Resource Management for Sustainable Power Grids and Wireless Networks: Distributed and Robust Designs

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Yu Zhang

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

Georgios B. Giannakis, Advisor

July, 2015

Dedication

To my beloved parents, for their unconditional love, support, and encouragement.

Acknowledgements

The winding road to a Ph.D. is full of obstacles, challenges, joy, and happiness. The dissertation completion marks the end of such an eventful journey for which there are many people I would like to acknowledge. Their kind support made the Ph.D. experience comparatively less arduous, and inspired me to be superior to my former self.

First and foremost, I would like to express my deepest gratitude and respect to my doctoral advisor Professor Georgios B. Giannakis for his invaluable guidance. Thanks to his incisive foresight and suggestion, I was strongly convinced to focus on the areas of the smart grid and renewable energy, which constitute the main threads of this dissertation. By generously offering excellent advice and enthusiastic encouragement, he has dedicated large amounts of time into motivating and training me to mature as a researcher with improved technical writing and presentation skills. From his critical thinking, brilliant insights, and cross-fertilization of ideas, I am lucky to have learnt quite a lot of things, not only useful for scientific research, but also for my career and life. He is a paragon of persistent enthusiasm and immeasurable energy, who serves as the best example for me by showing that "the only way to do great work is to love what you do." One simply could not wish for a better advisor and a lifelong friend.

Special thanks go to Professors Prodromos Daoutidis, Sairaj Dhople, Mostafa Kaveh, and Fotis Sotiropoulos for serving on the committees of my preliminary and final oral exams, as well as their valuable suggestions for revision.

The work in this dissertation would not have been possible without the help of the colleagues I have collaborated with during the last few years. I wish to express due credit and my warmest thanks to Drs. Emiliano Dall'Anese, Nikos Gatsis, Vassilis Kekatos, and Xin Wang for their significant contributions to our fruitful collaborations. The material here has also benefited from discussions with current, former members and

visiting scholars of the SPiNCOM group: Brian Baingana, Dr. Juan-Andrés Bazerque, Dimitris Berberidis, Dr. Alfonso Cano, Tianyi Chen, Sundeep Chepuri, Dr. Shahrokh Farahmand, Dr. Pedro Forero, Nitin Jain, Georgios Karanikolas, Dr. Seung-Jun Kim, Donghoon Lee, Prof. Geert Leus, Dr. Qing Ling, Dr. Morteza Mardani, Prof. Antonio Marques, Dr. Gonzalo Mateos, Dr. Eric Msechu, Dr. Ketan Rajawat, Dr. Ioannis Schizas, Fatemeh Sheikholeslami, Yanning Shen, Dr. Konstantinos Slavakis, Dr. Nasim Yahya Soltani, Panagiotis Traganitis, Gang Wang, Liang Zhang, and Dr. Hao Zhu. I am truly grateful to these people for their continuous help.

The beneficial research experience in Professor Zhi-Quan Luo's group is still vividly remembered during my initial stages at the University of Minnesota (UMN). My heartfelt gratitude goes to him for offering me the opportunity, as well as his group members: Drs. Shu Cai, Mingyi Hong, Qiang Li, Yafeng Liu, Meisam Razaviyayn, Maziar Sanjabi, Qingjiang Shi, Enbin Song, Ruoyu Sun, Hung-Wei Tseng, and Xiangfeng Wang.

I would like to take this opportunity to thank my colleagues at the ABB US Corporate Research Center, especially Drs. Mirrasoul Mousavi and James Stoupis, for the inspiring discussions during my 2014 summer internship there. My sincere gratitude also goes to many other friends along the way, including David Anastasiu, Dr. Alex Chen (UNC), Hongjian Fan, Dr. Huiyi Hu (UCLA), Kejun Huang, Swayambhoo Jain, Mojtaba Kadkhodaie, Wei-Cheng Liao, Hanchao Liu (RPI), Qi Shao, Dr. Yi Wang (Duke), Jiangmeng Zhang (UIUC), and Dr. Ren Zhu.

Finally, my appreciation is extended to the funding sources and awards that financially supported my Ph.D. research: the U.S. National Science Foundation, the Qatar National Research Fund, the UMN Initiative for Renewable Energy & the Environment Program, the UMN ECE Department Fellowship, and Student Travel Awards from the UMN ECE Department, the SIAM, and the IEEE Signal Processing Society.

Minneapolis, MN, July 2015 *Yu Zhang*

Abstract

Optimal management plays an indispensable role in judiciously allocating the surging demand of limited resources available to our modern society. Intelligent management schemes must be efficient, scalable, and even robust to the inherently uncertain and possibly adversarial nature. Leveraging state-of-the-art optimization and signal processing techniques, the present thesis addresses several fundamental issues and emerging challenges of cyber-physical systems, especially for the smart power grid and wireless networks.

Robust energy management is first dealt with for a grid-connected microgrid featuring distributed energy sources. To address the intrinsic challenge of maintaining the supply-demand balance due to stochastic availability of renewable energy sources (RES), a novel power scheduling strategy is introduced to minimize the microgrid operational cost including the worst-case energy transaction cost. The resulting optimization problem is solved in a distributed fashion by each local controller via the dual decomposition approach. In addition, for an islanded microgrid or the long-term planning of the bulk power system, risk-limiting energy management using the loss-of-load probability is developed. Day-ahead stochastic market clearing with high-penetration wind energy is further pursued based on the DC optimal power flow model. Capitalizing on the conditional value-at-risk, the novel model is able to mitigate the potentially high risk of the recourse actions to account for wind forecast errors. To cope with possibly large-scale dispatchable loads, fast distributed solvers are developed with guaranteed convergence.

This thesis also caters to distributed resource allocation in wireless networks. Robust transceiver design and energy scheduling are considered for multiple-input multiple-output cognitive radio networks, as well as smart-grid powered coordinated multipoint systems. Robust optimization problems are formulated to tackle the uncertainties from imperfect channel state information and the nondispatchable RES. Efficient distributed solvers are tailored to the resulting convex programs through the techniques of semidefinite relaxation, primal, and dual decomposition. Numerical results are reported to corroborate the merits of the novel framework, and assess performance of the proposed approaches.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Optimal resource management plays an indispensable role to catch up the surging demand in our resource-limited society. Resource allocation schemes must be efficient, scalable, reliable, and also resilient to inherent uncertainties involved in the systems of interest. Leveraging modern optimization and signal processing techniques, the present thesis contributes in several fundamental issues and emerging challenges of cyber-physical systems, including the smart power network and wireless communication networks.

The future "smart grid" is an automated electric power network that capitalizes on modern optimization, monitoring, communication, and control technologies to improve efficiency, sustainability, and reliability of generation, transmission, distribution, and consumption of electric energy. Limited supply and environmental impact of conventional power generation (e.g., oil, coal, and natural gas) raise major concerns worldwide, and compel industry to aggressively utilize the clean renewable energy sources (RES), including wind, sunlight, biomass, and geothermal heat, which are eco-friendly and price competitive. Growing at an annual rate of 20%, wind power generation already boasted a worldwide installed capacity of 369 GW by the end of 2014, and is widely embraced throughout the world [64]. Recently, both the U.S. Department of Energy and the European Union proposed ambitious blueprints towards a low-carbon economy by meeting 20% of the electricity consumption with renewables by 2030 and 2020, respectively [2,3]. High-penetration renewable energy clearly brings new variability and uncertainty to the power grid besides loads that change over time, and conventional resources which can

fail unexpectedly. Stochastic availability and intermittency of renewable energy make the power supply-demand balance harder to achieve, which must be accounted for by system operators during scheduling of generation, reserves, and controllable loads.

As regards the wireless communication networks, efficient allocation of resources (e.g., transmission power, frequency bands, and time slots) amounts to minimizing a system cost while satisfying certain quality of service metrics, and vice versa. For example, the system operator typically aims to minimize the transmission power, but guarantee that the network throughput is no less than a prescribed threshold. This type of tasks heavily depends on the available channel state information, which can not be perfectly known in practice. In addition, it is expected that the future communication systems will be powered by the smart grid, especially by RES. As a result, the full potential of "green communications" in next-generation heterogeneous wireless networks can only be exploited by uncertainty-aware resource allocation schemes.

The present thesis proposes practical frameworks and develops efficient algorithms for robust management of resources available to the smart grid (Chapters 2, 3, and 4) and the wireless cellular network (Chapter 5). The motivation, context, and contributions of this thesis are described in the ensuing sections.

1.1 Energy Management for Sustainable Microgrids

Microgrids are power systems comprising distributed energy resources (DERs) and electricity end-users, possibly with controllable (so-termed elastic) loads, all deployed across a limited geographical area [66]. Depending on their origin, DERs can come either from distributed generation (DG) or from distributed storage (DS). DG refers to small-scale power generators such as diesel generators, fuel cells, and RES, as in wind or photovoltaic (PV) generation. DS paradigms include batteries, flywheels, and pumped storage. Specifically, DG brings power closer to the point it is consumed, thereby incurring fewer thermal losses and bypassing limitations imposed by a congested transmission network. Moreover, the increasing tendency towards high penetration of RES stems from their environment-friendly and price-competitive advantages over conventional generation. Typical microgrid loads include critical non-dispatchable types and elastic controllable ones [57].

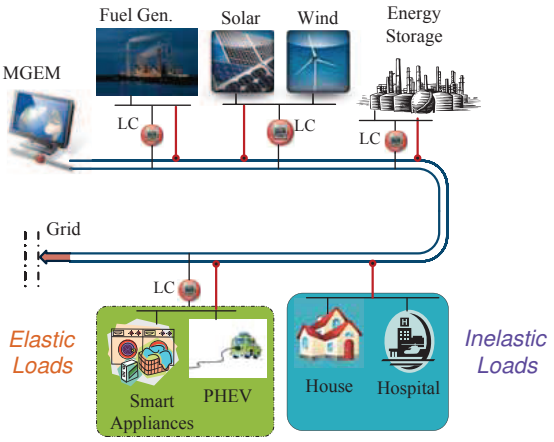


Figure 1.1: Distributed control and computation architecture of a microgrid featuring the microgrid energy manager (MGEM) coordinates the local controllers (LCs) of DERs and dispatchable loads.

Microgrids operate in grid-connected or island mode, and may entail distribution networks with residential or commercial end-users, in rural or urban areas. A typical configuration is depicted in Fig. 1.1; see also [66]. The microgrid energy manager (MGEM) coordinates the DERs and the controllable loads. Each of the DERs and loads has a local controller (LC), which coordinates with the MGEM the scheduling of resources through the communications infrastructure in a distributed fashion. The main challenge in energy scheduling is to account for the random and nondispatchable nature of the RES.

Optimal energy management for microgrids including economic dispatch (ED), unit commitment (UC), and demand-side management (DSM) is addressed in [129], but without pursuing a robust formulation against RES uncertainty. Based on the Weibull distribution for wind speed and the wind-speed-to-power-output mappings, an ED problem is formulated to minimize the risk of overestimation and underestimation of available wind power [67]. Stochastic programming is also used to cope with the variability of RES. Single-period chance-constrained ED problems for RES have been studied in [89], yielding probabilistic guarantees that the load will be served. Considering the uncertainties of demand profiles and PV generation, a stochastic program is formulated to minimize the overall cost of electricity and natural gas for a building in [62]. Without DSM, robust scheduling problems with penalty-based costs for uncertain supply and

demand have been investigated in [15]. Recent works explore energy scheduling with DSM and RES using only centralized algorithms [74, 171]. An energy source control and DS planning problem for a microgrid is formulated and solved using model predictive control in [75]. Using energy storage to mitigate fluctuation in generation due to time-varying RES, an optimal power flow problem is formulated in [31]. A two-level control scheme is developed for a renewable hybrid energy system in [136]. Distributed algorithms are developed in [46], but they only coordinate DERs to supply a given load without considering the stochastic nature of RES. In all the aforementioned works however, robust formulations accounting for the RES randomness are not pursued. Recently, a worst-case transaction cost based energy scheduling scheme has been proposed to address the variability of RESs through robust optimization that can also afford distributed implementation [162]. However, [162] considers only a single wind farm and no DS, and its approach cannot be readily extended to include multiple RESs and DS.

This thesis deals with optimal energy management for both supply and demand of a grid-connected microgrid incorporating RES. The objective of minimizing the microgrid net cost accounts for conventional DG cost, utility of elastic loads, penalized cost of DS, and a worst-case transaction cost. The latter stems from the ability of the microgrid to sell excess energy to the main grid, or to import energy in case of shortage. A *robust* formulation accounting for the worst-case amount of harvested RES is developed. A novel model is introduced in order to maintain the supply-demand balance arising from the intermittent RES. Moreover, a transaction-price-based condition is established to ensure convexity of the overall problem. The separable structure and strong duality of the resultant problem are leveraged to develop a low-overhead *distributed* algorithm based on dual decomposition, which is computationally efficient and resilient to communication outages or attacks. The distributed implementation relies upon message exchanges between the MGEM and LCs. For faster convergence, the proximal bundle method is employed for the non-smooth subproblem handled by the LC of RES.

Compared to the existing works, the contribution of the present work is threefold, and of critical importance for microgrids with high-penetration renewables. First, a detailed model for DS is included, and different design choices for storage cost functions are given to accommodate, for example, depth-of-discharge specifications. Second, with

the envisioned tide of high-penetration renewable energy, multiple wind farms are considered alongside two pertinent uncertainty models. Finally, a new class of controllable loads is added, with each load having a requirement of total energy over the scheduling horizon, as is the case with charging of plug-in hybrid electric vehicles (PHEVs). Detailed numerical tests are presented to illustrate the merits of the scheduling decisions for the DG, DS, and controllable loads.

A microgrid can also be operated in the so-called *islanded* mode that is basically isolated from the main grid, featuring a higher level of resilience to possible blackout of the latter. In such a case, the microgrid does not have the capability of trading energy with the main grid in order to keep the balance. Prior works include approaches of keeping power supply shortage with a very small probability, which is also legitimate for the long-term planning of the bulk system. Single-period chance-constrained ED is studied for a power system with both thermal generators and wind turbines in [89]. By using a here-and-now approach, a loss-of-load probability (LOLP)-guaranteed dispatch strategy is obtained. However, this approach is only applicable for single-period scheduling with a single wind turbine. Considering the uncertainty of PV generation, a stochastic program is formulated to minimize the overall cost of electricity and natural gas in [62]. A multi-period ED with spatio-temporal wind forecasts is pursued in [149] with the forecasted wind generation serving as the upper bound for the scheduled wind power, which results in a deterministic optimization formulation. However, the optimal solution to the proposed problem can be very sensitive to the forecasting accuracy of the wind power generation. Furthermore, minimizing a system net cost for the worst-transaction scenario, a distributed robust energy management for microgrids with renewables has been advocated very recently in [165]. Direct coupling of uncertain renewable energy supply with deferrable demand is advocated in [111] using stochastic dynamic programming. However, in all aforementioned works, risk-constrained formulations for *multi-period* scheduling with *multiple* renewable energy facilities have not been considered. Finally, relying upon Gaussianity assumptions for the wind power output and conic programming techniques, chance-constrained optimal power flow has been recently pursued in [17] and [127].

This thesis deals with optimal multi-period energy management for multiple wind farms, including ED and direct current optimal power flow (DC-OPF). To address the

intrinsically stochastic nature of non-dispatchable wind power, a chance-constrained optimization problem is formulated to limit the LOLP risk. Since the spatio-temporal joint distribution of the wind power generation is intractable to derive, a novel scenario approximation technique is introduced using Monte Carlo sampling, bypassing the need for Gaussianity assumptions. In order to guarantee a small LOLP in power systems with many generators and loads, many wind power samples are needed, which gives rise to an optimization problem with prohibitively many constraints. A key feature of this thesis is exploitation of the problem structure to obtain a sample-size-free problem formulation. Specifically, no matter how large the required sample size is, the resultant optimization problem entails just a single supply-demand balance constraint per time slot, which makes the problem efficiently solvable even for very small LOLP requirements over a long scheduling time horizon. Moreover, to capture the temporal and spatial correlation among power outputs of multiple wind farms, a vector autoregressive model is introduced to generate the required samples relying on wind speed distribution models as well as the wind-speed-to-power-output mappings.

1.2 Day-Ahead Electricity Market Clearing

Turning attention to power system economics, market clearing (MC) is one of the most important routines for a power market, which relies on security-constrained UC or OPF. Independent system operators (ISO) collect generation bids and consumption offers from the day-ahead (DA) electricity market. The MC process is then implemented to determine the market-clearing prices [124]. Deterministic MC without RES has been extensively studied; see e.g., [29,54,65]. Optimal wind power trading or contract offerings have been investigated from the perspective of wind power producers (WPPs) [10,20,23,100]. MC under uncertain power generation was recently pursued as well. As uncertainty of wind power is revealed on a continuous basis, ISOs are prompted to undertake corrective measures from the very beginning of the scheduling horizon [38]. One approach for an ISO to control the emerging risk is through the deployment of reserves following the contingencies [24]. Electricity pricing and power generation scheduling with uncertainties were accomplished via stochastic programming [97, 113]. In addition, one can co-optimize the competing objectives of generation cost and security indices [8]; see

also [9] for a stochastic security-constrained approach. Albeit computationally complex, stochastic bilevel programs are attractive because they can account for the coupling between DA and real-time (RT) markets [98, 99].

All existing MC approaches, however, are centralized. Moreover, they are not tailored to address the challenges of emerging large-scale dispatchable loads. Specifically, demand offers come from demand response (DR) aggregators serving large numbers of residential appliances that feature diverse utility functions and inter-temporal constraints. In this context, the present thesis deals with the DC-OPF based MC with high-penetration wind power. Instead of the worst-case or chance-constrained formulations, a novel stochastic optimization approach is proposed to maintain the nodal power balance while minimizing (maximizing) the grid-wide social cost (welfare). The social cost accounts for the conventional generation costs, the dis-utility of dispatchable loads, as well as a risk measure of the cost incurred by (over-) under-estimating the actual wind generation. This is essentially a cost of re-dispatching the system to compensate wind forecast errors, and is referred to as *transaction cost*. The transaction cost in the spot market is modulated through an efficient risk measure, namely the *conditional value-at-risk* (CVaR), which accounts not only for the *expected cost* of the recourse actions, but also for their "risks." A distribution-free sample average approximation (SAA) is employed to bypass the prohibitively burdensome integration involved in the CVaR-based convex minimization. To clear the market in a distributed fashion, a fast and provably convergent solver is developed using the alternating direction method of multipliers (ADMM).

The main contributions of the present thesis in this problem are: i) a CVaR-based transaction cost is introduced for the day-ahead MC to judiciously control the risk of under- and over-estimating the wind power generation; ii) a sufficient condition pertinent to transaction prices is established to effect convexity of the CVaR-based cost; and iii) a distributed solver of the resulting stochastic MC task is developed to be run by the market operator and DR aggregators while respecting the privacy of end users. Tailored to the problem structure, an efficient SAA-based method and a fast-converging ADMM solver are developed to bypass the high-dimensional integral and reduce the communication overhead involved.

1.3 Robust Resource Allocation for Wireless Networks

Cognitive radio (CR) is recognized as a disruptive technology with great potential to enhance spectrum efficiency. From the envisioned CR-driven applications, particularly promising is the hierarchical spectrum sharing [172], where CRs opportunistically reuse frequency bands licensed to primary users (PUs) whenever spectrum vacancies are detected in the time and space dimensions. Key enablers of a seamless coexistence of CR with PU systems are reliable sensing of the licensed spectrum [52,78], and judicious control of the interference that CRs inflict to PUs [172]. In this thesis, attention is focused on the latter aspect.

Recently, underlay multiple-input and multiple-output (MIMO) CR networks have attracted considerable attention thanks to their ability to mitigate both self- and PU-inflicted interference via beamforming, while leveraging spatial multiplexing and diversity to considerably increase transmission rates and reliability. On the other hand, wireless transceiver optimization has been extensively studied in the non-CR setup under different design criteria [36,116], and when either perfect or imperfect channel knowledge is available; see e.g., [45, 141], and references therein. In general, when network-wide performance criteria such as weighted sum-rate and sum mean squared error (MSE) are utilized, optimal beamforming is deemed challenging because the resultant optimization problems are typically non-convex. Thus, solvers assuring even first-order Karush-Kuhn-Tucker (KKT) optimality are appreciated in this context [36, 45, 116].

In the CR setup, the beamforming design problem is exacerbated by the presence of interference constraints [172]. In fact, while initial efforts in designing beamformers under PU interference constraints were made under the premise of *perfect* knowledge of the cognitive-to-primary propagation channels [79, 123, 157, 159], it has been recognized that obtaining accurate estimates of the CR-to-PU channels is challenging or even impossible. This is primarily due to the lack of full CR-PU cooperation [172], but also to estimation errors and frequency offsets between reciprocal channels when CR-to-PU channel estimation is attempted. It is therefore of paramount importance to take the underlying *channel uncertainties* into account, and develop prudent beamforming schemes that ensure protection of the licensed users.

Based on CR-to-PU channel statistics, probabilistic interference constraints were

employed in [42] for single-antenna CR links. Assuming imperfect knowledge of the CR-to-PU channel, the beamforming design in a multiuser CR system sharing resources with single-antenna PUs was considered in [55]; see also [173] for a downlink setup, where both CR and PU nodes have multiple antennas. The minimum CR signal-to-interference-plus-noise ratio (SINR) was maximized under a bounded norm constraint capturing uncertainty in the CR-to-PU links. Using the same uncertainty model, minimization of the overall MSE from all data streams in MIMO ad hoc CR networks was considered in [56]. However, identical channel estimation errors for different CR-to-PU links were assumed. This assumption was bypassed in [6], where the mutual information was maximized instead. Finally, a distributed algorithm based on a game-theoretic approach was developed in [144].

This thesis considers an underlay MIMO ad hoc CR network sharing spectrum bands licensed to PUs, which are possibly equipped with multiple antennas as well. CR-to-CR channels are assumed known perfectly, but this is not the case for CR-to-PU channels. Capitalizing on a norm-bounded uncertainty model to capture inaccuracies of the CR-to-PU channel estimates, a beamforming problem is formulated whereby CRs minimize the overall MSE, while limiting the interference inflicted to the PUs *robustly*. The resultant robust beamforming design confronts two major challenges: *a)* non-convexity of the total MSE cost function; and, *b)* the semi-infinite attribute of the robust interference constraint, which makes the optimization problem arduous to manage. To overcome the second hurdle, an equivalent re-formulation of the interference constraint as a linear matrix inequality (LMI) is derived by exploiting the S-Procedure [26]. On the other hand, to cope with the inherent non-convexity, a cyclic block coordinate ascent approach [14] is adopted along with local convex approximation techniques. This yields an iterative solution of the semi-definite programs (SDPs) involved, and generates a convergent sequence of objective function values. Moreover, when the CR-to-CR channel matrices have full column rank, every limit point generated by the proposed method is guaranteed to be a stationary point of the original non-convex problem. However, CR links where the transmitter is equipped with a larger number of antennas than the receiver, or spatially correlated MIMO channels [77], can lead to beamformers that are not necessarily optimal. For this reason, a proximal point-based regularization technique [118] is also employed to guarantee convergence to optimal operating points, regardless of the

channel rank and antenna configuration.

Interestingly, the schemes developed are suitable for *distributed* operation, provided that relevant parameters are exchanged among neighboring CRs. The algorithms can also be implemented in an *on-line* fashion which allows adaptation to (slow) time-varying propagation channels. In this case, CRs do not necessarily wait for the iterations to converge, but rather use the beamformer weights as and when they become available. This is in contrast to, e.g., [36, 45] and [6, 56, 173] in the non-CR and CR cases, respectively, where the relevant problems are solved *centrally* and in a *batch* form.

For the robust beamforming design, the interference power that can be tolerated by the PUs is initially assumed to be pre-partitioned in per-CR link portions, possibly according to quality-of-service (QoS) guidelines [123, 144]. However, extensions of the beamforming design are also provided when the PU interference limit is *not* divided a priori among CR links. In this case, primal decomposition techniques [14] are invoked to dynamically allocate the total interference among CRs. Compared to [56], the proposed scheme accounts for different estimation inaccuracies in the CR-to-PU links.

