note on maximizing a submodular set function subject to a knapsack constraint," *Operations Research Letters*, vol. 32, no. 1, pp. 41 – 43, 2004.

[214] J. Qin, J. Mather, J.-Y. Joo, K. Poolla, and P. Varaiya, "Automatic power exchange for distributed energy resource networks: Flexibility scheduling and pricing,"

[215] W. Tang, J. Qin, R. Jain, and R. Rajagopal, "Pricing sequential forward power contracts," in *Smart Grid Communications (SmartGridComm), 2015 IEEE International Conference on*, pp. 563–568, IEEE, 2015.

[216] H.-P. Chao and S. Peck, "A market mechanism for electric power transmission," *Journal of Regulatory Economics*, vol. 10, no. 1, pp. 25–59, 1996.

[217] R. A. Horn and C. R. Johnson, "Matrix analysis," *BOOK*, p. 607, 2013.

[218] N. Jakovčević Stor, I. Slapničar, and J. L. Barlow, "Forward stable eigenvalue decomposition of rank-one modifications of diagonal matrices," *Linear Algebra and Its Applications*, vol. 487, pp. 301–315, 2015.