To accommodate the explosive demand for wireless services, cellular systems are evolving into what are termed heterogeneous networks (HetNets) consisting of distributed macro/micro/pico base stations (BSs) to cover overlapping areas of different sizes [71]. Close proximity of many HetNet transmitters introduces severe inter-cell interference. For efficient interference management, coordinated multi-point processing (CoMP) has emerged as a promising technique for next-generation cellular networks such as LTE-Advanced [72].

To fully exploit the potential of CoMP at affordable overhead, coordinated beamforming and/or clustered BS cooperation for downlink systems were investigated in [40, 69, 81, 104, 134, 156]. Multiple BSs cooperate to beamform, and each user's data are only shared among a small number of BSs per cluster, thus greatly reducing the overall backhaul signalling cost.

The rapid development of small cells in HetNets has also driven the need for energy-efficient transmissions. Due to the growing number of BSs, the electricity bill has become a major part of the operational expenditure of cellular operators, and cellular networks contribute a considerable portion of the global "carbon footprint" [106]. These economic and ecological concerns advocate a "green communication" solution, where

BSs in cellular networks are powered by the electricity grid [40, 69, 81, 104, 106, 115, 126, 134, 156]. However, the current grid infrastructure is on the verge of a major paradigm shift, migrating from the traditional electricity grid to the so-termed "smart grid" equipped with a large number of advanced smart meters and state-of-the-art communication and control links. To decrease greenhouse gas emissions, an important feature of future power systems is integration of RES. This leads to high penetration of distributed generators equipped with energy harvesting modules, which can crop energy from the environmental resources (e.g., solar and wind), and possibly trade the harvested energy with the main grid. In addition to distributed generation, distributed storage, and two-way energy trading associated with RES, DSM including dynamic pricing and demand response, can further improve grid reliability and efficiency. Relying on pertinent tools, optimal energy management and scheduling with RES and/or DSM were proposed in [57, 62, 88, 165].

To take advantage of the aforementioned smart grid capabilities in next-generation cellular systems, only a few recent works have considered the smart-grid powered CoMP transmissions [27, 151, 152]. However, [27] only addressed dynamic pricing using a simplified smart grid level game, while [151] and [152] assumed that the energy amounts harvested from RES are precisely available *a priori* (e.g., through forecasting), and the harvested energy cannot be stored at the BSs. In addition, [152] assumed demand (or load) response based on different energy buying/selling prices across BSs without adapting power consumption to *time-varying* energy pricing. The smart grid models in the existing works [27, 151, 152] are somewhat (over-)simplified for the ease of exploration. While [27] simply addressed DSM with dynamic pricing, [151, 152] assumed that harvested energy cannot be stored at the BSs (i.e., the BS either consumes all energy or sells some to the grid) and all harvested energy amount as well as channel state information are precisely available *a priori*.

This thesis deals with optimal energy management and transmit-beamforming designs for the smart-grid powered, cluster-based CoMP downlink with the clustering carried by the HetNet's central processor. Each BS has local RES and can perform two-way energy trading with the grid based on *dynamic* buying and selling prices. Different from [27, 151, 152], we suppose that each BS has a local storage device, which can be charged to store the harvested (and even grid) energy and can be discharged to supply

electricity if needed. To account for the stochastic and nondispatchable nature of both RES and wireless channels, we assume that the actual harvested energy amounts and the wireless channel states are *unknown* and *time-varying*, yet lie in some known uncertainty regions. Building on realistic models, we develop robust energy management and transmit-beamforming designs that minimize the worst-case energy cost subject to the worst-case user QoS guarantees for the CoMP downlink. Leveraging a novel formulation accounting for the worst-case transaction cost with two-way energy trading, as well as the S-procedure in robust beamforming designs, we show how to (re-)formulate the task at hand as a convex problem. Strong duality of the latter is then utilized to develop a Lagrange dual based subgradient solver. It is shown that the resultant algorithm is guaranteed to find the desired optimal energy management strategy and transmit-beamforming vectors, and can also facilitate distributed implementations among the BSs.

Chapter 2

Robust Energy Management for Grid-Connected Microgrids

2.1 Microgrid Energy Management Problem

Consider a microgrid comprising M conventional (fossil fuel) generators, I RES facilities, and J DS units (see also Fig. 1.1). The scheduling horizon is $\mathcal{T} := \{1, 2, \ldots, T\}$ (e.g., one-day ahead). The particulars of the optimal scheduling problem are explained in the next subsections.

2.1.1 Load Demand Model

Loads are classified in two categories. The first comprises inelastic loads, whose power demand should be satisfied at all times. Examples are power requirements of hospitals or illumination demand from residential areas.

The second category consists of elastic loads, which are dispatchable, in the sense that their power consumption is adjustable, and can be scheduled. These loads can be further divided in two classes, each having the following characteristics:

i) The first class contains loads with power consumption $P_{D_n}^t \in [P_{D_n}^{\min}, P_{D_n}^{\max}]$, where $n \in \mathcal{N} := \{1, \ldots, N\}$, and $t \in \mathcal{T}$. Higher power consumption yields higher utility for the end user. The utility function of the nth dispatchable load, $U_{D_n}^t(P_{D_n}^t)$, is selected to be increasing and concave, with typical choices being piecewise linear

or smooth quadratic; see also [33]. An example from this class is an A/C.

ii) The second class includes loads indexed by $q \in \mathcal{Q} := \{1, \ldots, Q\}$ with power consumption limits $P_{E_q}^{\min}$ and $P_{E_q}^{\max}$, and prescribed total energy requirements E_q which have to be achieved from the start time S_q to termination time T_q; see e.g., [95]. This type of loads can be the plug-in hybrid electric vehicles (PHEVs). Power demand variables $\{P_{E_q}^t\}_{t=1}^T$ therefore are constrained as $\sum_{t=S_q}^{T_q} P_{E_q}^t = E_q$ and $P_{E_q}^t \in [P_{E_q}^{\min,t}, P_{E_q}^{\max,t}],\ t \in \mathcal{T}$, while $P_{E_q}^{\min,t} = P_{E_q}^{\max,t} = 0$ for $t \notin \{S_q, \ldots, T_q\}$. Higher power consumption in earlier slots as opposed to later slots may be desirable for a certain load, so that the associated task finishes earlier. This behavior can be encouraged by adopting for the qth load an appropriately designed time-varying concave utility function $U_{E_q}^t(P_{E_q}^t)$. An example is $U_{E_q}^t(P_{E_q}^t) := \pi_q^t P_{E_q}^t$, with weights $\{\pi_q^t\}$ decreasing in t from slots S_q to T_q. Naturally, $U_{E_q}^t(P_{E_q}^t) \equiv 0$ can be selected if the consumer is indifferent to how power is consumed across slots.

2.1.2 Distributed Storage Model

Let B_j^t denote the stored energy of the jth battery at the end of the slot t, with initial available energy B_j^0 while $B_j^{\max}$ denotes the battery capacity, so that $0 \leq B_j^t \leq B_j^{\max},\ j \in \mathcal{J} := \{1, \ldots, J\}$. Let $P_{B_j}^t$ be the power delivered to (drawn from) the jth storage device at slot t, which amounts to charging ($P_{B_j}^t \geq 0$) or discharging ($P_{B_j}^t \leq 0$) of the battery. Clearly, the stored energy obeys the dynamic equation

$$B_j^t = B_j^{t-1} + P_{B_j}^t,\ j \in \mathcal{J},\ t \in \mathcal{T}\ . \tag{2.1}$$

Variables $P_{B_j}^t$ are constrained in the following ways:

i) The amount of (dis)charging is bounded, that is

$$P_{B_j}^{\min} \leq P_{B_j}^t \leq P_{B_j}^{\max} \tag{2.2}$$

$$-\eta_j B_j^{t-1} \leq P_{B_j}^t \tag{2.3}$$

with bounds $P_{B_j}^{\min} < 0$ and $P_{B_j}^{\max} > 0$, while $\eta_j \in (0, 1]$ is the efficiency of DS unit j [7, 142]. The constraint in (2.3) means that a fraction η_j of the stored energy B_j^{t-1} is available for discharge.

ii) Final stored energy is also bounded for the sake of future scheduling horizons, that is $B_j^T \geq B_j^{\min}$.

To maximize DS lifetime, a storage cost $H_j^t(B_j^t)$ can be employed to encourage the stored energy to remain above a specified depth of discharge, denoted as $\mathsf{DOD}_j \in [0,1]$, where 100% (0%) depth of discharge means the battery is empty (full) [7]. Such a cost is defined as $H_j^t(B_j^t) := \psi_j^t[(1-\mathsf{DOD}_j)B_j^{\max} - B_j^t]$. Note that the storage cost $H_j^t(B_j^t)$ can be interpreted as imposing a soft constraint preventing large variations of the stored energy. Clearly, higher weights $\{\psi_j^t\}$ encourage smaller variation. If high power exchange is to be allowed, these weights can be chosen very small, or one can even select $H_j^t(B_j^t) \equiv 0$ altogether.

2.1.3 Worst-Case Transaction Cost

Let W_i^t denote the *actual* renewable energy harvested by the ith RES facility at time slot t, and also let $\mathbf{w}$ collect all W_i^t, i.e., $\mathbf{w} := [W_1^1, \ldots, W_1^T, \ldots, W_I^1, \ldots, W_I^T]$. To capture the intrinsically stochastic and time-varying availability of RES, it is postulated that $\mathbf{w}$ is unknown, but lies in a polyhedral uncertainty set $\mathcal{W}$. The following are two practical examples.

i) The first example postulates a separate uncertainty set $\mathcal{W}_i$ for each RES facility in the form

$$\mathcal{W}_i := \left\{ \{W_i^t\}_{t=1}^T \;\middle|\; \underline{W}_i^t \leq W_i^t \leq \overline{W}_i^t,\, W_{i,s}^{\min} \leq \sum_{t\in\mathcal{T}_{i,s}} W_i^t \leq W_{i,s}^{\max}, \mathcal{T} = \bigcup_{s=1}^{S} \mathcal{T}_{i,s} \right\} \tag{2.4}$$

where $\underline{W}_i^t$ ($\overline{W}_i^t$) denotes a lower (upper) bound on W_i^t; $\mathcal{T}$ is partitioned into consecutive but non-overlapping sub-horizons $\mathcal{T}_{i,s}$ for $i = 1, \ldots, I$, $s = 1, 2, \ldots, S$; the total renewable energy for the ith RES facility over the sth sub-horizon is assumed bounded by $W_{i,s}^{\min}$ and $W_{i,s}^{\max}$. In this example, $\mathcal{W}$ takes the form of Cartesian product $\mathcal{W} = \mathcal{W}_1 \times \ldots \times \mathcal{W}_I$.

ii) The second example assumes a joint uncertainty model across all RES facilities as

$$\mathcal{W} := \left\{ \mathbf{w} \;\middle|\; \underline{W}_i^t \leq W_i^t \leq \overline{W}_i^t,\, W_s^{\min} \leq \sum_{t\in\mathcal{T}_s} \sum_{i=1}^{I} W_i^t \leq W_s^{\max}, \mathcal{T} = \bigcup_{s=1}^{S} \mathcal{T}_s \right\} \tag{2.5}$$

where $\underline{W}_i^t$ ($\overline{W}_i^t$) denotes a lower (upper) bound on W_i^t; $\mathcal{T}$ is partitioned into consecutive but non-overlapping sub-horizons $\mathcal{T}_s$ for $s = 1, 2, \ldots, S$; the total renewable energy harvested by all the RES facilities over the sth sub-horizon is bounded by $W_s^{\min}$ and $W_s^{\max}$; see also [171].

The previous two RES uncertainty models are quite general and can take into account different geographical and meteorological factors. The only information required is the deterministic lower and upper bounds, namely $\underline{W}_i^t$, $\overline{W}_i^t$, $W_{i,s}^{\min}$, $W_{i,s}^{\max}$, $W_s^{\min}$, $W_s^{\max}$, which can be determined via inference schemes based on historical data [112].

Supposing the microgrid operates in a grid-connected mode, a transaction mechanism between the microgrid and the main grid is present, whereby the microgrid can buy/sell energy from/to the spot market. Let P_R^t be an auxiliary variable denoting the net power delivered to the microgrid from the renewable energy sources and the distributed storage in order to maintain the supply-demand balance at slot t. The shortage energy per slot t is given by $\left[P_R^t - \sum_{i=1}^I W_i^t + \sum_{j=1}^J P_{B_j}^t\right]^+$, while the surplus energy is $\left[\sum_{i=1}^I W_i^t - P_R^t - \sum_{j=1}^J P_{B_j}^t\right]^+$, where $[a]^+ := \max\{a, 0\}$.

The amount of shortage energy is bought with known purchase price α^t, while the surplus energy is sold to the main grid with known selling price β^t. The worst-case net transaction cost is thus given by

$$G(\{P_R^t\}, \{P_{B_j}^t\}) := \max_{\mathbf{w}\in\mathcal{W}} \sum_{t=1}^T \left(\alpha^t \left[P_R^t - \sum_{i=1}^I W_i^t + \sum_{j=1}^J P_{B_j}^t \right]^+ - \beta^t \left[\sum_{i=1}^I W_i^t - P_R^t - \sum_{j=1}^J P_{B_j}^t \right]^+ \right) \tag{2.6}$$

where $\{P_R^t\}$ collects P_R^t for $t \in \mathcal{T}$ and $\{P_{B_j}^t\}$ collects $P_{B_j}^t$ for $j \in \mathcal{J},\ t \in \mathcal{T}$.

Remark 2.1 *(Worst-case model versus stochastic model).* The worst-case robust model advocated here is particularly attractive when the probability distribution of the renewable power production is unavailable. This is for instance the case for multiple wind farms, where the spatio-temporal joint distribution of the wind power generation is intractable (see detailed discussions in [164] and [96]). If an accurate probabilistic model is available, an expectation-based stochastic program can be formulated to bypass the conservatism of worst-case optimization. In the case of wind generation, suppose that

wind power W_i^t is a function of the random wind velocity v_i^t, for which different models are available, and the wind-speed-to-power-output mappings $W_i^t(v_i^t)$ are known [30]. Then, the worst-case transaction cost can be replaced by the *expected* transaction cost

$$G(\{P_R^t\},\{P_{B_j}^t\}) := \mathbb{E}_{\mathbf{v}}\left[\sum_{t=1}^{T}\alpha^t\left[P_R^t-\sum_{i=1}^{I}W_i^t(v_i^t)+\sum_{j=1}^{J}P_{B_j}^t\right]^+ - \beta^t\left[\sum_{i=1}^{I}W_i^t-P_R^t-\sum_{j=1}^{J}P_{B_j}^t\right]^+\right]$$

where the vector $\mathbf{v}$ collects v_i^t for all i and t.

2.1.4 Robust Energy Management Formulation

Apart from RES, microgrids typically entail also conventional DG. Let $P_{G_m}^t$ be the power produced by the mth conventional generator, where $m \in \mathcal{M} := \{1,\ldots,M\}$ and $t \in \mathcal{T}$. The cost of the mth generator is given by an increasing convex function $C_m^t(P_{G_m}^t)$, which typically is either piecewise linear or smooth quadratic.

The energy management problem amounts to minimizing the microgrid social net cost; that is, the cost of conventional generation, storage, and the worst-case transaction cost (due to the volatility of RES) minus the utility of dispatchable loads:

$$\text{minimize} \quad \sum_{t=1}^{T}\left(\sum_{m=1}^{M}C_m^t(P_{G_m}^t)-\sum_{n=1}^{N}U_{D_n}^t(P_{D_n}^t)-\sum_{q=1}^{Q}U_{E_q}^t(P_{E_q}^t)+\sum_{j=1}^{J}H_j^t(B_j^t)\right) + G(\{P_R^t\},\{P_{B_j}^t\}) \tag{2.7a}$$

$$\text{subject to} \quad P_{G_m}^{\min} \le P_{G_m}^t \le P_{G_m}^{\max},\ m\in\mathcal{M},\ t\in\mathcal{T} \tag{2.7b}$$

$$P_{G_m}^t - P_{G_m}^{t-1} \le \mathsf{R}_m^{\text{up}},\ m\in\mathcal{M},\ t\in\mathcal{T} \tag{2.7c}$$

$$P_{G_m}^{t-1} - P_{G_m}^t \le \mathsf{R}_m^{\text{down}},\ m\in\mathcal{M},\ t\in\mathcal{T} \tag{2.7d}$$

$$\sum_{m=1}^{M}(P_{G_m}^{\max}-P_{G_m}^t) \ge \mathsf{SR}^t,\ t\in\mathcal{T} \tag{2.7e}$$

$$P_{D_n}^{\min} \le P_{D_n}^t \le P_{D_n}^{\max},\ n\in\mathcal{N},\ t\in\mathcal{T} \tag{2.7f}$$

$$P_{E_q}^{\min,t} \le P_{E_q}^t \le P_{E_q}^{\max,t},\ q\in\mathcal{Q},\ t\in\mathcal{T} \tag{2.7g}$$

$$\sum_{t=S_q}^{T_q}P_{E_q}^t = E_q,\ q\in\mathcal{Q} \tag{2.7h}$$

$$0 \le B_j^t \le B_j^{\max},\ B_j^T \ge B_j^{\min},\ j \in \mathcal{J},\ t \in \mathcal{T} \tag{2.7i}$$

$$P_{B_j}^{\min} \le P_{B_j}^t \le P_{B_j}^{\max},\ j \in \mathcal{J},\ t \in \mathcal{T} \tag{2.7j}$$

$$-\eta_j B_j^{t-1} \le P_{B_j}^t,\ j \in \mathcal{J},\ t \in \mathcal{T} \tag{2.7k}$$

$$B_j^t = B_j^{t-1} + P_{B_j}^t,\ j \in \mathcal{J},\ t \in \mathcal{T} \tag{2.7l}$$

$$P_R^{\min} \le P_R^t \le P_R^{\max},\ t \in \mathcal{T} \tag{2.7m}$$

$$\sum_{m=1}^{M} P_{G_m}^t + P_R^t = L^t + \sum_{n=1}^{N} P_{D_n}^t + \sum_{q=1}^{Q} P_{E_q}^t,\ t \in \mathcal{T} \tag{2.7n}$$

$$\text{variables} \quad \left\{P_{G_m}^t, P_{D_n}^t, P_{E_q}^t, B_j^t, P_{B_j}^t, P_R^t\right\}_{t\in\mathcal{T}}. \tag{2.7o}$$

Constraints (2.7b)–(2.7e) stand for the minimum/maximum power output, ramping up/down limits, and spinning reserves, respectively, which capture the typical physical requirements of a power generation system. Constraints (2.7f) and (2.7m) correspond to the minimum/maximum power of the flexible load demand and committed renewable energy. Constraint (2.7n) is the power supply-demand *balance equation* ensuring the total demand is satisfied by the power generation at any time.

Note that constraints (2.7b)–(2.7n) are linear, while $C_m^t(\cdot)$, $-U_{D_n}^t(\cdot)$, $-U_{E_q}^t(\cdot)$, and $H_j^t(\cdot)$ are convex (possibly non-differentiable or non-strictly convex) functions. Consequently, the convexity of (2.7) depends on that of $G(\{P_R^t\}, \{P_{B_j}^t\})$, which is established in the following proposition.

Proposition 2.1 *If the selling price β^t does not exceed the purchase price α^t for any $t \in \mathcal{T}$, then the worst-case transaction cost $G(\{P_R^t\}, \{P_{B_j}^t\})$ is convex in $\{P_R^t\}$ and $\{P_{B_j}^t\}$.*

Proof: Using the fact that $[a]^+ + [-a]^+ = |a|$ and $[a]^+ - [-a]^+ = a$, the worst-case transaction cost $G(\{P_R^t\}, \{P_{B_j}^t\})$ can be re-written as

$$G(\{P_R^t\}, \{P_{B_j}^t\}) = \max_{\mathbf{w}\in\mathcal{W}} \sum_{t=1}^{T} \left(\delta^t \left| P_R^t - \sum_{i=1}^{I} W_i^t + \sum_{j=1}^{J} P_{B_j}^t \right| + \gamma^t \left(P_R^t - \sum_{i=1}^{I} W_i^t + \sum_{j=1}^{J} P_{B_j}^t \right) \right) \tag{2.8}$$

with $\delta^t := (\alpha^t - \beta^t)/2$, and $\gamma^t := (\alpha^t + \beta^t)/2$. Since the absolute value function is convex, and the operations of nonnegative weighted summation and pointwise maximum (over an infinite set) preserve convexity [26, Sec. 3.2], the claim follows readily. □

An immediate corollary of Proposition 2.1 is that the energy management problem (2.7) is convex if $\beta^t \leq \alpha^t$ for all t. The next section focuses on this case, and designs an efficient decentralized solver for problem (2.7).

2.2 Distributed Algorithm

In order to facilitate a distributed algorithm for (2.7), a variable transformation is useful. Specifically, upon introducing $\tilde{P}_R^t := P_R^t + \sum_{j=1}^J P_{B_j}^t$, problem (2.7) can be re-written as

$$\underset{\mathbf{x}}{\text{minimize}} \quad \sum_{t=1}^{T} \Big(\sum_{m=1}^{M} C_m^t(P_{G_m}^t) - \sum_{n=1}^{N} U_{D_n}^t(P_{D_n}^t) - \sum_{q=1}^{Q} U_{E_q}^t(P_{E_q}^t) + \sum_{j=1}^{J} H_j^t(B_j^t) \Big) + G(\{\tilde{P}_R^t\}) \tag{2.9a}$$

$$\text{subject to} \quad \tilde{P}_R^t = P_R^t + \sum_{j=1}^{J} P_{B_j}^t, \ t \in \mathcal{T} \tag{2.9b}$$

$$(2.7\text{b}) - (2.7\text{n})$$

where $\mathbf{x}$ collects all the primal variables $\{P_{G_m}^t, P_{D_n}^t, P_{E_q}^t, P_{B_j}^t, B_j^t, P_R^t, \tilde{P}_R^t\}$ for $t \in \mathcal{T}$; and

$$G(\{\tilde{P}_R^t\}) := \max_{\mathbf{w} \in \mathcal{W}} \sum_{t=1}^{T} \left(\delta^t \left| \tilde{P}_R^t - \sum_{i=1}^{I} W_i^t \right| + \gamma^t \left(\tilde{P}_R^t - \sum_{i=1}^{I} W_i^t \right) \right). \tag{2.10}$$

The following proposition extends the result of Proposition 2.1 to the transformed problem, and asserts its strong duality.

Proposition 2.2 *If* (2.9) *is feasible, and the selling price* β^t *does not exceed the purchase price* α^t *for any* $t \in \mathcal{T}$*, then there is no duality gap.*

Proof: Due to the strong duality theorem for the optimization problems with linear constraints [cf. [12, Prop. 5.2.1]], it suffices to show that the cost function is convex over the entire space and its optimal value is finite. First, using the same argument, convexity of $G(\{\tilde{P}_R^t\})$ in $\{\tilde{P}_R^t\}$ is immediate under the transaction price condition. The finiteness of the optimal value is guaranteed by the fact that the continuous convex cost (2.9a) is minimized over a nonempty compact set specified by (2.7b)–(2.7n), and (2.9b). □

The strong duality asserted by Proposition 2.2 motivates the use of Lagrangian relaxation techniques in order to solve the scheduling problem. Moreover, problem (2.9) is clearly separable, meaning that its cost and constraints are sums of terms, with each term dependent on different optimization variables. The features of strong duality and separability imply that Lagrangian relaxation and dual decomposition are applicable to yield a decentralized algorithm; see also related techniques in power systems [39] and communication networks [35, 108]. Coordinated by dual variables, the dual approach decomposes the original problem into several separate subproblems that can be solved by the LCs in parallel. The development of the distributed algorithm is undertaken next.

2.2.1 Dual Decomposition

Constraints (2.7e), (2.7n), and (2.9b) couple variables across generators, loads, and the RES. Let $\mathbf{z}$ collect dual variables $\{\mu^t\}$, $\{\lambda^t\}$, and $\{\nu^t\}$, which denote the corresponding Lagrange multipliers. Keeping the remaining constraints implicit, the partial Lagrangian is given by

$$\begin{aligned} L(\mathbf{x},\mathbf{z}) = &\sum_{t=1}^{T}\Big(\sum_{m=1}^{M} C_m^t(P_{G_m}^t) - \sum_{n=1}^{N} U_{D_n}^t(P_{D_n}^t) - \sum_{q=1}^{Q} U_{E_q}^t(P_{E_q}^t) + \sum_{j=1}^{J} H_j^t(B_j^t)\Big) \\ &+ G(\{\tilde{P}_R^t\}) + \sum_{t=1}^{T}\Big\{\mu^t\Big(\mathsf{SR}^t - \sum_{m=1}^{M}(P_{G_m}^{\max} - P_{G_m}^t)\Big) - \nu^t\Big(\tilde{P}_R^t - P_R^t - \sum_{j=1}^{J} P_{B_j}^t\Big) \\ &- \lambda^t\Big(\sum_{m=1}^{M} P_{G_m}^t + P_R^t - \sum_{n=1}^{N} P_{D_n}^t - \sum_{q=1}^{Q} P_{E_q}^t - L^t\Big)\Big\}. \quad (2.11) \end{aligned}$$

Then, the dual function can be written as

$$\begin{aligned} D(\mathbf{z}) = \underset{\mathbf{x}}{\text{minimize}} \quad & L(\mathbf{x},\mathbf{z}) \\ \text{subject to} \quad & (2.7\text{b}) - (2.7\text{d}),\ (2.7\text{f}) - (2.7\text{m}) \end{aligned}$$

and the dual problem is given by

$$\begin{aligned} &\text{maximize} \quad D(\{\mu^t\},\{\lambda^t\},\{\nu^t\}) && (2.12\text{a}) \\ &\text{subject to} \quad \mu^t \geq 0, \lambda^t, \nu^t \in \mathbb{R},\ t \in \mathcal{T}. && (2.12\text{b}) \end{aligned}$$

The subgradient method will be employed to obtain the optimal multipliers and power schedules. The iterative process is described next, followed by its distributed implementation.

The subgradient method amounts to running the recursions [13, Sec. 6.3]

$$\mu^t(k+1) = [\mu^t(k) + a g_{\mu^t}(k)]^+ \tag{2.13a}$$

$$\lambda^t(k+1) = \lambda^t(k) + a g_{\lambda^t}(k) \tag{2.13b}$$

$$\nu^t(k+1) = \nu^t(k) + a g_{\nu^t}(k) \tag{2.13c}$$

where k is the iteration index; $a > 0$ is a constant stepsize; while $g_{\mu^t}(k)$, $g_{\lambda^t}(k)$, and $g_{\nu^t}(k)$ denote the subgradients of the dual function with respect to $\mu^t(k)$, $\lambda^t(k)$, and $\nu^t(k)$, respectively. These subgradients can be expressed in the following simple forms

$$g_{\mu^t}(k) = \mathsf{SR}^t - \sum_{m=1}^{M} (P_{G_m}^{\max} - P_{G_m}^t(k)) \tag{2.14a}$$

$$g_{\lambda^t}(k) = L^t + \sum_{n=1}^{N} P_{D_n}^t(k) + \sum_{q=1}^{Q} P_{E_q}^t(k) - \sum_{m=1}^{M} P_{G_m}^t(k) - P_R^t(k) \tag{2.14b}$$

$$g_{\nu^t}(k) = P_R^t(k) + \sum_{j=1}^{J} P_{B_j}^t(k) - \tilde{P}_R^t(k) \tag{2.14c}$$

where $P_{G_m}^t(k)$, $P_{D_n}^t(k)$, $P_{E_q}^t(k)$, $P_{B_j}^t(k)$, $P_R^t(k)$, and $\tilde{P}_R^t(k)$ are given as follows:

$$\{P_{G_m}^t(k)\}_{t=1}^T \in \underset{\substack{\{P_{G_m}^t\} \\ \text{subject to (2.7b)}-\text{(2.7d)}}}{\arg\min} \sum_{t=1}^{T} \left(C_m^t(P_{G_m}^t) + \big(\mu^t(k) - \lambda^t(k)\big) P_{G_m}^t \right) \tag{2.15}$$

$$\{P_{D_n}^t(k)\}_{t=1}^T \in \underset{\substack{\{P_{D_n}^t\} \\ \text{subject to (2.7f)}}}{\arg\min} \sum_{t=1}^{T} \left(\lambda^t(k) P_{D_n}^t - U_{D_n}^t(P_{D_n}^t) \right) \tag{2.16}$$

$$\{P_{E_q}^t(k)\}_{t=1}^T \in \underset{\substack{\{P_{E_q}^t\} \\ \text{subject to (2.7g)}-\text{(2.7h)}}}{\arg\min} \sum_{t=1}^{T} \left(\lambda^t(k) P_{E_q}^t - U_{E_q}^t(P_{E_q}^t) \right) \tag{2.17}$$

$$\{P_{B_j}^t(k)\}_{t=1}^T \in \underset{\substack{\{P_{B_j}^t, B_j^t\} \\ \text{subject to (2.7i)}-\text{(2.7l)}}}{\arg\min} \sum_{t=1}^{T} \left(\nu^t(k) P_{B_j}^t + H_j^t(B_j^t) \right) \tag{2.18}$$

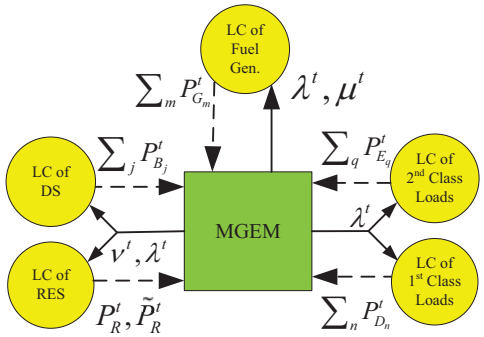


Figure 2.1: Decomposition and message exchange.

$$\{P_R^t(k), \tilde{P}_R^t(k)\}_{t=1}^T \in \underset{\substack{\{P_R^t, \tilde{P}_R^t\} \\ \text{subject to (2.7m)}}}{\arg\min} \left\{ \sum_{t=1}^T \left(\big(\nu^t(k) - \lambda^t(k)\big) P_R^t \right) + G(\{\tilde{P}_R^t\}) - \sum_{t=1}^T \nu^t(k) \tilde{P}_R^t \right\}. \quad (2.19)$$

Iterations are initialized with arbitrary $\lambda^t(0), \nu^t(0) \in \mathbb{R}$, and $\mu^t(0) \geq 0$. The iterates are guaranteed to converge to a neighborhood of the optimal multipliers [13, Sec. 6.3]. The size of the neighborhood is proportional to the stepsize, and can therefore be controlled by the stepsize.

When the primal objective is *not* strictly convex, a primal averaging procedure is necessary to obtain the optimal power schedules, which are then given by

$$\bar{\mathbf{x}}(k) = \frac{1}{k} \sum_{j=0}^{k-1} \mathbf{x}(j) = \frac{1}{k} \mathbf{x}(k-1) + \frac{k-1}{k} \bar{\mathbf{x}}(k-1). \quad (2.21)$$

The running averages can be recursively computed as in (2.21), and are also guaranteed to converge to a neighborhood of the optimal solution [102]. Note that other convergence-guaranteed stepsize rules and primal averaging methods can also be utilized; see [53] for detailed discussions.

The form of the subgradient iterations easily lends itself to a distributed implementation utilizing the control and communication capabilities of a typical microgrid. Specifically, the MGEM maintains and updates the Lagrange multipliers via (2.13). The LCs of conventional generation, dispatchable loads, storage units, and RES solve subproblems (2.15)–(2.19), respectively. These subproblems can be solved if the MGEM sends

Algorithm 2.1 Distributed Energy Management

1: Initialize Lagrange multipliers $\lambda^t = \mu^t = \nu^t = 0$
2: **repeat** $(k = 0, 1, 2, \ldots)$
3: **for** $t = 1, 2, \ldots, T$ **do**
4: Broadcast $\lambda^t(k)$, $\mu^t(k)$, and $\nu^t(k)$ to LCs of convectional generators, controllable loads, storage units, and RES facilities
5: Update power scheduling $P^t_{G_m}(k)$, $P^t_{D_n}(k)$, $P^t_{E_q}(k)$, $P^t_{B_j}(k)$, $P^t_R(k)$, and $\tilde{P}^t_R(k)$ by solving (2.15)–(2.19)
6: Update $\lambda^t(k)$, $\mu^t(k)$, and $\nu^t(k)$ via (2.13)
7: **end for**
8: Running averages of primal variables via (2.21)
9: **until** Convergence

the current multiplier iterates $\mu^t(k)$, $\lambda^t(k)$, and $\nu^t(k)$ to the LCs. The LCs send back to the MGEM the quantities $\sum_{m=1}^{M} P^t_{G_m}(k)$, $\sum_{n=1}^{N} P^t_{D_n}(k)$, $\sum_{q=1}^{Q} P^t_{E_q}(k)$, $\sum_{j=1}^{J} P^t_{B_j}(k)$, $P^t_R(k)$, and $\tilde{P}^t_R(k)$ which are in turn used to form the subgradients according to (2.14). The distributed algorithm using dual decomposition is tabulated as Algorithm 2.1, and the interactive process of message passing is illustrated in Fig. 2.1.

2.2.2 Local Controller Subproblems

This subsection shows how to solve each subproblem (2.15)–(2.19). Specifically, $C^t_m(\cdot)$, $-U^t_{D_n}(\cdot)$, $-U^t_{E_q}(\cdot)$, and $H^t_j(\cdot)$ are chosen either convex piece-wise linear or smooth convex quadratic. Correspondingly, the first four subproblems (2.15)–(2.18) are essentially linear programs (LPs) or quadratic programs (QPs), which can be solved efficiently. Therefore, the main focus is on solving (2.19).

The optimal solution of $P^t_R(k)$ in (2.19) is easy to obtain as

$$P^t_R(k) = \begin{cases} P^{\min}_R, & \text{if } \nu^t(k) \geq \lambda^t(k) \\ P^{\max}_R, & \text{if } \nu^t(k) < \lambda^t(k). \end{cases} \tag{2.22}$$

However, due to the absolute value operator and the maximization over $\mathbf{w}$ in the definition of $G(\{\tilde{P}^t_R\})$, subproblem (2.19) is a convex nondifferentiable problem in $\{\tilde{P}^t_R\}$, which can be challenging to solve.

As a state-of-the-art technique for convex nondifferentiable optimization problems [13,

Ch. 6], the bundle method is employed here to obtain $\{\tilde{P}_R^t(k)\}$. Define first the objective

$$\tilde{G}(\{\tilde{P}_R^t\}) := G(\{\tilde{P}_R^t\}) - \sum_{t=1}^{T} \nu^t(k)\tilde{P}_R^t. \tag{2.23}$$

By the generalization of Danskin's Theorem [13, Sec. 6.3], the subgradient of $\tilde{G}(\{\tilde{P}_R^t\})$ required by the bundle method can be obtained as

$$\partial\tilde{G}(\{\tilde{P}_R^t\}) = \begin{cases} \alpha^t - \nu^t(k), & \text{if } \tilde{P}_R^t \geq \sum_{i=1}^{I}(W_i^t)^* \\ \beta^t - \nu^t(k), & \text{if } \tilde{P}_R^t < \sum_{i=1}^{I}(W_i^t)^* \end{cases} \tag{2.24}$$

where for given $\{\tilde{P}_R^t\}$ it holds that

$$\mathbf{w}^* \in \arg\max_{\mathbf{w}\in\mathcal{W}} \sum_{t=1}^{T} \left(\delta^t \left| \tilde{P}_R^t - \sum_{i=1}^{I} W_i^t \right| + \gamma^t \left(\tilde{P}_R^t - \sum_{i=1}^{I} W_i^t \right) \right). \tag{2.25}$$

With $\mathbf{p} := [\tilde{P}_R^1, \ldots, \tilde{P}_R^T]$, the bundle method generates a sequence $\{\mathbf{p}_\ell\}$ with guaranteed convergence to the optimal $\{\tilde{P}_R^t(k)\}$; see e.g., [51], [13, Ch. 6]. The iterate $\mathbf{p}_{\ell+1}$ is obtained by minimizing a polyhedral approximation of $\tilde{G}(\mathbf{p})$ with a quadratic proximal regularization as follows

$$\mathbf{p}_{\ell+1} := \arg\min_{\mathbf{p}\in\mathbb{R}^T} \left\{ \hat{G}_\ell(\mathbf{p}) + \frac{\rho_\ell}{2}\|\mathbf{p} - \mathbf{y}_\ell\|^2 \right\} \tag{2.26}$$

where $\hat{G}_\ell(\mathbf{p}) := \max\{\tilde{G}(\mathbf{p}_0) + \mathbf{g}_0'(\mathbf{p} - \mathbf{p}_0), \ldots, \tilde{G}(\mathbf{p}_\ell) + \mathbf{g}_\ell'(\mathbf{p} - \mathbf{p}_\ell)\}$; $\mathbf{g}_\ell$ is the subgradient of $\tilde{G}(\mathbf{p})$ evaluated at the point $\mathbf{p} = \mathbf{p}_\ell$, which is calculated according to (2.24); proximity weight ρ_ℓ is to control stability of the iterates; and the proximal center $\mathbf{y}_\ell$ is updated according to a query for descent

$$\mathbf{y}_{\ell+1} = \begin{cases} \mathbf{p}_{\ell+1}, & \text{if } \tilde{G}(\mathbf{y}_\ell) - \tilde{G}(\mathbf{p}_{\ell+1}) \geq \theta\eta_\ell \\ \mathbf{y}_\ell, & \text{if } \tilde{G}(\mathbf{y}_\ell) - \tilde{G}(\mathbf{p}_{\ell+1}) < \theta\eta_\ell \end{cases} \tag{2.27}$$

where $\eta_\ell = \tilde{G}(\mathbf{y}_\ell) - \left(\hat{G}_\ell(\mathbf{p}_{\ell+1}) + \frac{\rho_\ell}{2}\|\mathbf{p}_{\ell+1} - \mathbf{y}_\ell\|^2 \right)$ and $\theta \in (0, 1)$.

Algorithms for solving (2.25) depend on the form of the uncertainty set $\mathcal{W}$, and are elaborated in the next section.

Algorithm 2.2 Enumerate all the vertices of a polytope $\mathcal{A}$

1: Initialize vertex set $\mathcal{V} = \emptyset$
2: Generate set $\tilde{\mathcal{A}} := \{\tilde{\mathbf{a}} \in \mathbb{R}^n | \tilde{a}_i = \underline{a}_i \text{ or } \overline{a}_i,\ i = 1, \ldots, n\}$; check the feasibility of all the points in set $\tilde{\mathcal{A}}$, i.e., if $a^{\min} \leq \mathbf{1}'\tilde{\mathbf{a}} \leq a^{\max}\}$, then $\mathcal{V} = \mathcal{V} \cup \{\tilde{\mathbf{a}}\}$
3: Generate set $\hat{\mathcal{A}} := \{\hat{\mathbf{a}} \in \mathbb{R}^n | \hat{a}_i = a^{\min} - \sum_{j \neq i} \hat{a}_j \text{ or } a^{\max} - \sum_{j \neq i} \hat{a}_j,\ \hat{a}_j = \underline{a}_j \text{ or } \overline{a}_j,\ i, j = 1, \ldots, n, j \neq i\}$; check the feasibility of all the points in set $\hat{\mathcal{A}}$, i.e., if $\underline{\mathbf{a}} \preceq \hat{\mathbf{a}} \preceq \overline{\mathbf{a}}$, then $\mathcal{V} = \mathcal{V} \cup \{\hat{\mathbf{a}}\}$

Algorithm 2.3 Enumerate all the vertices of a polytope $\mathcal{B}$

1: **for** $i = 1, 2, \ldots, S$ **do**
2: Obtain vertex set $\mathcal{V}_s$ by applying Algorithm 2.2 to $\mathcal{B}_s$
3: **end for**
4: Generate vertices $\mathbf{b}^{\text{v}}$ for $\mathcal{B}$ by concatenating all the individual vertices $\mathbf{b}_s$ as $\mathbf{b}^{\text{v}} = [(\mathbf{b}_1^{\text{v}})', \ldots, (\mathbf{b}_S^{\text{v}})']',\ \mathbf{b}_s \in \mathcal{V}_s$

2.3 Vertex Enumeration Algorithms

In order to obtain $\mathbf{w}^*$, the convex nondifferentiable function in (2.25) should be maximized over $\mathcal{W}$. This is generally an NP-hard convex maximization problem, meaning the global optimum $\{W_*^t\}$ can not be obtained in polynomial time. However, for the specific problem here, the special structure of the problem can be utilized to obtain a computationally efficient approach. Specifically, the global solution is attained at the extreme points of the polytope [13, Sec. 2.4]. Therefore, the objective in (2.25) can be evaluated at all (finitely many) vertices of $\mathcal{W}$ to obtain the global solution.

For the polytopes $\mathcal{W}$ with special structure [cf. (2.4), (2.5)], characterizations of vertices are established in Propositions 2.3 and 2.4. Capitalizing on these propositions, vertex enumerating procedures are designed consequently, and are tabulated as Algorithms 2.2 and 2.3. To this end, the following lemma is first needed, which establishes the necessary and sufficient condition for vertices of a polytope represented as a linear system [49, Sec. 3.5].

Lemma 2.1 *For a polytope $\mathcal{P} := \{\mathbf{x} \in \mathbb{R}^n | \mathbf{A}\mathbf{x} \preceq \mathbf{c}\}$, a point $\mathbf{v} \in \mathcal{P}$ is a vertex if and only if there exists a subsystem $\tilde{\mathbf{A}}\mathbf{x} \preceq \tilde{\mathbf{c}}$ of $\mathbf{A}\mathbf{x} \preceq \mathbf{c}$ so that rank$(\tilde{\mathbf{A}}) = n$ and $\mathbf{v}$ is the unique (feasible) solution of $\tilde{\mathbf{A}}\mathbf{v} = \tilde{\mathbf{c}}$.*

Proposition 2.3 *For a polytope $\mathcal{A} := \{\mathbf{a} \in \mathbb{R}^n | \underline{\mathbf{a}} \preceq \mathbf{a} \preceq \overline{\mathbf{a}}, a^{\min} \leq \mathbf{1}'\mathbf{a} \leq a^{\max}\}$, $\mathbf{a}^{\text{v}} \in \mathcal{A}$*

is a vertex (extreme point) of $\mathcal{A}$ if and only if it has one of the following forms: i) $a_i^{\rm v} = \underline{a}_i$ or $\overline{a}_i$ for $i = 1, \ldots, n$; or ii) $a_i^{\rm v} = a^{\min} - \sum_{j \neq i} a_j^{\rm v}$ or $a^{\max} - \sum_{j \neq i} a_j^{\rm v}$, $a_j^{\rm v} = \underline{a}_j$ or $\overline{a}_j$, for $i, j = 1, \ldots, n, j \neq i$.

Proof: The polytope $\mathcal{A} := \{\mathbf{a} \in \mathbb{R}^n | \underline{\mathbf{a}} \preceq \mathbf{a} \preceq \overline{\mathbf{a}}, a^{\min} \leq \mathbf{1}'\mathbf{a} \leq a^{\max}\}$ can be re-written as $\mathcal{A} := \{\mathbf{a} \in \mathbb{R}^n | \mathbf{A}\mathbf{a} \preceq \mathbf{c}\}$, where $\mathbf{A} := [\mathbf{I}_{n\times n}, -\mathbf{I}_{n\times n}, \mathbf{1}, -\mathbf{1}]'$ and $\mathbf{c} := [\overline{\mathbf{a}}', -\underline{\mathbf{a}}', a^{\max}, -a^{\min}]'$. By Lemma 2.1, enumerating vertices of $\mathcal{A}$ is equivalent to finding all feasible solutions of the linear subsystems $\tilde{\mathbf{A}}\mathbf{a} = \tilde{\mathbf{c}}$, such that rank-$n$ matrix $\tilde{\mathbf{A}}$ is constructed by extracting rows of $\mathbf{A}$. It can be seen that such full column-rank matrix $\tilde{\mathbf{A}}$ can only have two forms (with row permutation if necessary): i) $\tilde{\mathbf{A}}_1 = \text{diag}(\mathbf{d})$ with $d_i \in \{-1, 1\}$, $i = 1, \ldots, n$; ii) $\tilde{\mathbf{A}}_2(i,:) = \pm\mathbf{1}'$, $i = 1, \ldots, n$, and $\tilde{\mathbf{A}}_2(j,:) = \tilde{\mathbf{A}}_1(j,:)$, $\forall j \neq i$. Basically, $\tilde{\mathbf{A}}_1$ is constructed by choosing n vectors as a basis of $\mathbb{R}^n$ from the first $2n$ rows of $\mathbf{A}$. Substituting any row of $\tilde{\mathbf{A}}_1$ with $\pm\mathbf{1}'$, forms $\tilde{\mathbf{A}}_2$. Finally, by solving all the linear subsystems of the form $\tilde{\mathbf{A}}_k\mathbf{a} = \tilde{\mathbf{c}}_k$, for $k = 1, 2$, Proposition 2.3 follows readily. □

Essentially, Proposition 2.3 verifies the geometric characterization of vertices. Since $\mathcal{W}$ is the part of a hyperrectangle (orthotope) between two parallel hyperplanes, its vertices can only either be the hyperrectangle's vertices which are not cut away, or, the vertices of the intersections of the hyperrectangle and the hyperplanes, which must appear in some edges of the hyperrectangle.

Next, the vertex characterization of a polytope in a Cartesian product formed by many lower-dimensional polytopes like $\mathcal{A}$ is established, which is needed for the uncertainty set (2.4).

Proposition 2.4 *Assume $\mathbf{b} \in \mathbb{R}^n$ is divided into S consecutive and non-overlapping blocks as $\mathbf{b} = [\mathbf{b}_1', \ldots, \mathbf{b}_S']'$, where $\mathbf{b}_s \in \mathbb{R}^{n_s}$ and $\sum_{s=1}^S n_s = n$. Consider a polytope $\mathcal{B} := \{\mathbf{b} \in \mathbb{R}^n | \underline{\mathbf{b}} \preceq \mathbf{b} \preceq \overline{\mathbf{b}}, b_s^{\min} \leq \mathbf{1}_{n_s}'\mathbf{b}_s \leq b_s^{\max}, s = 1, \ldots, S\}$. Then $\mathbf{b}^{\rm v} = [(\mathbf{b}_1^{\rm v})', \ldots, (\mathbf{b}_S^{\rm v})']'$ is a vertex of $\mathcal{B}$ if and only if for $s = 1, \ldots, S$, $\mathbf{b}_s^{\rm v}$ is the vertex of a lower-dimensional polytope $\mathcal{B}_s := \{\mathbf{b}_s \in \mathbb{R}^{n_s} | \underline{\mathbf{b}}_s \preceq \mathbf{b}_s \preceq \overline{\mathbf{b}}_s, b_s^{\min} \leq \mathbf{1}_{n_s}'\mathbf{b}_s \leq b_s^{\max}\}$.*

Proof: The polytope $\mathcal{B} := \{\mathbf{b} \in \mathbb{R}^n | \underline{\mathbf{b}} \preceq \mathbf{b} \preceq \overline{\mathbf{b}}, b_s^{\min} \leq \mathbf{1}_{n_s}'\mathbf{b}_s \leq b_s^{\max}, s = 1, \ldots, S\}$ can be re-written as $\mathcal{B} := \{\mathbf{b} \in \mathbb{R}^n | \mathbf{B}\mathbf{b} \preceq \mathbf{c}\}$, where $\mathbf{B} := \text{diag}(\mathbf{B}_1, \ldots, \mathbf{B}_S)$, $\mathbf{c} := [\mathbf{c}_1', \ldots, \mathbf{c}_S']'$, $\mathbf{B}_s := [\mathbf{I}_{n_s\times n_s}, -\mathbf{I}_{n_s\times n_s}, \mathbf{1}, -\mathbf{1}]'$, and $\mathbf{c}_s := [\overline{\mathbf{a}}_s', -\underline{\mathbf{a}}_s', b_s^{\max}, -b_s^{\min}]'$ for $s = 1, \ldots, S$.

Table 2.1: Generating capacities, ramping limits, and cost coefficients. The units of a_m and b_m are \$/(kWh)^2 and \$/kWh, respectively.

Unit	$P_{G_m}^{\min}$	$P_{G_m}^{\max}$	$R_{m,\text{up(down)}}$	a_m	b_m
1	10	50	30	0.006	0.5
2	8	45	25	0.003	0.25
3	15	70	40	0.004	0.3

Similarly by Lemma 2.1, all the vertices of $\mathcal{B}$ can be enumerated by solving $\tilde{\mathbf{B}}\mathbf{b} = \tilde{\mathbf{c}}$, where the rank-$n$ matrix $\tilde{\mathbf{B}}$ is formed by extracting rows of $\mathbf{B}$. Due to the block diagonal structure of $\mathbf{B}$, it can be seen that the only way to find its n linear independent rows is to find n_s linear independent vectors from the rows corresponding to $\mathbf{B}_s$ for $s = 1, \ldots, S$. In other words, the vertices $\mathbf{b}^{\text{v}}$ can be obtained by concatenating all the individual vertices $\mathbf{b}_s$ as stated in Proposition 2.4. □

Algorithms 2.2 and 2.3 can be used to to generate the vertices of uncertainty sets (2.4) and (2.5) as described next.

i) For uncertainty set (2.4), first use Algorithm 2.2 to obtain the vertices corresponding to each sub-horizon $\mathcal{T}_{i,s}$ for all the RES facilities. Then, concatenate the obtained vertices to get the ones for each RES facility by Step 4 in Algorithm 2.3. Finally, run this step again to form the vertices of (2.4) by concatenating the vertices of each $\mathcal{W}_i$.

ii) For uncertainty sets (2.5), use Algorithm 2.2 to obtain the vertices for each sub-horizon $\mathcal{T}_s$. Note that concatenating step in Algorithm 2.3 is not needed in this case because problem (2.25) is decomposable across sub-horizons $\mathcal{T}_s$, $s = 1, \ldots, S$, and can be independently solved accordingly.

After the detailed description of vertex enumerating procedures for RES uncertainty sets, a discussion on the complexity of solving (2.25) follows.

Remark 2.2 *(Complexity of solving* (2.25)*).* Vertex enumeration incurs exponential complexity because the number of vertices can increase exponentially with the number of variables and constraints [16, Ch. 2]. However, if the cardinality of each sub-horizon $\mathcal{T}_s$ is not very large (e.g., when 24 hours are partitioned into 4 sub-horizons each comprising

Table 2.2: Class-1 dispatchable loads parameters. The units of c_n and d_n are \$/(kWh)^2 and \$/kWh, respectively.

	Load 1	Load 2	Load 3	Load 4	Load 5	Load 6
$P_{D_n}^{\min}$	0.5	4	2	5.5	1	7
$P_{D_n}^{\max}$	10	16	15	20	27	32
c_n	-0.002	-0.0017	-0.003	-0.0024	-0.0015	-0.0037
d_n	0.2	0.17	0.3	0.24	0.15	0.37

Table 2.3: Class-2 dispatchable loads parameters.

	Load 1	Load 2	Load 3	Load 4
$P_{E_q}^{\min}$	0	0	0	0
$P_{E_q}^{\max}$	1.2	1.55	1.3	1.7
$E_q^{\max}$	5	5.5	4	8
S_q	6PM	7PM	6PM	6PM
T_q	12AM	11PM	12AM	12AM

6 time slots), then the complexity is affordable. Most importantly, the vertices of $\mathcal{W}$ need only be listed once offline, before solving the optimization problem (2.9).

2.4 Numerical Tests

In this section, numerical results are presented to verify the performance of the robust and distributed energy scheduler. The Matlab-based modeling package `CVX` [59] along with the solver `MOSEK` [5] are used to specify and solve the proposed robust energy management problem. The considered microgrid consists of $M = 3$ conventional generators, $N = 6$ class-1 dispatchable loads, $Q = 4$ class-2 dispatchable loads, $J = 3$ storage units, and $I = 2$ renewable energy facilities (wind farms). The time horizon spans $T = 8$ hours, corresponding to the interval 4PM–12AM. The generation costs $C_m(P_{G_m}) = a_m P_{G_m}^2 + b_m P_{G_m}$ and the utilities of class-1 elastic loads $U_n(P_{D_n}) = c_n P_{D_n}^2 + d_n P_{D_n}$ are set to be quadratic and time-invariant. Generator parameters are given in Table 2.1, while $\mathsf{SR}^t = 10$kWh. The relevant parameters of two classes of dispatchable loads are listed in Tables 2.2 and 2.3 (see also [53]). The utility of class-2 loads is $U_{E_q}^t(P_{E_q}^t) := \pi_q^t P_{E_q}^t$ with weights $\pi_q^t = 4, 3.5, \ldots, 1, 0.5$ for $t = 4\text{PM}, \ldots, 11\text{PM}$ and $q \in \mathcal{Q}$.

Table 2.4: Limits of forecasted wind power.

Slot	1	2	3	4	5	6	7	8
$\underline{W}_1^t$	2.47	2.27	2.18	1.97	2.28	2.66	3.1	3.38
$\overline{W}_1^t$	24.7	22.7	21.8	19.7	22.8	26.6	31	33.8
$\underline{W}_2^t$	2.57	1.88	2.16	1.56	1.95	3.07	3.44	3.11
$\overline{W}_2^t$	25.7	18.8	21.6	15.6	19.5	30.7	34.4	31.1

Table 2.5: Fixed loads demand and transaction prices. The units of α^t and β^t are ¢/kWh.

Slot	1	2	3	4	5	6	7	8
L^t	57.8	58.4	64	65.1	61.5	58.8	55.5	51
α^t (Case A)	2.01	2.2	3.62	6.6	5.83	3.99	2.53	2.34
β^t (Case A)	1.81	1.98	3.26	5.94	5.25	3.59	2.28	2.11
α^t (Case B)	40.2	44	72.4	132	116.6	79.8	50.6	46.8
β^t (Case B)	36.18	39.6	65.16	118.8	104.94	71.82	45.54	42.12

Three batteries have capacity $B_j^{\max} = 30$kWh (similar to [62]). The remaining parameters are $P_{B_j}^{\min} = -10$kWh, $P_{B_j}^{\max} = 10$kWh, $B_j^0 = B_j^{\min} = 5$kWh, and $\eta_j = 0.95$, for all $j \in \mathcal{J}$. The battery costs $H_j^t(B_j^t)$ are set to zero. The joint uncertainty model with $S = 1$ is considered for $\mathcal{W}$ [cf. (2.5)], where $W_1^{\min} = 40$kWh, and $W_1^{\max} = 360$kWh. In order to obtain $\underline{W}_i^t$ and $\overline{W}_i^t$ listed in Table 2.4, day-ahead wind forecast data of the Midcontinent independent system operator (MISO) [93] are rescaled to the order of 1 kWh to 40 kWh, which is a typical wind power generation for a microgrid [147].

Similarly, the fixed load L^t in Table 2.5 is a rescaled version of the cleared load provided by MISO's daily report [50]. For the transaction prices, two different cases are studied as given in Table 2.5, where $\{\alpha^t\}$ in Case A are real-time prices of the Minnesota hub in MISO's daily report. To evaluate the effect of high transaction prices, $\{\alpha^t\}$ in Case B is set as 20 times of that in Case A. For both cases, $\beta^t = 0.9\alpha^t$, which satisfies the convexity condition for (2.7) given in Proposition 2.1.

First, convergence of the Lagrange multiplier $\{\lambda^t\}$ corresponding to the balance equation (2.7n) is confirmed for Case A by Fig. 2.2. It can be seen that λ^t converges for all $t \in \mathcal{T}$ within a couple of hundred iterations, which verifies the validity of the dual decomposition approach using the subgradient projection method. With the running

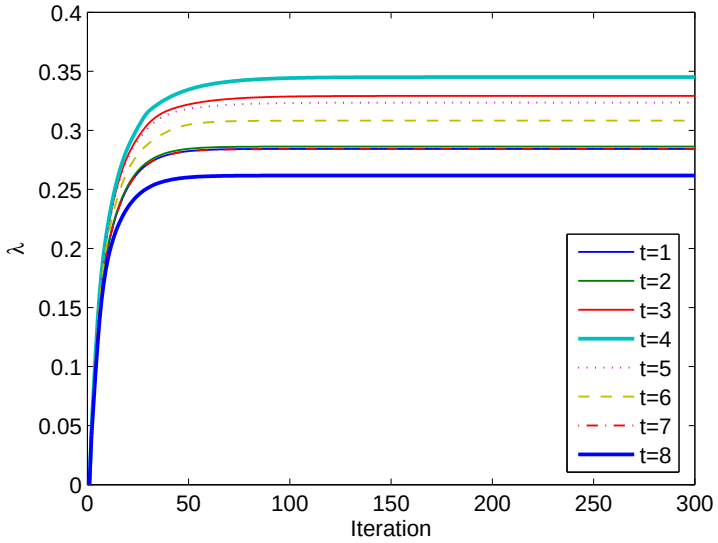


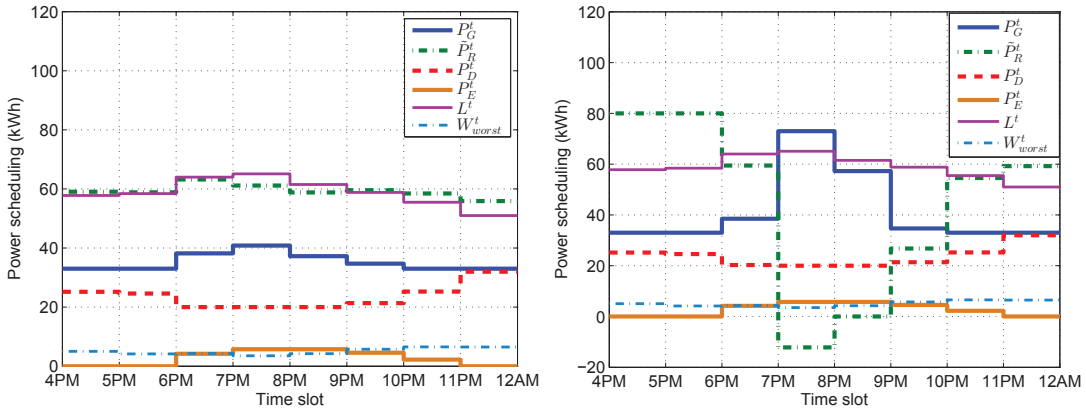
Figure 2.2: Convergence of $\{\lambda^t\}$.

averages, convergence of the other primal and dual variables as well as for Case B was also verified.

The optimal microgrid power schedules of two cases are shown in Figs. 2.3(a) and 2.3(b). The stairstep curves include $P_G^t := \sum_m P_{G_m}^t$, $P_D^t := \sum_n P_{D_n}^t$, and $P_E^t := \sum_q P_{E_q}^t$ denoting the total conventional power generation, and total elastic demand for classes 1 and 2, respectively, which are the optimal solutions of problem (2.9). Quantity W_{worst}^t denotes the total worst-case wind energy at slot t, which is the optimal solution of (2.25) with optimal $\tilde{P}_R^t$.

A common observation from Figs. 2.3(a) and 2.3(b) is that the total conventional power generation P_G^t varies with the same trend across t as the fixed load demand L^t, while the class-1 elastic load exhibits the opposite trend. Because the conventional generation and the power drawn from the main grid are limited, the optimal scheduling by solving (2.9) dispatches less power for P_D^t when L^t is large (from 6PM to 10PM), and vice versa. This behavior indeed reflects the load shifting ability of the proposed design for the microgrid energy management.

Furthermore, by comparing two cases in Figs. 2.3(a) and 2.3(b), it is interesting to illustrate the effect of the transaction prices. Remember that the difference between $\tilde{P}_R^t$



(a) Case A (b) Case B

Figure 2.3: Optimal power schedules.

and W^t_{worst} is the shortage power needed to purchase (if positive) or the surplus power to be sold (if negative), Figs. 2.3(a) shows that the microgrid always purchases energy from the main grid because $\tilde{P}^t_R$ is more than W^t_{worst}. This is because for Case A, the purchase price α^t is much lower than the marginal cost of the conventional generation [cf. Tables 2.1 and 2.5]. The economic scheduling decision is thus to reduce conventional generation while purchasing more power to keep the supply-demand balance. For Case B, since α^t is much higher than that in Case A, less power should be purchased which is reflected in the relatively small gap between $\tilde{P}^t_R$ and W^t_{worst} across time slots. It can also be seen that $\tilde{P}^t_R$ is smaller than W^t_{worst} from 7PM to 9PM, meaning that selling activity happens and is encouraged by the highest selling price β^t in these slots across the entire time horizon. Moreover, selling activity results in the peak conventional generation from 7PM to 9PM. Fig. 2.4 compares the optimal costs for the two cases. It can be seen that the optimal costs of conventional generation and worst-case transaction of Case B are higher than those of Case A, which can be explained by the higher transaction prices and the resultant larger DG output for Case B.

The optimal power scheduling of class-2 elastic load is depicted in Fig. 2.5 for Case A. Due to the start time S_q [cf. Table 2.3], zero power is scheduled for the class-2 load 1, 3, and 4 from 4PM to 6PM while from 4PM to 7PM for the load 2. The decreasing trend for all such loads is due to the decreasing weights $\{\pi^t_q\}$ from S_q to T_q, which is established from the fast charging motivation for the PHEVs, for example.

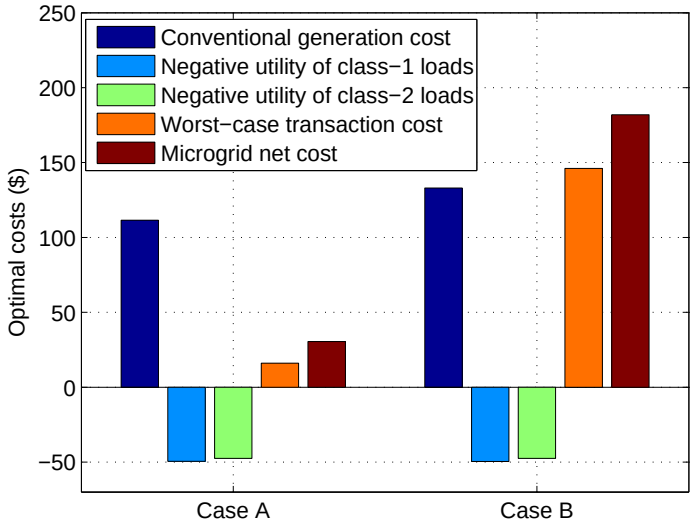


Figure 2.4: Optimal costs: Cases A and B.

Figs. 2.6 depicts the optimal charging or discharging power of the DSs for Case B. Clearly, all DSs are discharging during the three slots of 7PM, 8PM, and 9PM. This results from the motivation of selling more or purchasing less power because both purchase and selling prices are very high during these slots [cf. Table 2.5]. The charging (discharging) activity can also be reflected by the stored energy of the battery devices shown in Fig. 2.7. Note that, starting from the initial energy 5kWh at 4PM, the optimal stored energy of all units are scheduled to have 5kWh at 12AM, which satisfies the minimum stored energy requirement for the next round of scheduling time horizons.

Finally, Fig. 2.8 shows the effect of different selling prices $\{\beta^t\}$ on the optimal energy costs, where Case B is studied with fixed purchase prices $\{\alpha^t\}$. It can be clearly seen that the net cost decreases with the increase of the selling-to-purchase-price ratio β^t/α^t. When this ratio increases, the microgrid has a higher margin for revenue from the transaction mechanism, which yields the reduced worst-case transaction cost.

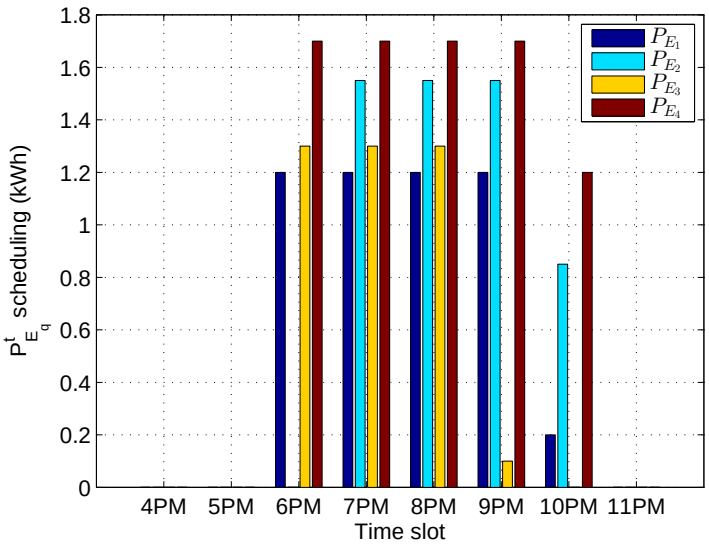


Figure 2.5: Optimal power schedule for $P^t_{E_q}$: Case A.

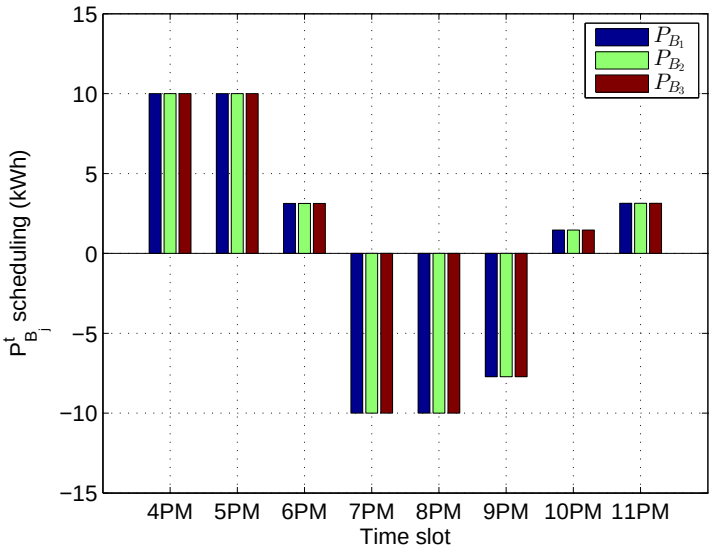


Figure 2.6: Optimal power schedule for $P^t_{B_j}$: Case B.

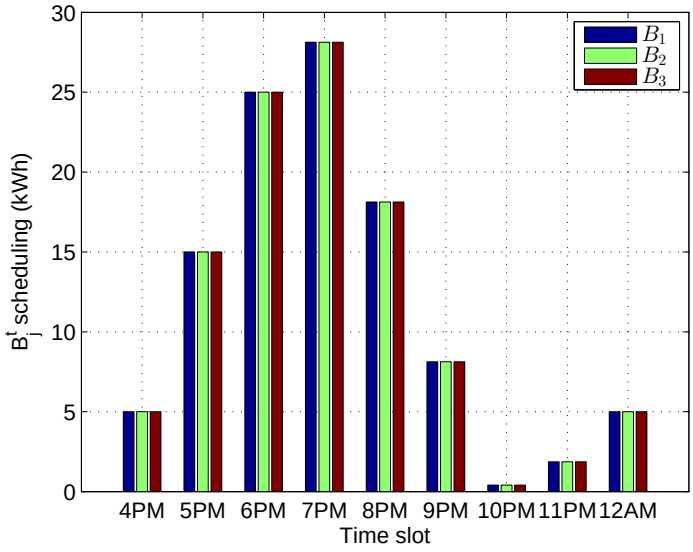


Figure 2.7: Optimal power schedule for B_j^t: Case B.

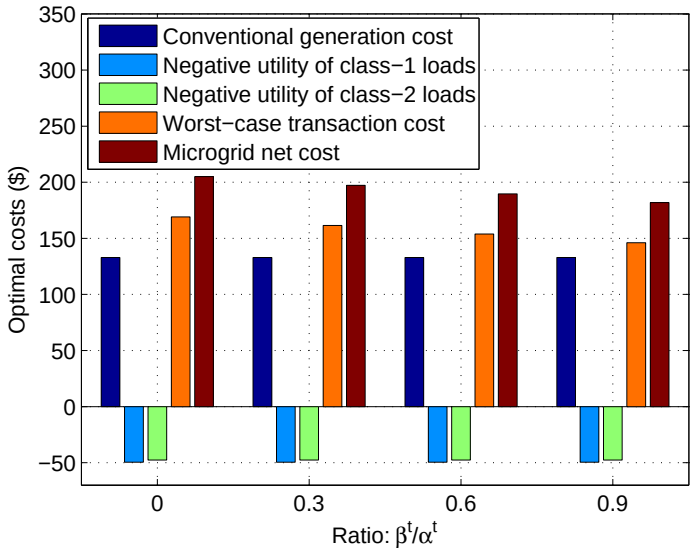


Figure 2.8: Optimal costs: Case B.

Chapter 3

Risk-Limiting Energy Management with Renewables

3.1 Risk-Constrained Economic Dispatch

Consider a power system comprising M conventional generators, N dispatchable loads, and I wind farms. The scheduling horizon is $\mathcal{T} := \{1, 2, \ldots, T\}$ (e.g., one day ahead). Let $P^t_{G_m}$ be the power produced by the mth conventional generator, and $P^t_{D_n}$ the power consumed by the nth dispatchable load at slot t, where $m \in \mathcal{M} := \{1, \ldots, M\}$, $n \in \mathcal{N} := \{1, \ldots, N\}$, and $t \in \mathcal{T}$. Besides dispatchable loads, there is also a fixed demand from critical loads, denoted by L^t. The actual wind power generated by the ith wind farm at slot t is denoted by $W^t_i,\ i \in \mathcal{I} := \{1, \ldots, I\}$. The ensuing section describes the risk-limiting model capturing the stochastic nature of wind power W^t_i. Section 3.1.2 formulates the risk-constrained energy management problem, which boils down to optimally scheduling the variables $P^t_{G_m}$ and $P^t_{D_n}$ for all $m \in \mathcal{M}$, $n \in \mathcal{N}$, and $t \in \mathcal{T}$.

3.1.1 Loss-of-Load Probability

In the analysis of the power system operations, the LOLP is often utilized as a statistical metric evaluating how often the system generating capacity cannot meet the total load demand during a given period. The supply-demand imbalance typically results from

uncertainties inherent to generators and loads (e.g., sudden loss, derating of generation, or, sudden variation of the load). Moreover, with the envisioned tide of high penetration wind power, a new uncertainty factor appears because of the intermittent nature of the wind. It is worth mentioning that supply-demand balance is maintained by the automatic generation control mechanism in the real-time operation (seconds timescale level) [73]. However, for larger timescale levels (e.g., one-day ahead ED), the supply-demand imbalance probability should be considered in order to make economic and risk-limiting power planning decisions.

Consider the supply shortage function at time t defined as

$$g^t(\mathbf{p}_G^t, \mathbf{p}_D^t, \mathbf{p}_R^t) := L^t + \sum_{n=1}^{N} P_{D_n}^t - \sum_{m=1}^{M} P_{G_m}^t - \sum_{i=1}^{I} W_i^t \tag{3.1}$$

where for the time slot t, vectors $\mathbf{p}_G^t$, $\mathbf{p}_D^t$, and $\mathbf{p}_R^t$ collect $p_{G_m}^t$, $p_{D_n}^t$, and W_i^t over all $m \in \mathcal{M}$, $n \in \mathcal{N}$, and $i \in \mathcal{I}$, respectively. Therefore, the LOLP constraint at time t can be equivalently written as

$$\text{Prob}\{g^t(\mathbf{p}_G^t, \mathbf{p}_D^t, \mathbf{p}_R^t) \leq 0\} \geq 1 - \alpha \tag{3.2}$$

where $\text{Prob}\{A\}$ denotes the probability of an event A; and $\alpha \in [0, 1]$ is a pre-selected tolerance denoting the LOLP threshold, which should be chosen small (e.g., 1–5%) for a practical risk-limiting energy management. Equation (3.2) can also be interpreted as the per-slot reliability; that is, satisfaction-of-load probability (SOLP) must be greater than or equal to $1 - \alpha$.

In the context of multi-period energy management, it is important to relate the notions of joint vis-à-vis per-slot SOLP, which are elaborated in the next subsection.

With the function (3.1), the joint SOLP can be defined as

$$\text{Prob}\left\{g^1(\mathbf{p}_G^1, \mathbf{p}_D^1, \mathbf{p}_R^1) \leq 0, \ldots, g^T(\mathbf{p}_G^T, \mathbf{p}_D^T, \mathbf{p}_R^T) \leq 0\right\} \geq 1 - \alpha. \tag{3.3}$$

Clearly, the joint SOLP expression (3.3) reflects the desire of the power system operator to have joint probability of load-satisfaction events for all time slots $t = 1, \ldots, T$, no less than a threshold close to 1. For notational brevity, define a vector-valued function $\mathbf{g}(\cdot)$ as

$$\mathbf{g}(\mathbf{p}_G, \mathbf{p}_D, \mathbf{p}_R) := (g^1(\mathbf{p}_G^1, \mathbf{p}_D^1, \mathbf{p}_R^1), \ldots, g^T(\mathbf{p}_G^T, \mathbf{p}_D^T, \mathbf{p}_R^T))' \tag{3.4}$$

where $\mathbf{p}_G$, $\mathbf{p}_D$, and $\mathbf{p}_R$ collect $\mathbf{p}_G^t$, $\mathbf{p}_D^t$, and $\mathbf{p}_R^t$ across the entire time $t \in \mathcal{T}$, respectively. ($\mathbf{a}'$ denotes the transpose of the vector $\mathbf{a}$). The joint SOLP (3.3) can be recast as [cf. (3.4)]

$$\text{Prob}\{\mathbf{g}(\mathbf{p}_G, \mathbf{p}_D, \mathbf{p}_R) \preceq \mathbf{0}\} \geq 1 - \alpha \tag{3.5}$$

where the notation $\preceq$ denotes element-wise inequality.

An important relationship between joint SOLP (3.3) and per-slot SOLP (3.2) is established in the following proposition.

Proposition 3.1 *If the joint SOLP* (3.3) *holds, then the per-slot SOLP* (3.2) *also holds for all* $t \in \mathcal{T}$. *Moreover, if* $\{\mathbf{p}_R^t\}_{t=1}^T$ *are independent and identically distributed (i.i.d.) across time, then each per-slot SOLP is lower-bounded by* $(1-\alpha)^{1/T}$.

Proof: Using the fact that $\text{Prob}\{A_1 A_2 \dots A_n\} \leq \min\limits_{i=1,\dots,n} \{\text{Prob}\{A_i\}\}$, it can be seen that

$$\min_{t=1,\dots,T} \{\text{Prob}\{g^t(\mathbf{p}_G^t, \mathbf{p}_D^t, \mathbf{p}_R^t) \leq 0\}\} \geq \text{Prob}\{\mathbf{g}(\mathbf{p}_G, \mathbf{p}_D, \mathbf{p}_R) \preceq \mathbf{0}\} \geq 1 - \alpha. \tag{3.6}$$

If $\{\mathbf{p}_R^t\}_{t=1}^T$ are i.i.d., then for all $t \in \mathcal{T}$, it follows that

$$\begin{aligned} \text{Prob}\{\mathbf{g}(\mathbf{p}_G, \mathbf{p}_D, \mathbf{p}_R) \preceq \mathbf{0}\} &= \Pi_{t=1}^T \text{Prob}\{g^t(\mathbf{p}_G^t, \mathbf{p}_D^t, \mathbf{p}_R^t) \leq 0\} \\ &= \left[\text{Prob}\{g^t(\mathbf{p}_G^t, \mathbf{p}_D^t, \mathbf{p}_R^t) \leq 0\}\right]^T \end{aligned}$$

based on which the following lower bound is obtained

$$\text{Prob}\{g^t(\mathbf{p}_G^t, \mathbf{p}_D^t, \mathbf{p}_R^t) \leq 0\} \geq (1-\alpha)^{1/T}, \ \forall\ t \in \mathcal{T}.$$

□

Remark 3.1 *(SOLP for the i.i.d. and distribution-free cases).* If wind power production across time is i.i.d., the per-slot SOLP lower bound $(1-\alpha)^{1/T}$ is increasing in T. Leveraging this property, each per-slot LOLP decreases (goes to 0) as the total scheduling time T increases (goes to infinity). In fact, for a very small value of α, Taylor's expansion implies the lower bound $(1-\alpha)^{1/T} \approx 1 - \alpha/T$. Unfortunately, the i.i.d. condition of wind power across time is very strict and not practical in real power systems since typically the wind speed (and hence the wind power) is correlated across time.

However, the per-slot lower bound in (3.6) is distribution-free, meaning that this bound always holds no matter what the distributions of $\{\mathbf{p}_R^t\}_{t=1}^T$ are and whether or not they are independent. Hence, if the joint SOLP can be satisfied with probability $1-\alpha$, then the per-slot SOLP can be also guaranteed to be at least $1-\alpha$. Therefore, for the energy management optimization problem which is formulated in the ensuing section, it suffices to include just a single joint SOLP risk constraint, instead of multiple per-slot SOLP constraints in terms of the system reliability consideration.

3.1.2 Social Cost Minimization

Let $C_m^t(P_{G_m}^t)$ and $U_n^t(P_{D_n}^t)$ denote the cost of the mth conventional generator and the utility function of the nth dispatchable load, respectively. Typically, the increasing function $C_m^t(P_{G_m}^t)$ ($U_n^t(P_{D_n}^t)$) is chosen either convex (concave) quadratic or piecewise linear.

Energy management with multiple wind farms amounts to minimizing the power system net cost, which is the cost of conventional generation minus the load utility:

$$\underset{\{\mathbf{p}_G,\mathbf{p}_D\}}{\text{minimize}} \quad \sum_{t=1}^{T}\left(\sum_{m=1}^{M} C_m^t(P_{G_m}^t) - \sum_{n=1}^{N} U_n^t(P_{D_n}^t)\right) \tag{3.7a}$$

$$\text{subject to} \quad P_{G_m}^{\min} \le P_{G_m}^t \le P_{G_m}^{\max},\ m\in\mathcal{M},\ t\in\mathcal{T} \tag{3.7b}$$

$$P_{G_m}^t - P_{G_m}^{t-1} \le \mathsf{R}_m^{\text{up}},\ m\in\mathcal{M},\ t\in\mathcal{T} \tag{3.7c}$$

$$P_{G_m}^{t-1} - P_{G_m}^t \le \mathsf{R}_m^{\text{down}},\ m\in\mathcal{M},\ t\in\mathcal{T} \tag{3.7d}$$

$$P_{D_n}^{\min} \le P_{D_n}^t \le P_{D_n}^{\max},\ n\in\mathcal{N},\ t\in\mathcal{T} \tag{3.7e}$$

$$\text{Prob}\{\mathbf{g}(\mathbf{p}_G,\mathbf{p}_D,\mathbf{p}_R)\preceq\mathbf{0}\}\ge 1-\alpha. \tag{3.7f}$$

Constraints (3.7b)–(3.7d) stand for the power generation bounds and ramping up/down limits, capturing the typical physical constraints of power generation systems. Constraints (3.7e) correspond to the minimum/maximum limits of the dispatchable load demand. Constraint (3.7f) is the *risk-limiting* SOLP constraint that was defined and analyzed in the previous section. Mathematically, problem (3.7) is a so-called chance constrained program, which is widely used for dealing with random parameters in optimization problems [19].

Risk-constrained ED (3.7) arises naturally due to the high SOLP requirement. Regarding the possible transmission congestion, a risk-limiting ED with power flow constraints will be discussed in the ensuing section.

3.2 Risk-Constrained DC-OPF

Consider a power system with M buses. Again, let P_{G_m} denote the power output of a thermal generator and P_{D_m} the power dissipation of a load, both residing at bus m. While P_{G_m} is a decision variable, load P_{D_m} is considered fixed here for simplicity. Furthermore, if a renewable energy producer is located at bus m, two quantities will be associated with it: the predicted wind power generation W_m, and the power P_{R_m} scheduled to be injected to bus m. Note that the former is a random variable, whereas the latter is a decision variable. For notational simplicity, let the following M-dimensional vectors collect pertinent quantities $\mathbf{p}_G := [P_{G_1}, \ldots, P_{G_M}]'$, $\mathbf{p}_D := [P_{D_1}, \ldots, P_{D_M}]'$, $\mathbf{w} := [W_1, \ldots, W_M]'$, $\mathbf{p}_R := [P_{R_1}, \ldots, P_{R_M}]'$, $\mathbf{p}_G^{\min} := [P_{G_1}^{\min}, \ldots, P_{G_M}^{\min}]'$, and $\mathbf{p}_G^{\max} := [P_{G_1}^{\max}, \ldots, P_{G_M}^{\max}]'$. With these definitions, the nodal injections into the transmission grid can be succinctly expressed in a vector form as $\mathbf{p}_G + \mathbf{p}_R - \mathbf{p}_D$.

Focusing next on the transmission network, let L denote the number of lines in the grid and x_l the reactance of the l-th line. Define then the $L \times L$ diagonal matrix $\mathbf{D} := \text{diag}\left(\{x_l^{-1}\}_{l=1}^{N_l}\right)$; and the $L \times M$ branch-bus incidence matrix $\mathbf{A}$, such that if its l-th row $\mathbf{a}_l^T$ corresponds to the branch (m, n), then $[\mathbf{a}_l]_m := +1$, $[\mathbf{a}_l]_n := -1$, and zero elsewhere.

Flow conservation dictates that the aggregate power injected per bus should equal the power flowing away from the bus. The DC power flow model gives rise to the *nodal balance constraint* [58]

$$\mathbf{p}_G + \mathbf{p}_R - \mathbf{p}_D = \mathbf{B}\boldsymbol{\theta} \tag{3.8}$$

where $\boldsymbol{\theta} := [\theta_1, \ldots, \theta_M]'$ is the vector of nodal voltage phases $\{\theta_m\}_{m=1}^M$, and $\mathbf{B} := \mathbf{A}^T\mathbf{D}\mathbf{A}$ is the bus admittance matrix. Since the all-ones vector belongs to the nullspace of $\mathbf{B}$, the node balance equation (3.8) is invariant to nodal phase shifts. Hence, without loss of generality, the first bus can be the reference bus with phase set to zero, that is, $\theta_1 = 0$. The power flows on all transmission lines can be expressed as $\mathbf{H}\boldsymbol{\theta}$ with $\mathbf{H} := \mathbf{D}\mathbf{A}$. Physical considerations enforce a limit $\mathbf{f}^{\max}$ on the transmision power flows leading to

the *line flow constraint*

$$-\mathbf{f}^{\max} \preceq \mathbf{H}\boldsymbol{\theta} \preceq \mathbf{f}^{\max}. \tag{3.9}$$

Recall that the power system is dispatched several hours or even one day prior to the operation period of interest. Given a wind power generation forecast $\mathbf{w}$, the system operator wishes to schedule an injection $\mathbf{p}_R$ that is expected to be furnished. This requirement is captured here by allowing the vector inequality $\mathbf{w} \succeq \mathbf{p}_R$ to be violated with very low risk α. Specifically, the following chance constraint is imposed:

$$\text{Prob}\left(\mathbf{w} \succeq \mathbf{p}_R\right) \geq 1 - \alpha \tag{3.10}$$

Let $C_m(P_{G_m})$ denote the cost associated with the m-th thermal generator, which is convex and strictly increasing. The scheduling problem amounts to minimizing the total production cost subject to the constraints presented earlier, that is,

$$\begin{aligned}
\underset{\mathbf{p}_G, \mathbf{p}_R, \boldsymbol{\theta}}{\text{minimize}} \quad & \sum_{m=1}^{M} C_m(P_{G_m}) && \text{(3.11a)}\\
\text{subject to} \quad & \mathbf{p}_G + \mathbf{p}_R - \mathbf{p}_D = \mathbf{B}\boldsymbol{\theta} && \text{(3.11b)}\\
& -\mathbf{f}^{\max} \preceq \mathbf{H}\boldsymbol{\theta} \preceq \mathbf{f}^{\max} && \text{(3.11c)}\\
& \mathbf{p}_G^{\min} \preceq \mathbf{p}_G \preceq \mathbf{p}_G^{\max} && \text{(3.11d)}\\
& \theta_1 = 0 && \text{(3.11e)}\\
& \text{Prob}\left(\mathbf{w} \succeq \mathbf{p}_R\right) \geq 1 - \alpha. && \text{(3.11f)}
\end{aligned}$$

Formulation (3.11) extends to the DC-OPF problem (see e.g., [37]) to account for uncertain renewable energy injections. The scheduled renewable energy $\mathbf{p}_R$ is used as a basis for optimizing the power outputs of thermal generators based on (3.11b). The risk that the produced renewable energy will not be adequate to provide the scheduled one is limited as per constraint (3.11f). If during the actual system operation the harvested renewable energy exceeds the scheduled value, then curtailment is effected.

3.3 Sampling-based Scenario Approximation

In this section, we first develop numerically tractable convex approximation for problem (3.7). The novel approach can be naturally extended to solve (3.11), which will be discussed in Section (3.3.3).

Recall that the cost in problem (3.7) involves the convex (possibly non-differentiable) functions $C_m^t(\cdot)$ and $-U_n^t(\cdot)$, while the constraints (3.7b)–(3.7e) are linear. Consequently, the difficulty of problem (3.7) depends on the constraint (3.7f), whose tractability is discussed next.

3.3.1 Tractability Issue

Clearly, (3.7f) is not in a *computationally tractable* form. To convert it into a deterministically tractable form, the corresponding probability must be computable for a given distribution of the random vector $\mathbf{w}$. For a single wind turbine with single-period scheduling, this is possible [89]. However, for the multi-period power scheduling with multiple wind farms in problem (3.7), the joint distribution function of $\mathbf{w}$ is very hard to obtain, if not impossible.

One may also consider approximating $g^t(\cdot)$ as a Gaussian random variable for relatively large values of I, by appealing to the central limit theorem (CLT). Specifically, the total wind power at time t is obtained by aggregating the wind powers $\{W_i^t\}$ across all wind farms [cf. (3.1)]. Hence, the distribution of $\mathbf{g}(\cdot)$, which is approximated to be multivariate Gaussian, can be evaluated if time correlations are known. Note CLT relies on vanishing dependence of the random variables summed. Unfortunately, this does not hold for geographically close wind farms (e.g., in a microgird with many distributed renewable energy resources). In this case, wind speed is spatially correlated across different wind farms which are not far away from each other. Clearly, constraint (3.11f) in the risk-constrained DC-OPF problem has the same challenges.

3.3.2 Scenario Approximation for Economic Dispatch

As discussed in the previous subsection, in order to make problem (3.7) solvable, a computationally tractable replacement of the chance constraint (3.7f) must be employed, ideally of the convex type. Bypassing the challenges of the possible techniques described in Section 3.3.1, a straightforward heuristic method based on the *scenario approximation* is introduced, which turns out to be very efficient in solving problem (3.7). For convenience, consider first the following chance-constrained optimization problem in a

generic form [103]

$$\underset{\mathbf{x}\in X}{\text{minimize}} \quad f(\mathbf{x}) \tag{3.12a}$$

$$\text{subject to} \quad \text{Prob}\{\mathbf{h}(\mathbf{x},\boldsymbol{\xi}) \preceq \mathbf{0}\} \geq 1-\alpha \tag{3.12b}$$

where the real-valued objective $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function, and $X \subset \mathbb{R}^n$ is a nonempty convex set. Random vector $\boldsymbol{\xi}$ has probability distribution supported on a set $\Xi \subset \mathbb{R}^d$, and enters the problem through a vector-valued constraint function $\mathbf{h} : \mathbb{R}^n \times \Xi \to \mathbb{R}^m$.

As a general way to construct a tractable form of a chance constraint, the scenario approximation approach based on the Monte Carlo sampling technique amounts to generating S independent realizations of the random vector $\boldsymbol{\xi}$, denoted as $\boldsymbol{\xi}(1), \ldots, \boldsymbol{\xi}(S)$. Then, problem (3.12) is approximated as follows:

$$\underset{\mathbf{x}\in X}{\text{minimize}} \quad f(\mathbf{x}) \tag{3.13a}$$

$$\text{subject to} \quad \mathbf{h}(\mathbf{x},\boldsymbol{\xi}(s)) \preceq \mathbf{0},\ s = 1, \ldots, S. \tag{3.13b}$$

Note that a remarkable feature of this heuristic approach is that there are no specific requirements on the distribution of $\boldsymbol{\xi}$, or, on how it enters the constraints.

However, problem (3.13) itself is random in the sense that its solution varies with different sample realizations. Hence, the solution of problem (3.13) may not satisfy the original chance constraint (3.12b). Fortunately, facing this challenge, in [28] an elegant result established that regardless of the distribution of the random vector $\boldsymbol{\xi}$, if the sample size S is no less than the quantity of $\lceil 2n\alpha^{-1}\ln(2\alpha^{-1}) + 2\alpha^{-1}\ln(\delta^{-1}) + 2n \rceil$, then the optimal solution to problem (3.13) is feasible for the original problem (3.12) with probability at least $1-\delta$. ($\lceil a \rceil$ denotes the smallest integer greater than or equal to a.)

Applying this result to problem (3.7) with a prescribed LOLP risk level α, the sample size S should satisfy

$$S \geq S^* = \lceil 2T(M+N)\alpha^{-1}\ln(2\alpha^{-1}) + 2\alpha^{-1}\ln(\delta^{-1}) + 2T(M+N) \rceil. \tag{3.14}$$

The upshot of this sample bound is that it is distribution-free, which is particularly useful for multi-period power scheduling with multiple wind farms, because the joint

spatio-temporal distribution of the wind power is unknown. To this end, leveraging the scenario sampling approach, problem (3.7) can be approximated with the problem

$$\begin{aligned}
\underset{\{\mathbf{p}_G,\mathbf{p}_D\}}{\text{minimize}} \quad & \sum_{t=1}^{T}\left(\sum_{m=1}^{M} C_m^t(P_{G_m}^t) - \sum_{n=1}^{N} U_n^t(P_{D_n}^t)\right) && \text{(3.15a)}\\
\text{subject to} \quad & (3.7\text{b}) - (3.7\text{e}) \\
& L^t + \sum_{n=1}^{N} P_{D_n}^t - \sum_{m=1}^{M} P_{G_m}^t - \sum_{i=1}^{I} W_i^t(s) \le 0, \\
& \qquad s = 1,\ldots,S^*,\ t = 1,\ldots,T. && \text{(3.15b)}
\end{aligned}$$

Considering the sample bound (3.14), a potential drawback of the scenario approximation is that S^* grows linearly with the number of generators M and dispatchable loads N, as well as the scheduling time length T. Moreover, it is at least inversely proportional to the risk level α, which could augment problem (3.15) with many constraints, and hence render it difficult to solve. For example, for one-day ($T = 24$) ahead energy management of a small power system with $M = 2$ generators, $N = 4$ controllable loads, LOLP $\alpha = 0.05$, and feasibility risk $\delta = 0.05$, the bound results in $S^* = 21,656$. However, by inspecting the structure of the sampled constraints (3.15b) carefully, it is clear that problem (3.15) can be equivalent re-written as:

$$\begin{aligned}
\underset{\{\mathbf{p}_G,\mathbf{p}_D\}}{\text{minimize}} \quad & \sum_{t=1}^{T}\left(\sum_{m=1}^{M} C_m^t(P_{G_m}^t) - \sum_{n=1}^{N} U_n^t(P_{D_n}^t)\right) && \text{(3.16a)}\\
\text{subject to} \quad & (3.7\text{b}) - (3.7\text{e}) \\
& \sum_{n=1}^{N} P_{D_n}^t - \sum_{m=1}^{M} P_{G_m}^t \le \min_{s=1,\ldots,S^*}\left\{\sum_{i=1}^{I} W_i^t(s)\right\} - L^t,\ t \in \mathcal{T}. && \text{(3.16b)}
\end{aligned}$$

In practice, if the required minimum sample size S^* is very large, the value of $\min_{s=1,\ldots,S^*}\left\{\sum_{i=1}^{I} W_i^t(s)\right\}$ may become very small. In this case, the problem boils down to scheduling under the worst-case scenario where essentially the wind power output is zero. In order to avoid this situation, but still be able to guarantee the prescribed LOLP, a small positive quantity can be added to the right hand side of (3.16b). The effectiveness of this adjustment will be demonstrated numerically.

Remark 3.2 *(Algorithm scalability).* By exploiting how the random data $\mathbf{w}$ enter the

chance constraint (3.7f), i.e., the separability of the supply shortage function across $\mathbf{p}_G$, $\mathbf{p}_D$, and $\mathbf{w}$, the original S^* constraints per time instant, which can be potentially very large as shown in (3.14), are equivalently reformulated into just a single constraint. Therefore, by exploiting the problem structure, the seemingly intractable chance constraint (3.7f) is finally approximated by only T linear inequality constraints as in (3.16b). Essentially, an optimization problem with uncertain parameters is converted to a deterministic one without increasing the problem size. This sample-size-free feature yields a linear or convex quadratic program problem (3.16), which is efficiently solvable no matter how small α is.

3.3.3 Scenario Approximation for DC-OPF

It is not difficult to see that the scenario approximation developed in the previous section can be leveraged for solving (3.11f). Specifically, let $\{\mathbf{w}(s)\}_{s=1}^{S^*}$ denote S^* independent samples available. The scenario approximation approach relies on substituting (3.11f) with its sampled version

$$\mathbf{p}_R \preceq \mathbf{w}(s), \quad s = 1, \ldots, S^*. \tag{3.17}$$

Then, the optimization problem consisting of (3.11a)–(3.11d) and (3.17) is solved.

As (3.17) is an approximation of (3.11f), the question of whether the solution of the resultant optimization problem is feasible for the original problem is raised. In fact, notice that the solution of the approximate problem is a random variable, because the samples $\{\mathbf{w}(s)\}_{s=1}^{S^*}$ are random. Reference [28] develops a bound on the sample size S^* as a function of the risk level α which guarantees that the solution of the approximate problem is feasible for the original one with high probability.

Notice that (3.17) is linear in $\mathbf{p}_R$, a fact that renders the overall scheduling problem convex. On the negative side, the required S^* to achieve feasibility of the approximate solution is typically very large. This implies that the resultant optimization problem will have a very large number of constraints [cf. (3.17)], which may pose significant computational burden. It is possible to exploit the structure of the problem at hand, in order to overcome this difficulty and come up with a sample size free approximation. Specifically, it is not hard to see that (3.17) is equivalent to

$$P_{R_m} \leq \min_{s=1,\ldots,S^*} \{W_m(s)\}, \quad m = 1, \ldots, M. \tag{3.18}$$

A complication of this sampling mechanism is that the right-hand side of (3.18) can become very small as S^* grows. Recall that W_m is the power output of the mth renewable energy producer. As such, it is lower bounded by zero, and there is in fact nonzero probability that $W_m = 0$. This shortcoming can drive the decision variable P_{R_m} to very small values or even to zero. In a nutshell, there is a degree of conservatism inherent to the scenario approximation method.

A straightforward modification of (3.18) can alleviate the aforementioned conservatism. Specifically, a small quantity $\delta_m > 0$ can be added to $\min_{s=1,\ldots,S^*}\{W_m(s)\}$, in which case (3.18) is surrogated by

$$P_{R_m} \leq \min_{s=1,\ldots,S^*}\{W_m(s)\} + \delta_m, \quad m = 1, \ldots, M. \tag{3.19}$$

The effectiveness of this adjustment will be demonstrated numerically. With $\mathbf{w}^{\text{res}} = [W_1^{\text{res}}, \ldots, W_M^{\text{res}}]$ denoting the right-hand side of (3.19), the following problem is solved instead of (3.11):

$$\begin{aligned}
\underset{\mathbf{p}_G, \mathbf{p}_R, \boldsymbol{\theta}}{\text{minimize}} \quad & \sum_{m=1}^{M} C_m(P_{G_m}) && \text{(3.20a)}\\
\text{subject to} \quad & \mathbf{p}_G + \mathbf{p}_R - \mathbf{p}_D = \mathbf{B}\boldsymbol{\theta} && \text{(3.20b)}\\
& -\mathbf{f}^{\max} \preceq \mathbf{H}\boldsymbol{\theta} \preceq \mathbf{f}^{\max} && \text{(3.20c)}\\
& \mathbf{p}_G^{\min} \preceq \mathbf{p}_G \preceq \mathbf{p}_G^{\max} && \text{(3.20d)}\\
& \theta_1 = 0 && \text{(3.20e)}\\
& \mathbf{p}_R \preceq \mathbf{w}^{\text{res}}. && \text{(3.20f)}
\end{aligned}$$

To this end, solving problems (3.7) and (3.11) via scenario approximation requires S^* independent realizations of the random spatio-temporal correlated wind power output $\mathbf{w}$. This random sampling task is undertaken next.

3.3.4 Sampling Techniques

In addition to seasonal and diurnal trends, wind speed is clearly temporally correlated across short time horizons (e.g., a few hours). Moreover, since smart grids with multiple wind farms become more widespread, the spatial correlation of wind speed should be also taken into account. Generally, the spatial as well as temporal correlation of the wind

speed has considerable influence on the performance of power networks [44]. Therefore, in order to obtain the required samples of wind power outputs, a simple but effective approach proposed in [139] is briefly described in this section.

The *Weibull distribution* is the most widely accepted model for the stochastic wind speed V. Its advantage over alternative distributions (e.g., lognormal, generalized Gamma) has been well documented in a comprehensive review [30].

A (c, k)-parametrized Weibull random variable (RV) v can be generated from a standard normal RV y via the following Normal-to-Weibull transformation

$$v = c\left[-\ln\left(\frac{1}{2} - \frac{1}{2}\text{erf}\left(\frac{y}{\sqrt{2}}\right)\right)\right]^{\frac{1}{k}} \tag{3.21}$$

where the error function $\text{erf}(y)$ is defined as $\text{erf}(y) := \frac{2}{\sqrt{\pi}}\int_0^y \exp(-t^2)\,\mathrm{d}t$.

For simplicity, consider I decoupled AR(1) process, one for each wind farm output, as follows:

$$x_i^t = \phi_i x_i^{t-1} + \epsilon_i^t,\ i = 1, \ldots, I \tag{3.22}$$

where ϕ_i controls the temporal correlation of $\{x_i^t\}$. It is clear that x_i^t is standard normal, if the white noise ϵ_i^t is normally distributed with zero mean and variance $\sigma_{\epsilon_i}^2 = 1 - \phi_i^2$.

With decoupled AR models, the obtained $\mathbf{x}^t := [x_1^t, \ldots, x_I^t]$ is a Gaussian random vector with uncorrelated elements. A correlated Gaussian random vector $\mathbf{y}^t$ can thus be obtained by the linear transformation $\mathbf{y}^t = \mathbf{C}^{\frac{1}{2}}\mathbf{x}^t$ for any $t \in \mathcal{T}$, where the matrix $\mathbf{C} \in \mathbb{R}^{I\times I}$ is the desired correlation coefficient matrix of $\mathbf{y}^t$. Then, the spatio-temporal wind speed data $\{\mathbf{v}^t\}_{t=1}^T$ can be generated from $\{\mathbf{y}^t\}_{t=1}^T$ using the transformation (3.21). Finally, samples of the wind power output $\{\mathbf{w}(s)\}_{s=1}^{S^*}$ can be obtained by passing vectors $\{\mathbf{v}^t\}_{t=1}^T$ through a wind-turbine-specific mapping curve relating the wind speed to the wind power output [174]. In the ensuing simulation section, a simplified model will be utilized to implement the speed-to-power ($V \to W$) conversion as (see also [89])

$$W = \begin{cases} 0, & V < v_{\text{in}} \text{ or } V \geq v_{\text{out}} \\ \left(\frac{V - v_{\text{in}}}{v_{\text{rated}} - v_{\text{in}}}\right) w_{\text{rated}}, & v_{\text{in}} \leq V < v_{\text{rated}} \\ w_{\text{rated}}, & v_{\text{rated}} \leq V < v_{\text{out}} \end{cases}$$

where v_{in}, v_{rated}, and v_{out} represent the cut-in, rated, and cut-out wind velocity; and w_{rated} the rated wind power output.

Table 3.1: Generation limits, ramping rates, and cost coefficients. The units of a_m and b_m are \$/(kWh)^2 and \$/kWh, respectively.

Unit	$P_{G_m}^{\min}$	$P_{G_m}^{\max}$	R_m^{up}	$\mathsf{R}_m^{\text{down}}$	a_m	b_m
1	10	35	15	15	0.006	0.5
2	8	25	10	10	0.003	0.25
3	15	50	20	20	0.004	0.3

Table 3.2: Parameters of dispatchable loads. The units of c_n and d_n are \$/(kWh)^2 and \$/kWh, respectively.

Load	1	2	3	4	5	6
$P_{D_n}^{\min}$	1.5	3.3	2	5.7	4	9
$P_{D_n}^{\max}$	8	10	15	24	20	35
c_n	-0.0045	-0.0111	-0.0186	-0.0132	-0.0135	-0.0261
d_n	0.15	0.37	0.62	0.44	0.45	0.87

3.4 Numerical Tests

In this section, numerical tests are implemented to verify the performance of the proposed approaches. The simulation results for the risk-limiting ED (Sec. 3.1) and DC-OPF (Sec. 3.2) will be discussed in Sections 3.4.1 and 3.4.2, respectively. The Matlab-based package `CVX` [59] with solvers `SDPT3` [133] and `SeDuMi` [131] are used to solve the resulting convex problems.

3.4.1 Simulations for Economic Dispatch

The tested power system consists of $M = 3$ conventional generators, $N = 6$ dispatchable loads, and $I = 4$ wind farms. The scheduling horizon spans $T = 8$ hours, corresponding to the interval 4pm–12am. The generation costs $C_m(P_{G_m}) = a_m P_{G_m}^2 + b_m P_{G_m}$, and the utilities of controllable loads $U_n(P_{D_n}) = c_n P_{D_n}^2 + d_n P_{D_n}$ are set to be quadratic and time-invariant. The relevant parameters of generator and dispatchable loads are listed in Tables 3.1 and 3.2. The LOLP and feasibility risks are chosen to be $\alpha = \delta = 0.1$; the parameters of the Weibull distribution are $c = 10$ and $k = 2.2$, and $v_{\{\text{in,rated,out}\}} = 3, 14, 26$ m/s, $w_{\text{rated}} = 30$ kWh for the wind energy conversion; the fixed demand $L^t = [28.9, 29.2, 32, 32.55, 30.75, 29.4, 27.75, 25.5]$ kWh is the rescaled cleared load in a MISO daily report [50]; the lag-one temporal correlations are selected as

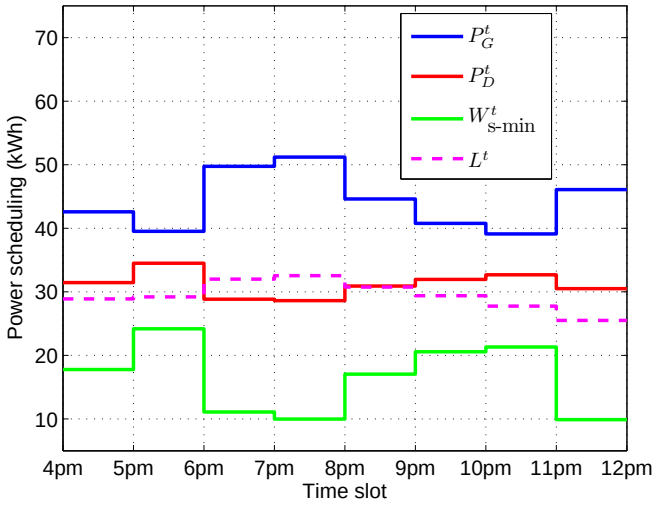


Figure 3.1: Optimal power schedule.

$\{\phi_i\}_{i=1}^I := \{0.15, 0.43, 0.67, 0.59\}$; and the spatial correlation coefficient matrix is set to

$$\mathbf{C} = \begin{bmatrix} 1 & 0.1432 & 0.4388 & -0.0455 \\ 0.1432 & 1 & -0.4555 & 0.8097 \\ 0.4388 & -0.4555 & 1 & -0.7492 \\ -0.0455 & 0.8097 & -0.7492 & 1 \end{bmatrix}.$$

Upon solving problem (3.16), the optimal power schedules are depicted in Fig. 3.1. The stairstep curves include $P_G^t := \sum_m P_{G_m}^t$ and $P_D^t = \sum_n P_{D_n}^t$ denoting the total conventional power and total elastic demand, respectively. Quantity $W^t_{\text{s-min}} := \min\limits_{s=1,\ldots,S^*} \left\{\sum_{i=1}^I W_i^t(s)\right\}$ denotes the minimum value of the total wind power over the required $S^* = 4,504$ samples. A key observation from Fig. 3.1 is that across time t, the total conventional power generation P_G^t varies complementarily to the one of the worst-sampled total wind power $W^t_{\text{s-min}}$. Clearly, this result makes intuitive sense since the conventional power P_G^t will decrease to reduce the conventional generation cost whenever the wind power $W^t_{\text{s-min}}$ is large (see e.g., the slot 5pm–6pm). Comparing P_G^t with $W^t_{\text{s-min}}$, it is clear that as the major supply source, the conventional power P_G^t should exhibit similar trend with the fixed loads demand L^t. Moreover, the elastic demand P_D^t exhibits opposite trend to the fixed demand L^t. This is because when L^t is low,

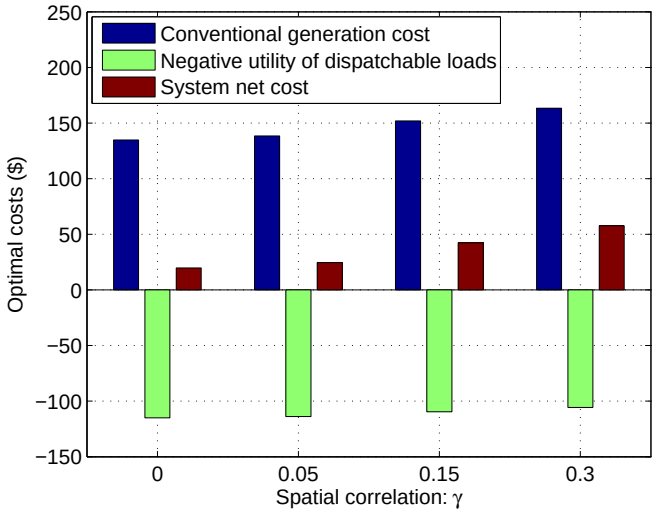


Figure 3.2: Optimal costs for different spatial correlation.

P_D^t increases to obtain more utility. This behavior indeed reflects the load adjustment ability of the proposed design.

Fig. 3.2 illustrates the effect of spatial correlation on the optimal cost. A number of correlation coefficient values $\gamma = \{0, 0.05, 0.15, 0.3\}$ is utilized for all wind-farm pairs; that is $\mathbf{C}_{ij} = \gamma,\ \forall\ i \neq j$. Clearly, the optimal net cost increases along with the spatial correlation. This can be explained by the effect of sampling as follows. Compared to the low-spatial correlation case, it is more likely that wind speeds at all wind farms happen to be small at the high-correlation scenario. Therefore, after the wind-speed-to-wind-power conversion, and with an increasing number of samples combined in the minimum-operation [cf. (3.16b)], $W_{\text{s-min}}^t$ in the high-correlation case will be smaller than that in the low-correlation case. This makes P_G^t and P_D^t more constrained in the high-correlation scenario (i.e., the feasible set is smaller), and yields a worse net cost. Thus, for the sampling approach to the risk-constrained energy management problem, the low correlation is in favor of obtaining a lower net cost due to the wind power output diversity effect.

Fig. 3.3 illustrates the effect of the LOLP risk level α on the optimal costs. As expected, the optimal net cost decreases as α increases. Because higher risk (LOLP) is allowed, less conventional power and more flexible demand will be scheduled, in order to

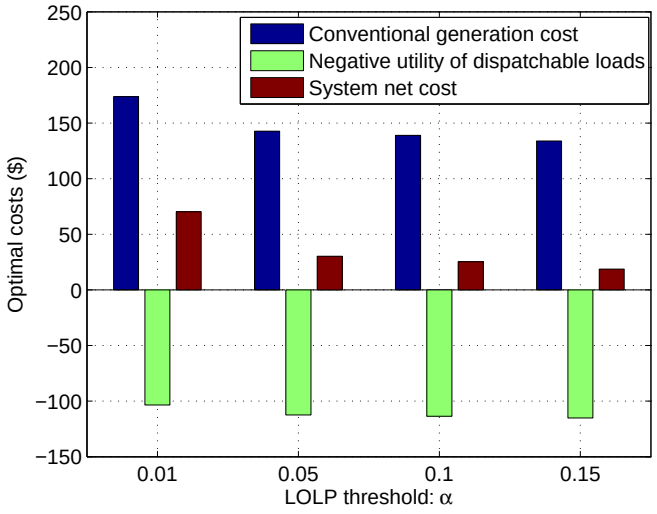


Figure 3.3: Optimal costs for different LOLP values.

reduce the generation cost and increase load utility. As in the discussion after problem (3.16), $W^t_{\text{s-min}}$ can be lifted by a small quantity in order to avoid the case of essentially worst-case scheduling (i.e., where no wind power is taken into account for the power scheduling). To this end, the wind speed v was increased by 2 m/s before converting it to the wind power in the aforementioned simulation setup. This boosting can be justified by examining the actual LOLP α_{ac} inferred from the probability in (3.3). The exact value of this probability can be simply approximated by its empirical counterpart, by evaluating the supply shortage function $\mathbf{g}(\mathbf{p}_G, \mathbf{p}_D, \mathbf{p}_R)$ for a large number of i.i.d. samples of $\{\mathbf{p}^t_R\}^T_{t=1}$, which should be different from the ones in (3.16b). Table 3.3 shows the validation results, which are obtained by checking (3.3) with 10^6 i.i.d. wind power samples, and the optimal power schedules $\{\mathbf{p}^*_G, \mathbf{p}^*_D\}$ obtained based on the prescribed LOLP α_{pr}. Clearly, the fact that the actual LOLPs are always smaller than the predefined ones verifies the validity of the boosting, and the effectiveness of the proposed scenario sampling approach for the risk-limited energy management.

Table 3.3: Prescribed LOLP α_{pr} versus actual LOLP α_{ac}.

α_{pr}	0.01	0.05	0.1	0.15
α_{ac}	0.0002	0.0346	0.0464	0.0739

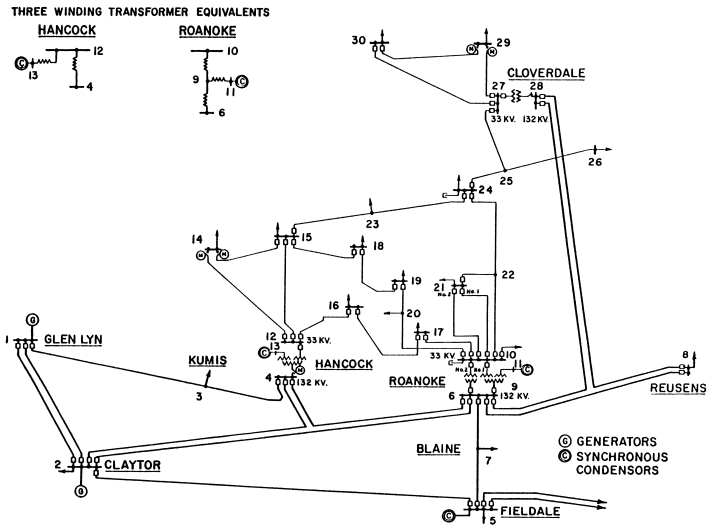


Figure 3.4: Seven wind farms have been added to the IEEE 30-bus grid system [138].

3.4.2 Simulations for DC-OPF

The performance of the novel scheduling approach is corroborated via numerical tests using the IEEE 30-bus power system [138]. The system includes 41 transmission lines and 6 conventional generators residing at buses $\{1, 2, 13, 22, 23, 27\}$ [cf. Fig. 3.4]. Load demands, generation costs, generator capacities, and transmission line ratings, are all specified in [175]. Seven wind farms have been added on buses $\{1, 2, 5, 9, 15, 24, 30\}$.

To simulate wind farm operation, real data originally provided for a wind energy forecasting competition organized by Kaggle platform were utilized [1]. Among other data, the specific dataset contains the actual hourly power output of seven wind farms over three years. To eliminate possible non-stationarities, only the interval from May 1^{st} to June 26^{th} of 2012 was considered, yielding a total of 589 hours due to missing entries.

Wind power outputs have been normalized per farm due to privacy concerns. To preserve the total installed generation capacity fixed after adding the wind farms, the conventional capacity is scaled down by 80%. Then, all wind farm outputs are scaled to contribute equally to the rest of the installed capacity, hence yielding a 20% wind energy penetration.

Recall that the developed scenario approximation-based scheduling requires drawing

Table 3.4: Optimal costs and LMPs for high-wind and low-wind scenarios ($\alpha = 0.05$).

Scenario	Cost	LMP
High-wind	481.42	364.97
Low-wind	565.21	378.91

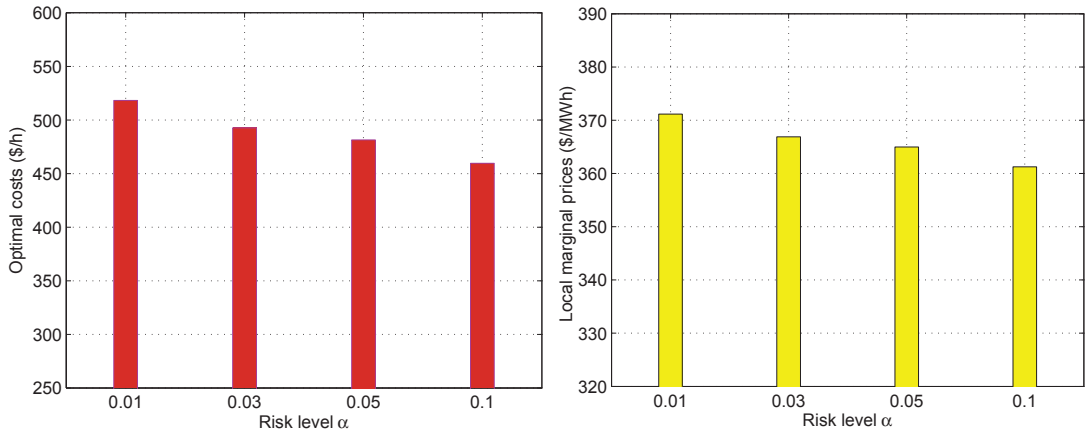
Table 3.5: Prescribed risk level and actual risk (high-wind scenario).

α	0.01	0.03	0.05	0.1
Actual risk	0.0072	0.0075	0.0076	0.0087

independent samples from the wind energy forecast $\mathbf{w}$. As a proof of concept, it is assumed here that $\mathbf{w}$ is Gaussian distributed. Its expected value is considered to be the actual wind power generated. A "low-wind" and a "high-wind" scenario were considered. The low-wind scenario yields $\boldsymbol{\mu}_l = [1.15\ 1.37\ 0.47\ 1.05\ 1.45\ 1.64\ 0.00]'$ and corresponds to May 19^{th} at 8 a.m. The high-wind scenario has $\boldsymbol{\mu}_h = [6.00\ 0.31\ 7.66\ 8.01\ 8.42\ 8.44\ 8.46]'$ and is observed on May 22^{nd} at 8 a.m. To model correlation across farms, it is further postulated that the covariance of $\mathbf{w}$ is that of the wind farm power outputs. The latter is empirically estimated as the sample covariance and it is denoted by $\hat{\boldsymbol{\Sigma}}$. Samples of $\mathbf{w}$ can then be drawn from $\mathcal{N}(\boldsymbol{\mu}_l, \hat{\boldsymbol{\Sigma}})$ and $\mathcal{N}(\boldsymbol{\mu}_h, \hat{\boldsymbol{\Sigma}})$, respectively for the two scenarios.

Before solving (3.20), the boosting parameters $\{\delta_m\}_{m=1}^M$ introduced in (3.19) must be selected. An intuitive and easily-implementable heuristic for doing so is described next. Instead of constraining P_{R_m} to be no larger than all samples $W_m(s)$ as dictated by (3.18), it is natural to require P_{R_m} to be no larger than only the $(1-\alpha)\%$ largest samples. Algorithmically, if $\left\{\{W_m^{[s]}\} \mid W_m^{[1]} \geq W_m^{[2]} \geq \cdots \geq W_m^{[S^*]}\right\}$ denote the order statistics of the original samples $\{W_m(s)\}_{s=1}^{S^*}$ for $m = 1, \ldots, M$, the right-hand side of (3.19) can be selected as $W_m^{\text{res}} = W_m^{[\lceil (1-\alpha)\times S^* \rceil]}$. Negative-valued entries of $\mathbf{w}^{\text{res}}$ are truncated to zero.

Dispatching the IEEE 30-bus power system for a risk level of $\alpha = 0.05$ yields the optimal costs listed in Table 3.4. The Lagrange multipliers corresponding to (3.20b), also known as *locational marginal prices* (LMPs), are also listed in the same table. LMPs are important components of electricity markets since they represent the cost of selling or buying electricity at a particular bus; see e.g., [82], [58]. Due to lack of transmission line congestion, all LMPs turn out to be equal to the value provided in Table 3.4.



(a) Optimal costs (b) LMPs

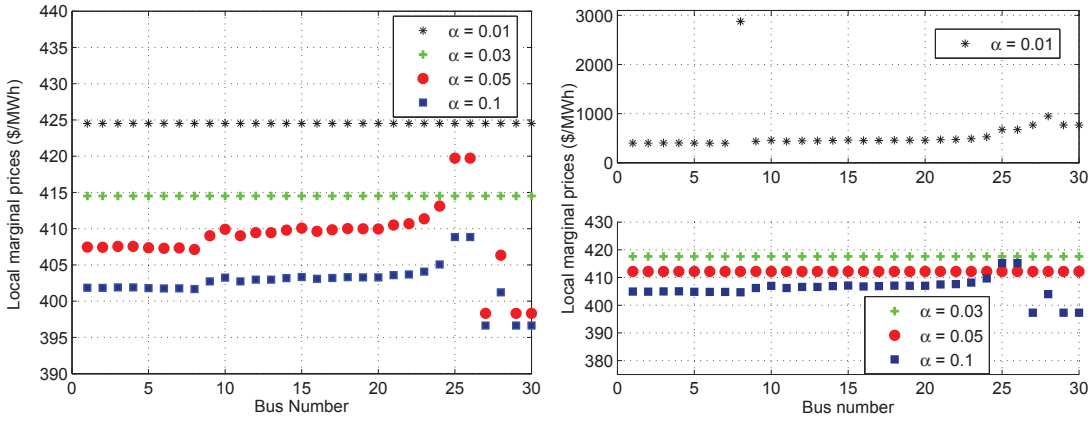
Figure 3.5: Effects of the prescribed risk level α (high-wind scenario)

The high-wind scenario attains lower cost and LMPs than the low-wind scenario, since less conventional power is needed when more free wind power is available. It is worth mentioning that due to the risk-aware constraint, the low-wind scenario essentially boils down to scheduling with no wind power at all.

Figs. 3.5(a) and 3.5(b) illustrate the effect of the prescribed risk level α on the optimal costs and the LMPs, respectively. The optimal net cost decreases with increasing α, since higher risk allows more wind power to be committed.

To justify the heuristic boosting procedure, the risk incurred by the optimal $\mathbf{p}_R$ that minimizes (3.20) is empirically evaluated by drawing 10^5 independent wind forecast samples $\mathbf{w}$, and checking whether (3.10) holds. Table 3.5 shows the validation results. The actual risk is always smaller than the predefined one, hence numerically validating the boosting step.

The effect of the risk level α on LMPs under transmission network congestion is investigated next. To simulate congestion, load demand at all buses is scaled up by β. The optimal costs listed in Table 3.6 decrease with decreasing β and/or increasing α, as expected. The corresponding 30 LMPs (one per bus) obtained when $\beta = 1.330$ and $\beta = 1.342$ and for varying values of α are plotted in Figs. 3.6(a) and 3.6(b), respectively. The two figures indicate that high risk levels result in lower prices in general. However, by varying α and β, different congestion patterns may occur due to the grid topology.



(a) $\beta = 1.330$ (b) $\beta = 1.342$

Figure 3.6: Locational marginal prices.

Table 3.6: Optimal costs for varying α and β (high-wind scenario).

α \ β	1.05	1.1	1.2	1.3
0.01	553.2936	589.0077	662.2263	738.4597
0.03	527.7588	565.2449	636.9299	712.1263
0.05	515.6234	549.3798	623.0269	697.1422
0.1	496.8349	530.5984	603.5701	677.1949

Chapter 4

Day-Ahead Electricity Market Clearing

4.1 Deterministic Market Clearing

In a day-ahead electricity market, participants including power generation companies and load service entities (LSEs) first submit their hourly supply bids and demand offers to market operators for the next operating day. Then, the ISO or regional transmission organization (RTO) clear the forward markets yielding least-cost unit commitment decisions, power dispatch outputs, and the corresponding DA clearing prices. The market clearing procedure proceeds in two stages. A security-constrained unit commitment (SCUC) is performed first by solving a large-scale mixed integer program to commit generation resources after simplifying or omitting transmission constraints. The second stage involves security-constrained economic dispatch (SCED) obtaining the economical power generation outputs and the LMPs as a byproduct. With unit commitment decisions fixed, SCED is usually in the form of DC-OPF, including the transmission network constraints [21].

Consider a power network comprising N_g generators, N_b buses, N_l lines, and N_a aggregators, each serving a large number of residential end-users with controllable smart appliances. The scheduling horizon of interest is $\mathcal{T} := \{1, 2, \ldots, T\}$ (e.g., one day ahead). Let $\mathbf{p}_G^t := [P_{G_1}^t, \ldots, P_{G_{N_g}}^t]'$ and $\mathbf{p}_{\text{DRA}}^t := [P_{\text{DRA}_1}^t, \ldots, P_{\text{DRA}_{N_a}}^t]'$ denote the

generator power outputs, and the power consumption of the aggregators at slot t, respectively. Define further the sets $\mathcal{N}_a := \{1, 2, \ldots, N_a\}$ and $\mathcal{N}_g := \{1, 2, \ldots, N_g\}$. Each aggregator $j \in \mathcal{N}_a$ serves a set $\mathcal{R}_j$ of residential users, and each user $r \in \mathcal{R}_j$ has a set $\mathcal{S}_{rj}$ of controllable smart appliances. Let $\mathbf{p}_{jrs}$ be the power consumption of smart appliance s and user r corresponding to aggregator j across the horizon. The power consumption $\mathbf{p}_{jrs}$ of each smart appliance across the horizon must typically satisfy operating constraints captured by a set $\mathcal{P}_{jrs}$, and may also give rise to user satisfaction represented by a concave utility function $U_{jrs}(\mathbf{p}_{jrs})$. Moreover, the generation cost is captured by convex functions $\{C_i(\cdot)\}_i$, and the fixed base load demands across the network buses at slot t is denoted by the vector $\mathbf{p}^t_{\text{BL}}$.

For brevity, vector $\mathbf{p}_0$ is used to collect all $p^t_{G_i}$, $P^t_{\text{DRA}_j}$, and network nodal angles θ^t_n; while vector $\{\mathbf{p}_j\}_{j\in\mathcal{N}_a}$ collects all smart appliance consumptions corresponding to aggregator j. With the goal of minimizing the system net cost, the DC optimal power flow (OPF) based MC stands as follows:

$$
\begin{aligned}
f^* = \underset{\{\mathbf{p}_j\}_{j=0}^{N_a}}{\text{minimize}} \quad & \sum_{t=1}^{T}\sum_{i=1}^{N_g} C_i(P^t_{G_i}) - \sum_{j=1}^{N_a}\sum_{r\in\mathcal{R}_j}\sum_{s\in\mathcal{S}_{jr}} U_{jrs}(\mathbf{p}_{jrs}) && \text{(4.1a)}\\
\text{subject to} \quad & \mathbf{A}_g\mathbf{p}^t_G - \mathbf{A}_a\mathbf{p}^t_{\text{DRA}} - \mathbf{p}^t_{\text{BL}} = \mathbf{B}_n\boldsymbol{\theta}^t,\ t\in\mathcal{T} && \text{(4.1b)}\\
& P^{\min}_{G_i} \le P^t_{G_i} \le P^{\max}_{G_i},\ i\in\mathcal{N}_g,\ t\in\mathcal{T} && \text{(4.1c)}\\
& -\mathsf{R}^{\text{down}}_i \le P^t_{G_i} - P^{t-1}_{G_i} \le \mathsf{R}^{\text{up}}_i,\ i\in\mathcal{N}_g,\ t\in\mathcal{T} && \text{(4.1d)}\\
& \mathbf{f}^{\min} \le \mathbf{B}_f\boldsymbol{\theta}^t \le \mathbf{f}^{\max},\ t\in\mathcal{T} && \text{(4.1e)}\\
& \theta^t_1 = 0,\ t\in\mathcal{T} && \text{(4.1f)}\\
& 0 \le P^t_{\text{DRA}_j} \le P^{\max}_{\text{DRA}_j},\ j\in\mathcal{N}_a,\ t\in\mathcal{T} && \text{(4.1g)}\\
& P^t_{\text{DRA}_j} = \sum_{r\in\mathcal{R}_j,\, s\in\mathcal{S}_{jr}} p^t_{jrs},\ j\in\mathcal{N}_a,\ t\in\mathcal{T} && \text{(4.1h)}\\
& \mathbf{p}_{jrs}\in\mathcal{P}_{jrs},\ r\in\mathcal{R}_j, s\in\mathcal{S}_{jr},\ j\in\mathcal{N}_a && \text{(4.1i)}
\end{aligned}
$$

where the nodal susceptance matrix $\mathbf{B}_n := -\mathbf{A}'_n\mathbf{B}_s\mathbf{A}_n \in \mathbb{R}^{N_b\times N_b}$ and the angle-to-flow matrix $\mathbf{B}_f := -\mathbf{B}_s\mathbf{A}_n \in \mathbb{R}^{N_l\times N_b}$. The ℓth row of the branch-node incidence matrix $\mathbf{A}_n \in \mathbb{R}^{N_l\times N_b}$ has 1 and -1 in its entry corresponding to the from and to nodes of branch ℓ, and 0 elsewhere; and the square diagonal matrix $\mathbf{B}_s := \text{diag}(b_1, \ldots, b_{N_l})$ is the branch susceptance matrix collecting the primitive susceptance across all branches.

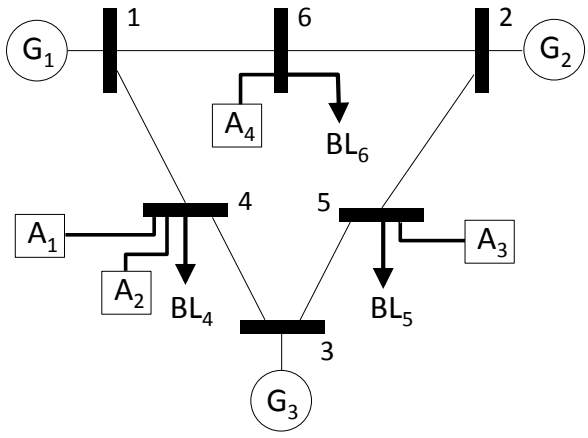


Figure 4.1: Power system example featuring 6 buses, 3 generators, 4 aggregators, and 3 base loads.

Matrices $\mathbf{A}_g \in \mathbb{R}^{N_b \times N_g}$ and $\mathbf{A}_a \in \mathbb{R}^{N_b \times N_a}$ in (4.1b) are the incidence matrices of the conventional generators and the aggregators, respectively. Take $\mathbf{A}_g$ as an example, $(\mathbf{A}_g)_{mn} = 1$ if the nth generator is injected to the mth bus, and $(\mathbf{A}_g)_{mn} = 0$, otherwise. Matrix $\mathbf{A}_a$ can be constructed likewise. A power network with $N_b = 6$, $N_l = 6$, $N_g = 3$, and $N_a = 4$ is shown in Fig. 4.1. For this adapted Western Electricity Coordinating Council (WECC) system [122], matrices $\mathbf{A}_g$ and $\mathbf{A}_a$ are given in the following forms

$$\mathbf{A}_g = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{A}_a = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Linear equality (4.1b) represents the *nodal balance* constraint. Limits of generator outputs and ramping rates are specified in constraints (4.1c) and (4.1d). Network line flow constraints are accounted for in (4.1e). Without loss of generality, the first bus can be set as the reference bus with zero phase (4.1f). Constraint (4.1g) captures the lower and upper bounds on the energy consumed by the aggregators. Equality (4.1h) amounts to the *aggregator-users power balance* equation; finally, (4.1i) gives the smart appliance constraints.

A smart appliance example is charging a PHEV, which typically amounts to consuming a prescribed total energy E_{jrs} over a specific horizon from a start time T^{st}_{jrs} to a termination time T^{end}_{jrs}. The consumption must remain within a range between $p^{\min}_{jrs}$ and $p^{\max}_{jrs}$ per period. With $\mathcal{T}^E_{jrs} := \{T^{\text{st}}_{jrs}, \ldots, T^{\text{end}}_{jrs}\}$, set $\mathcal{P}_{jrs}$ takes the form:

$$\mathcal{P}_{jrs} = \Bigg\{ \mathbf{p}_{jrs} \Bigg| \sum_{t \in \mathcal{T}^E_{jrs}} p^t_{jrs} = E_{jrs},\, p^t_{jrs} \in [p^{\min}_{jrs}, p^{\max}_{jrs}],\ \forall t \in \mathcal{T}^E_{jrs},$$
$$p^t_{jrs} = 0,\ \forall\ t \in \mathcal{T} \setminus \mathcal{T}^E_{jrs} \Bigg\}. \quad (4.2)$$

Further examples of $\mathcal{P}_{jrs}$ and $U_{jrs}(\mathbf{p}_{jrs})$ can be found in [54], where it is argued that $\mathcal{P}_{jrs}$ is a convex set for several appliance types of interest.

Problem (4.1) can be principally solved at the market operator (MO) in a central fashion. However, there are two major challenges when it comes to solving (4.1) with large-scale DR: i) functions $U_{jrs}(\mathbf{p}_{jrs})$ and sets $\mathcal{P}_{jrs}$ are private, and cannot be revealed to the MO; ii) including the sheer number of variables $\mathbf{p}_{jrs}$ would render the overall problem intractable for the MO, regardless of the privacy issue. The aggregator plays a critical role in successfully addressing these two challenges through decomposing the optimization tasks that arises, as detailed in the ensuing section.

4.2 Decomposition Algorithm

4.2.1 Dual Decomposition

Leveraging the dual decomposition technique, problem (4.1) can be decoupled into simpler subproblems tackled by the MO and the aggregators. Specifically, consider dualizing the linear coupling constraint (4.1h) with corresponding Lagrange multiplier μ^t_j. Upon straightforward re-arrangements, the partial Lagrangian can be written as

$$L(\{\mathbf{p}_j\}_{j=0}^{N_a}, \boldsymbol{\mu}) = L_0(\mathbf{p}_0, \boldsymbol{\mu}) + \sum_{j=1}^{N_a} L_j(\mathbf{p}_j, \boldsymbol{\mu}) \quad (4.3)$$

where

$$L_0(\mathbf{p}_0, \boldsymbol{\mu}) := \sum_{t=1}^{T} \left[\sum_{i=1}^{N_g} C_i(P^t_{G_i}) - \sum_{j=1}^{N_a} \mu^t_j P^t_{\text{DRA}_j} \right] \quad (4.4)$$

$$L_j(\mathbf{p}_j, \boldsymbol{\mu}) := \sum_{r\in\mathcal{R}_j}\sum_{s\in\mathcal{S}_{jr}}\left[\sum_{t=1}^{T}\mu_j^t p_{jrs}^t - U_{jrs}(\mathbf{p}_{jrs})\right]. \tag{4.5}$$

The dual function is thus obtained by minimizing the partial Lagrangian over the primal variables $\{\mathbf{p}_j\}_{j=0}^{N_a}$ as

$$D(\boldsymbol{\mu}) := D_0(\boldsymbol{\mu}) + \sum_{j=1}^{N_a} D_j(\boldsymbol{\mu}) \tag{4.6a}$$

$$= \min_{\text{s.t. (4.1b)–(4.1g)}} L_0(\mathbf{p}_0, \boldsymbol{\mu}) + \sum_{j=1}^{N_a} \min_{\text{s.t. (4.1i)}} L_j(\mathbf{p}_j, \boldsymbol{\mu}). \tag{4.6b}$$

The dual decomposition essentially iterates between two steps: S1) Lagrangian minimization with respect to $\{p_j\}_{j=0}^{N_a}$ given the current multipliers, and S2) multiplier update, using the obtained primal minimizers. It is clear from (4.3) that the Lagrangian minimization can be decoupled into $1 + N_a$ minimizations, where one is performed by the MO, and the rest by the corresponding aggregators.

Specifically, let $k = 1, 2, \ldots$ index iterations. Given the multipliers $\boldsymbol{\mu}(k)$, the subproblems at iteration k solved by the MO and each residential end-user are given as follows

$$\mathbf{p}_0(k) = \underset{\substack{\mathbf{p}_0 \\ \text{s. t. (4.1b)}-\text{(4.1g)}}}{\arg\min} \quad L_0(\mathbf{p}_0, \boldsymbol{\mu}(k)) \tag{4.7a}$$

$$\{\mathbf{p}_{jrs}(k)\}_s = \underset{\{\mathbf{p}_{jrs}\in\mathcal{P}_{jrs}\}_s}{\arg\min} \sum_{s\in\mathcal{S}_{jr}}\Big[\sum_{t=1}^{T}\mu_j^t(k) p_{jrs}^t - U_{jrs}(\mathbf{p}_{jrs})\Big]. \tag{4.7b}$$

Note that subproblem (4.7a) is a standard DC-OPF while the convex subproblem (4.7b) can be handled efficiently by the smart meters. In fact, with the feasible set in (4.2) and upon setting $U_{jrs}(\mathbf{p}_{jrs}) \equiv 0$, (4.7b) boils down to the *fractional knapsack* problem, which can be solved in closed form. To this end, the multipliers $\mu_j^t(k)$ needed can be transmitted to the user's smart meter via the AMI.

With the obtained quantities of $\mathbf{p}_0(k)$, $\{\mathbf{p}_{jrs}(k)\}_s$, and $\{D_j(\boldsymbol{\mu}(k))\}_{j=0}^{N_a}$, the ensuing section develops the approach to updating the multipliers $\{\mu_j^t\}_{j,t}$ using the so-termed bundle methods.

4.2.2 Multiplier Update via Bundle Methods

The choice of the multiplier update method is crucial, because fewer update steps imply less communications between the MO and the aggregators. A popular method of choice in the context of dual decomposition is the subgradient method, which is very slow typically. In this section, the bundle method with disaggregated cuts is proposed for the multiplier update. It is better suited to the problem of interest yielding faster convergence, because it exploits the special structure of the dual function which can be written as a sum of separate terms [cf. (4.6)], while it overcomes the drawbacks of the cutting plane one developed in [54]. Numerical tests in Section 4.5 illustrate differences in terms of convergence speed.

The following overview of the disaggregated bundle method in a general form is useful to grasp its role in the present context; see e.g., [13, Ch. 6] for detailed discussions. Consider the following separable convex minimization problem with n_c linear constraints:

$$f^* = \underset{\{\mathbf{x}_j \in \mathcal{X}_j\}_{j=0}^{N_a}}{\text{minimize}} \quad \sum_{j=0}^{N_a} f_j(\mathbf{x}_j) \tag{4.8a}$$

$$\text{subject to} \quad \sum_{j=0}^{N_a} \mathbf{A}_j \mathbf{x}_j = \mathbf{0}. \tag{4.8b}$$

For problem (4.1), constraint (4.8b) corresponds to the aggregator-level balance (4.1h). Set $\mathcal{X}_0$ captures constraints (4.1b)–(4.1g), while $\{\mathcal{X}_j\}_{j\in\mathcal{N}_a}$ corresponds to (4.1i).

The dual function $D(\boldsymbol{\mu}) = \sum_{j=0}^{N_a} D_j(\boldsymbol{\mu})$ can be obtained by dualizing constraint (4.8b) with the multiplier vector $\boldsymbol{\mu}$. Thus, the dual problem is to maximize the dual objective as

$$\max_{\boldsymbol{\mu}\in\mathbb{R}^{n_c}} \sum_{j=0}^{N_a} D_j(\boldsymbol{\mu}) = \max_{\boldsymbol{\mu}\in\mathbb{R}^{n_c}} \sum_{j=0}^{N_a} \left[\min_{\mathbf{x}_j} \{ f_j(\mathbf{x}_j) + \boldsymbol{\mu}' \mathbf{A}_j \mathbf{x}_j \} \right] \tag{4.9}$$

where strong duality holds here due to the polyhedral feasible set (4.8b).

The basic idea of bundle methods as well as the cutting plane method (CPM) is to approximate the epigraph of a convex (possibly non-smooth) objective function as the intersection of a number of supporting hyperplanes (also called cuts in this context). The approximation is gradually refined by generating additional cuts based on subgradients of the objective function.

Specifically, suppose that the method has so far generated the iterates $\{\boldsymbol{\mu}(\ell)\}_{\ell=1}^{k}$ after k steps. Let $\mathbf{x}_j(\ell)$ be the primal minimizer corresponding to $\boldsymbol{\mu}(\ell)$. Observe that the vector $\mathbf{g}_j(\ell) := \mathbf{A}_j\mathbf{x}_j(\ell)$ is a subgradient of function $D_j(\boldsymbol{\mu})$ at point $\boldsymbol{\mu}(\ell)$, and it thus holds for all $\boldsymbol{\mu}$ such that

$$D_j(\boldsymbol{\mu}) \leq D_j(\boldsymbol{\mu}(\ell)) + (\boldsymbol{\mu} - \boldsymbol{\mu}(\ell))'\mathbf{g}_j(\ell). \tag{4.10}$$

Clearly, the minimum of the right-hand side of (4.10) over $\ell = 1, \ldots, k$ is a polyhedral approximation of $D_j(\boldsymbol{\mu})$, and is essentially a concave and piecewise linear overestimator of the dual function.

The bundle method with disaggregated cuts generates a sequence $\{\boldsymbol{\mu}(k)\}$ with guaranteed convergence to an optimal solution. Specifically, the iterate $\boldsymbol{\mu}(k+1)$ is obtained by maximizing the polyhedral approximations of $\{D_j(\boldsymbol{\mu})\}_j$ with a proximal regularization:

$$D_{\text{ap}}(\boldsymbol{\mu}(k+1)) := \max_{\boldsymbol{\mu}, \{v_j\}_{j=1}^{N_a}} \quad \sum_{j=0}^{N_a} v_j - \frac{\rho(k)}{2}\|\boldsymbol{\mu} - \check{\boldsymbol{\mu}}(k)\|_2^2 \tag{4.11a}$$

$$\text{s.t.} \quad v_j \leq D_j(\boldsymbol{\mu}(\ell)) + (\boldsymbol{\mu} - \boldsymbol{\mu}(\ell))'\mathbf{g}_j(\ell),$$
$$j = 0, \ldots, N_a,\ \ell = 1, \ldots, k \tag{4.11b}$$

where the proximity weight $\rho(k) > 0$ is to control stability of the iterates; and the proximal center $\check{\boldsymbol{\mu}}(k)$ is updated according to a query for ascent:

$$\check{\boldsymbol{\mu}}(k+1) = \begin{cases} \boldsymbol{\mu}(k+1), & \text{if } D(\boldsymbol{\mu}(k+1)) - D(\check{\boldsymbol{\mu}}(k)) \geq \beta\eta(k) \\ \check{\boldsymbol{\mu}}(k), & \text{if } D(\boldsymbol{\mu}(k+1)) - D(\check{\boldsymbol{\mu}}(k)) < \beta\eta(k) \end{cases}$$

where $\eta(k) = D_{\text{ap}}(\boldsymbol{\mu}(k+1)) - D(\check{\boldsymbol{\mu}}(k))$, and $\beta \in (0, 1)$. Finally, the bundle algorithm can be terminated when $\eta(k) < \epsilon$ holds for a prescribed tolerance ϵ [cf. [13, Ch. 6]].

Remark 4.1 *(Bundle methods versus CPM).* When $\rho(k) \equiv 0$, problem (4.11) boils down to the CPM with disaggregated cuts for solving the dual, which is however known to be unstable and converges slowly on some practical instances [68]. The proximal regularization in the bundle methods is thus introduced to improve stability of the iterates, while the *smart* prox-center updating rule enhances further the convergence speed compared with the proximal CPM. A further limitation of CPM is that a compact set containing the optimal solution has to be included, as is the case with $\boldsymbol{\mu} \in [\boldsymbol{\mu}^{\min}, \boldsymbol{\mu}^{\max}]$ in [54].

The CPM convergence performance depends on the choice of this set, while there is no such issue for the bundle methods. Note further that the dual problem of (4.11) is a QP over a probability simplex. Such a special structure can be exploited by off-the-shelf QP solvers, and hence it is efficiently solvable. As a result, solving (4.11) does not require much more computational work than solving a linear program (LP), which is the case for the CPM. Finally, it is worth stressing that the disaggregated bundle method takes advantage of the *separability* of (4.8). In a nutshell, offering state-of-the-art algorithms for solving non-smooth convex programs, the stable and fast convergent bundle methods are well motivated here for clearing the market distributedly.

Specifically, applying the disaggregated bundle method to problem (4.9) at hand, the multiplier update at iteration k amounts to solving the following problem:

$$\underset{\{\mu_j^t, v_j\}_{j,t}}{\text{maximize}} \quad \sum_{j=0}^{N_a} v_j - \frac{\rho(k)}{2} \sum_{j=1}^{N_a} \sum_{t=1}^{T} (\mu_j^t - \check{\mu}_j^t)^2 \tag{4.12a}$$

$$\text{subject to} \quad v_0 \leq D_0(\boldsymbol{\mu}(\ell)) - \sum_{j=1}^{N_a} \sum_{t=1}^{T} P^t_{\text{DRA}_j}(\ell)[\mu_j^t - \mu_j^t(\ell)],\ \ell = 1, \ldots, k \tag{4.12b}$$

$$v_j \leq D_j(\boldsymbol{\mu}(\ell)) + \sum_{t=1}^{T} \sum_{r,s} p^t_{jrs}(\ell)[\mu_j^t - \mu_j^t(\ell)],\ j \in \mathcal{N}_a,\ \ell = 1, \ldots, k. \tag{4.12c}$$

Problem (4.12) that yields the updated multipliers $\boldsymbol{\mu}(k+1)$ can be solved at the ISO. The quantities $\{D_j(\boldsymbol{\mu}(k)), \sum_{r,s} p^t_{jrs}(k)\}_j$ are needed from each aggregator per iteration k as the problem input. Note that $D_j(\boldsymbol{\mu}(k)) := \sum_{r \in \mathcal{R}_j} D_{jr}(\boldsymbol{\mu}(k))$, where $D_{jr}(\boldsymbol{\mu}(k))$ is the optimal value of problem (4.7b). Thus, it is clear that all these required quantities can be formed at the aggregator level as summations over all end-users, and then transmitted to the ISO. The highlight here is that the proposed decomposition scheme respects *user privacy*, since $U_{jrs}(\mathbf{p}_{jrs})$ and $\mathcal{P}_{jrs}$ are never revealed.

4.3 Stochastic Market Clearing

The MC process is implemented with a goal of minimizing the system net cost, or equivalently maximizing the social welfare. With the trend of increasing penetration of

renewables, WPPs are able to directly bid in the forward market [121]. Under uncertainty of wind generation, it now becomes challenging but imperative for the ISOs/RTOs and market participants to extract forecast information and make efficient decisions, including reserve requirements, day-ahead scheduling, market clearing, reliability commitments, as well as the real-time dispatch [22]. In this section, a stochastic MC approach using the CVaR-based transaction cost will be developed as follows.

4.3.1 CVaR Revisited: A Convex Risk Measure

Value-at-risk (VaR) and conditional value-at-risk (CVaR) are widely used in various real-world applications, especially in the finance area, as the popular tools to evaluate the credit risk of a portfolio, and reduce the probability of large losses [48, 119, 120]. The following revisit is useful to grasp their role in the present context.

Consider a loss function $L(\mathbf{x}, \boldsymbol{\xi}) : X \times \Xi \mapsto \mathbb{R}$ denoting the real-valued cost associated with the decision variable $\mathbf{x} \in X \subset \mathbb{R}^n$; and the random vector $\boldsymbol{\xi}$ with probability density function $p(\boldsymbol{\xi})$ supported on a set $\Xi \subset \mathbb{R}^d$. In the context of power grids, $\mathbf{x}$ can represent the power schedules of generators, while $\boldsymbol{\xi}$ collects the sources of uncertainty due to for instance renewable energy and forecasted load demand.

Clearly, the probability of $L(\mathbf{x}, \boldsymbol{\xi})$ not exceeding a threshold η is given by the right-continuous cumulative distribution function (CDF)

$$\Psi(\mathbf{x}, \eta) = \int_{L(\mathbf{x}, \boldsymbol{\xi}) \leq \eta} p(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi}. \tag{4.13}$$

Definition 4.1 *(VaR). Given a prescribed confidence level $\beta \in (0, 1)$, the β-VaR is the generalized inverse of Ψ defined as*

$$\eta_\beta(\mathbf{x}) := \min\{\eta \in \mathbb{R} \mid \Psi(\mathbf{x}, \eta) \geq \beta\}. \tag{4.14}$$

β-VaR is essentially the β-quantile of the random $L(\mathbf{x}, \boldsymbol{\xi})$. Since Ψ is non-decreasing in η, $\eta_\beta(\mathbf{x})$ comes out as the lower endpoint of the solution interval satisfying $\Psi(\mathbf{x}, \eta) = \beta$, and the commonly chosen values of β are, e.g., 0.99, 0.95, and 0.9. Clearly, VaR determines a *maximum tolerable* loss of an investment, i.e., a threshold the loss will not exceed with a high probability β. Hence, given the confidence level β, investors are motivated to solve the so-termed *portfolio optimization* problem which yields the

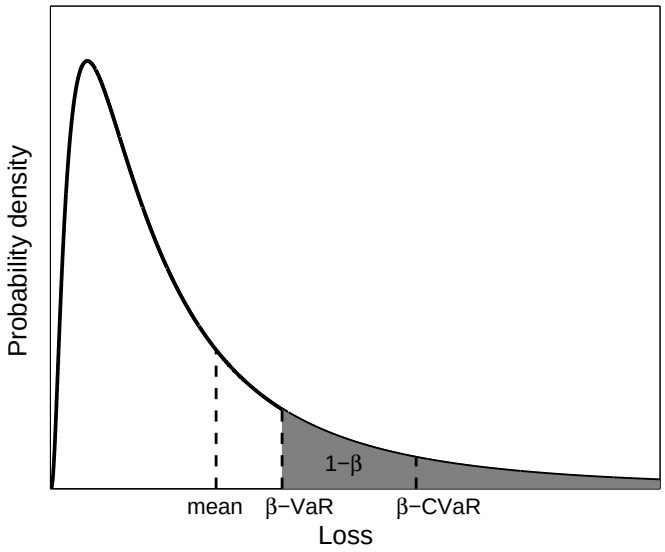


Figure 4.2: Illustration of VaR and CVaR: β-VaR is the threshold that the loss exceeds with at most a small probability $1-\beta$. β-CVaR is the conditional expectation of the loss beyond the β-VaR.

optimal investment decisions minimizing the VaR value. $\eta_\beta(\mathbf{x})$ is proportional to the standard deviation if Ψ is Gaussian. However, for general distributions, β-VaR is non-subadditive which means the VaR of a combined portfolio can be larger than the sum of the VaRs of each component. This violates the common principle "diversification reduces risk". Moreover, it is generally non-convex rendering the optimization task hard to tackle.

Because of these conceptual and practical drawbacks, CVaR (a.k.a. "tail VaR", "mean shortfall", or "mean excess loss") was proposed as an alternative risk metric that has many superior properties over VaR.

Definition 4.2 *(CVaR). The β-CVaR is the mean of the β-tail distribution of $L(\mathbf{x}, \boldsymbol{\xi})$, which is given as*

$$\Psi_\beta(\mathbf{x}, \eta) := \begin{cases} 0, & \text{if } \eta < \eta_\beta(\mathbf{x}) \\ \frac{\Psi(\mathbf{x},\eta)-\beta}{1-\beta}, & \text{if } \eta \geq \eta_\beta(\mathbf{x}) \end{cases} . \tag{4.15}$$

Truncated and re-scaled from Ψ, function Ψ_β is non-decreasing, right-continuous, and in fact a distribution function. If Ψ is continuous everywhere (without jumps),

β-CVaR coincides with the lower CVaR $\phi_\beta^-(\mathbf{x}) := \mathbb{E}_{\boldsymbol{\xi}}\left[L \mid L \geq \eta_\beta(\mathbf{x})\right]$, that is the conditional expectation of the loss beyond the β-VaR. Hence, roughly speaking, β-CVaR is the expected loss in the worst $100(1-\beta)\%$ scenarios; i.e., cases of such severe losses occur only $100(1-\beta)$ percent of the time.

The β-CVaR can be also defined as the optimal value of the following optimization problem

$$\phi_\beta(\mathbf{x}) := \min_{\eta \in \mathbb{R}} \left\{ \eta + \frac{1}{1-\beta} \mathbb{E}_{\boldsymbol{\xi}} \left[L(\mathbf{x}, \boldsymbol{\xi}) - \eta \right]^+ \right\}. \tag{4.16}$$

Let $F_\beta(\mathbf{x}, \eta)$ denote the objective function in (4.16). Key properties of F_β and its relationship with $\eta_\beta(\mathbf{x})$ and $\phi_\beta(\mathbf{x})$ are summarized next.

Theorem 4.1 ([120], pp. 1454–1457) *Function $F_\beta(\mathbf{x}, \eta)$ is finite and convex in η. Values $\eta_\beta(\mathbf{x})$ and $\phi_\beta(\mathbf{x})$ are linked through $F_\beta(\mathbf{x}, \eta)$ as*

$$\eta_\beta(\mathbf{x}) = \lfloor \arg\min_{\eta \in \mathbb{R}} F_\beta(\mathbf{x}, \eta) \rfloor \tag{4.17}$$

$$\phi_\beta(\mathbf{x}) = F_\beta(\mathbf{x}, \eta_\beta(\mathbf{x})) \tag{4.18}$$

$$\min_{\mathbf{x} \in X} \phi_\beta(\mathbf{x}) = \min_{(\mathbf{x}, \eta) \in X \times \mathbb{R}} F_\beta(\mathbf{x}, \eta). \tag{4.19}$$

Moreover, if $L(\mathbf{x}, \boldsymbol{\xi})$ is convex in $\mathbf{x}$, then $F_\beta(\mathbf{x}, \eta)$ is jointly convex in $(\mathbf{x}, \eta)$, while $\phi_\beta(\mathbf{x})$ is convex in $\mathbf{x}$.

From Definition 4.2, it can be seen that CVaR is an upper bound of VaR, implying that portfolios with small CVaR also have small VaR. As a consequence of Theorem 4.1, minimizing the convex $\phi_\beta(\mathbf{x})$ amounts to minimizing $F_\beta(\mathbf{x}, \eta)$, which is not only convex, but also easier to approximate. A readily implementable approximation of the expectation function F_β is its empirical estimate using N_s Monte Carlo samples $\{\boldsymbol{\xi}_s\}_{s=1}^{N_s}$, namely

$$\hat{F}_\beta(\mathbf{x}, \eta) = \eta + \frac{1}{N_s(1-\beta)} \sum_{s=1}^{N_s} \left[L(\mathbf{x}, \boldsymbol{\xi}_s) - \eta \right]^+. \tag{4.20}$$

Clearly, the sample average approximation method is distribution free, and the law of large numbers ensures $\hat{F}_\beta$ approximates well F_β for N_s large enough. Furthermore, $\hat{F}_\beta(\mathbf{x}, \eta)$ is convex with respect to $(\mathbf{x}, \eta)$ if $L(\mathbf{x}, \boldsymbol{\xi}_s)$ is convex in $\mathbf{x}$. The nondifferentiability due to the projection operator can be readily overcome by leveraging the epigraph form of $\hat{F}$, which will be shown explicitly in Section 4.3.4.

With the function $F_\beta(\mathbf{x}, \eta)$, it is now possible to develop the CVaR-based stochastic market clearing, as detailed in the next section.

4.3.2 CVaR-based Energy Transaction Cost

Consider a power system comprising N_b buses, N_l lines, N_g conventional generators, N_w wind farms and N_a aggregators, each serving a large number of residential end-users with controllable smart appliances. Let $\mathcal{T} := \{1, 2, \ldots, T\}$ denote the scheduling horizon of interest, e.g., one day ahead. If a wind farm is located at bus m, two quantities will be associated with it: the actual wind power generation w_m, and the power scheduled to be injected p_{W_m}. Note that the former is random, whereas the latter is a decision variable. For notational simplicity, define also two N_w-dimensional vectors $\mathbf{w}^t := [w_1^t, \ldots, w_{N_w}^t]'$, and $\mathbf{p}_W^t := [p_{W_1}^t, \ldots, p_{W_{N_w}}^t]'$.

Since $\mathbf{w}^t$ varies randomly, either energy surplus or shortage should be included to satisfy the nodal balance with the committed quantity $\mathbf{p}_W^t$. When surplus occurs, the wind farms can sell the excess wind energy back to the spot market, or simply curtail it. For the case of shortage, in order to accomplish the promised bid in the DA contract, farms can buy the energy shortfall from the RT market in the form of ancillary services.

Let $\mathbf{b}^t := [b_1^t, \ldots, b_{N_w}^t]'$ and $\mathbf{s}^t := [s_1^t, \ldots, s_{N_w}^t]'$ collect the purchase and selling prices at time t, respectively. Clearly, with the power shortfall and surplus being $[\mathbf{p}_W^t - \mathbf{w}^t]^+$ and $[\mathbf{w}^t - \mathbf{p}_W^t]^+$ at time t, the grid-wide net transaction cost is

$$\begin{aligned} T(\mathbf{p}_W, \mathbf{w}) &= \sum_{t=1}^{T} \Big(\mathbf{b}^t \cdot [\mathbf{p}_W^t - \mathbf{w}^t]^+ - \mathbf{s}^t \cdot [\mathbf{w}^t - \mathbf{p}_W^t]^+ \Big) \\ &= \sum_{t=1}^{T} \Big(\boldsymbol{\varpi}^t \cdot |\mathbf{p}_W^t - \mathbf{w}^t| + \boldsymbol{\vartheta}^t \cdot (\mathbf{p}_W^t - \mathbf{w}^t) \Big) \end{aligned} \tag{4.21}$$

where $\boldsymbol{\varpi}^t := \frac{\mathbf{b}^t - \mathbf{s}^t}{2}$ and $\boldsymbol{\vartheta}^t := \frac{\mathbf{b}^t + \mathbf{s}^t}{2}$; $\mathbf{p}_W$ and $\mathbf{w}$ collect $\mathbf{p}_W^t$ and $\mathbf{w}^t$ for all $t \in \mathcal{T}$, respectively.

Replacing $L(\cdot, \cdot)$ in (4.16) with $T(\cdot, \cdot)$, function F_β can be expressed through the conditional expected transaction cost as

$$F_\beta(\mathbf{p}_W, \eta) = \eta + \frac{1}{1-\beta} \mathbb{E}_{\mathbf{w}} \left[\sum_{t=1}^{T} \Big(\boldsymbol{\varpi}^t \cdot |\mathbf{p}_W^t - \mathbf{w}^t| + \boldsymbol{\vartheta}^t \cdot (\mathbf{p}_W^t - \mathbf{w}^t) \Big) - \eta \right]^+ . \tag{4.22}$$

A condition guaranteeing convexity of $F_\beta(\mathbf{p}_W, \eta)$ is established next.

Proposition 4.1 *If the selling price s_m^t does not exceed the purchase price b_m^t for any $m \in \mathcal{N}_w$ and $t \in \mathcal{T}$, function $F_\beta(\mathbf{p}_W, \eta)$ is jointly convex with respect to $(\mathbf{p}_W, \eta)$.*

Proof: Thanks to Theorem 4.1, it suffices to show that $T(\mathbf{p}_W, \mathbf{w}) = \sum_{t=1}^{T} \Big(\boldsymbol{\varpi}^t \cdot |\mathbf{p}_W^t - \mathbf{w}^t| + \boldsymbol{\vartheta}^t \cdot (\mathbf{p}_W^t - \mathbf{w}^t) \Big)$ is convex in $\mathbf{p}_W$ under the proposition's condition. Clearly, the stated condition is equivalent to $\boldsymbol{\varpi}^t \succeq \mathbf{0}$ for all $t \in \mathcal{T}$. Thus, by the convexity of the absolute value function, and the convexity-preserving operators of summation and expectation [26, Sec. 3.2], the claim follows readily. □

In this thesis, a perfectly competitive market is assumed such that all participants act as price takers. That is, every competitor is *atomistic* to have small enough market share so that there is no market power affecting the price [130]. For American electricity markets, a single pricing mechanism is used such that $\mathbf{s}^t \equiv \mathbf{b}^t$ holds in most of the scenarios. This is a special case of the pricing condition in Proposition 4.1, which facilitates calculating the function (4.22) since the absolute value functions vanish. Note that it is possible that different WPPs may buy (sell) wind energy from (to) different sellers (purchasers) in a competitive electricity pool as an ancillary service, which can yield different purchase and selling prices.

For most of the European markets including UK, France, Italy, and Netherlands, the imbalance prices $\{\mathbf{b}^t, \mathbf{s}^t\}_t$ are commonly set in an *ex-post* way that is known as *dual imbalance pricing* [114]. Specifically, if the system RT imbalance is negative, i.e., the overall market is short, then $\mathbf{s}^t = \boldsymbol{\chi}^t \preceq \mathbf{b}^t$ holds, where $\boldsymbol{\chi}^t := [\chi_1^t, \ldots, \chi_{N_w}^t]'$ collects the DA prices at the buses attached with all N_w wind farms. In this case, the RT purchase price is typically higher than the DA price, reflecting the cost of acquiring the balancing energy [107]. Wind farms with excess energy can sell this part to reduce the system imbalance but only be paid the DA prices. On the other hand, we have $\mathbf{s}^t \preceq \boldsymbol{\chi}^t = \mathbf{b}^t$ if the market is long. Hence, market participants selling excess energy receive a balancing price which is lower than the DA one, while those running negative imbalance pay the DA price. Note that the relationship $\mathbf{s}^t \preceq \boldsymbol{\chi}^t \preceq \mathbf{b}^t$ always holds even when the market imbalance outcome is unknown at the time of the DA bids. Such a pricing mechanism drives bidders to match their forward offers with the true forecasts of generation or consumption.

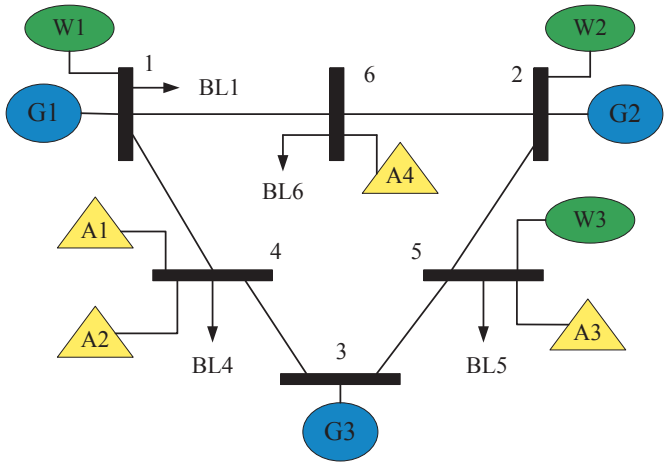


Figure 4.3: Modified WECC system featuring 6 buses, 3 generators, 4 aggregators, 4 base loads, and 3 wind farms.

Leveraging the CVaR-based transaction cost, a stochastic MC problem based on the DC-OPF will be formulated next.

4.3.3 CVaR-based Market Clearing

Let $\mathbf{p}_G^t := [P_{G_1}^t, \ldots, P_{G_{N_g}}^t]'$ and $\mathbf{p}_{\text{DRA}}^t := [P_{\text{DRA}_1}^t, \ldots, P_{\text{DRA}_{N_a}}^t]'$ denote the power outputs of the thermal generators, and the power consumption of the aggregators at slot t, respectively. Define further the sets $\mathcal{N}_a := \{1, 2, \ldots, N_a\}$ and $\mathcal{N}_g := \{1, 2, \ldots, N_g\}$. Each aggregator $j \in \mathcal{N}_a$ serves a set $\mathcal{R}_j$ of residential users, and each user $r \in \mathcal{R}_j$ has a set $\mathcal{S}_{rj}$ of controllable appliances. Let $\mathbf{p}_{jrs}$ be the power consumption of appliance s with user r corresponding to aggregator j across the slots. The operational constraints of $\mathbf{p}_{jrs}$ are captured by a set $\mathcal{P}_{jrs}$, while the end user satisfaction is modeled by a concave utility function $U_{jrs}(\mathbf{p}_{jrs})$. Furthermore, let convex functions $\{C_i(\cdot)\}_i$ denote the generation costs, and $\mathbf{p}_{\text{BL}}^t$ the base load demand. For brevity, let vector $\mathbf{p}_0$ collect variables η and $\{\mathbf{p}_G^t, \mathbf{p}_{\text{DRA}}^t, \mathbf{p}_W^t, \boldsymbol{\theta}^t\}_{t\in\mathcal{T}}$; and vector $\{\mathbf{p}_j\}_{j\in\mathcal{N}_a}$ the power consumption of all appliances with the aggregator j. Based on the WECC system (see Fig. 4.1), a modified one with wind power integration is shown in Fig. 4.3.

Hinging on three assumptions: a1) lossless lines, a2) small voltage phase differences, and a3) approximated one p.u. voltage magnitudes, the DC-OPF based stochastic MC

stands with the goal of minimizing the social cost:

$$\text{minimize} \quad \sum_{t=1}^{T}\sum_{i=1}^{N_g} C_i(P_{G_i}^t) - \sum_{j=1}^{N_a}\sum_{\substack{r\in\mathcal{R}_j,\\ s\in\mathcal{S}_{jr}}} U_{jrs}(\mathbf{p}_{jrs}) + \mu F_\beta(\mathbf{p}_W, \eta) \tag{4.23a}$$

$$\text{subject to} \quad \mathbf{A}_g\mathbf{p}_G^t + \mathbf{A}_w\mathbf{p}_W^t - \mathbf{A}_a\mathbf{p}_{\text{DRA}}^t - \mathbf{p}_{\text{BL}}^t = \mathbf{B}_n\boldsymbol{\theta}^t,\ t\in\mathcal{T} \tag{4.23b}$$

$$P_{G_i}^{\min} \le P_{G_i}^t \le P_{G_i}^{\max},\ i\in\mathcal{N}_g,\ t\in\mathcal{T} \tag{4.23c}$$

$$-\mathsf{R}_i^{\text{down}} \le P_{G_i}^t - P_{G_i}^{t-1} \le \mathsf{R}_i^{\text{up}},\ i\in\mathcal{N}_g,\ t\in\mathcal{T} \tag{4.23d}$$

$$\mathbf{f}^{\min} \preceq \mathbf{B}_f\boldsymbol{\theta}^t \preceq \mathbf{f}^{\max},\ t\in\mathcal{T} \tag{4.23e}$$

$$\theta_1^t = 0,\ t\in\mathcal{T} \tag{4.23f}$$

$$\mathbf{0} \preceq \mathbf{p}_W \preceq \mathbf{p}_W^{\max} \tag{4.23g}$$

$$0 \le P_{\text{DRA}_j}^t \le P_{\text{DRA}_j}^{\max},\ j\in\mathcal{N}_a,\ t\in\mathcal{T} \tag{4.23h}$$

$$P_{\text{DRA}_j}^t = \sum_{r\in\mathcal{R}_j,\, s\in\mathcal{S}_{jr}} p_{jrs}^t,\ j\in\mathcal{N}_a,\ t\in\mathcal{T} \tag{4.23i}$$

$$\mathbf{p}_{jrs}\in\mathcal{P}_{jrs},\ s\in\mathcal{S}_{jr},\ r\in\mathcal{R}_j,\ j\in\mathcal{N}_a \tag{4.23j}$$

$$\text{variables} \quad \{\mathbf{p}_j\}_{j=0}^{N_a}$$

Linear equality (4.23b) is the *nodal balance* constraint; i.e., the load balance at bus levels dictated by the law of conservation of power. Limits of generator outputs and ramping rates are specified in constraints (4.23c) and (4.23d). Network power flow constraints are accounted for in (4.23e). Without loss of generality, the first bus can be set as the reference bus with zero phase in (4.23f). Constraints (4.23h) and (4.23g) capture the lower and upper limits of the energy consumed by the aggregators and the committed wind power, respectively. Equality (4.23i) amounts to the *aggregator-user power balance* equation; and constraints (4.23j) define the feasible set of appliances. Finally, the pre-determined risk-aversion parameter $\mu > 0$ controls the trade off between the transaction cost and the generation cost as well as the end-user utility.

Remark 4.2 *(Availability of real-time prices).* In this thesis, the real-time prices $\{\mathbf{b}^t, \mathbf{s}^t\}_{t\in\mathcal{T}}$ are assumed to be perfectly known to the ISO for the DA market clearing. However, such an assumption can be readily extended to a more practical setup by taking the price stochasticity into account. Specifically, imperfect price information can be modeled by appropriately designing the function $T(\mathbf{p}_W, \mathbf{w})$ [cf. (4.21)]. For

example, the expectation can be also taken over the random RT prices in (4.22) as $F_\beta(\mathbf{p}_W, \eta) = \eta + \frac{1}{1-\beta}\mathbb{E}_{\{\mathbf{w},\{\mathbf{b}^t,\mathbf{s}^t\}_t\}}[T(\mathbf{p}_W, \mathbf{w}) - \eta]^+$. The dependence between $\{\mathbf{b}^t, \mathbf{s}^t\}$ and $\mathbf{w}$ can be further investigated. In addition, worst-case analysis is available upon postulating an uncertainty set Δ for $\{\mathbf{b}^t, \mathbf{s}^t\}$. This results in a novel risk measure given as $F_\beta(\mathbf{p}_W, \eta) = \eta + \frac{1}{1-\beta}\mathbb{E}_{\mathbf{w}}[\sup_{\{\mathbf{b}^t,\mathbf{s}^t\}_t \in \Delta} T(\mathbf{p}_W, \mathbf{w}) - \eta]^+$.

It is worth mentioning that SCED and SCUC yield two different market pricing systems: locational marginal pricing and convex hull pricing (a.k.a. extended LMP). The ED formulation produces the LMPs given by the dual variables associated with the supply-demand balance constraint. Prices supporting the equilibrium solution are found at the intersection of the supply marginal cost curve with the demand bids. However, if discrete operations of UC are involved, there is no exact price that supports such an economic equilibrium. This issue prompted the introduction of the convex hull pricing to reduce the uplift payments [60]. In the present thesis, the core ED model is considered to deal with the high penetration of renewables and large-scale DR programs. Therefore, the formulation (4.23) relies on re-solving the dispatch problem with fixed UC decisions.

Remark 4.3 *(Reliability assessment commitment).* The proposed dispatch model can be cast as a two-stage program. The first stage is the DA MC, and the second is simply the balancing operation (recourse action) dealing with differences between the pre-dispatch amount and the actual wind power generation. Between the DA and RT markets, ISOs implement the reliability assessment commitment (RAC) as a reliability backstop tool to ensure sufficient resources are available and cover the adjusted forecast load online. One principle of the RAC process is to commit the capacity deemed necessary to reliably operate the grid at the least commitment cost. In this step, based on the updated information of the wind power forecast, WPPs have an opportunity to feedback to the ISO if they are able to commit the scheduled wind power decided by the DA MC. Then, the ISO is able to adjust UC decisions as necessary to ensure reliability.

To this end, reformulation of problem (4.23) as a smooth convex minimization is useful for developing distributed solvers, as detailed next.

4.3.4 Smooth Convex Minimization Reformulation

It is clear that under the condition of Proposition 4.1, the objective and the constraints of (4.23) are convex, which renders it not hard to solve in principle. Nevertheless, due to the high-dimensional integration present in $F_\beta(\mathbf{p}_W, \eta)$ [cf. (4.22)], an analytical solution is typically impossible. To this end, it is necessary to re-write the resulting problem in a form suitable for off-the-shelf solvers.

First, as shown in (4.20), an efficient approximation of $F_\beta(\mathbf{p}_W, \eta)$ is offered by the empirical expectation using i.i.d. samples $\{\mathbf{w}_s\}_{s=1}^{N_s}$; that is,

$$\hat{F}_\beta(\mathbf{p}_W, \eta) = \eta + \frac{1}{N_s(1-\beta)} \sum_{s=1}^{N_s} \left[\sum_{t=1}^{T} \left(\boldsymbol{\varpi}^t \cdot |\mathbf{p}_W^t - \mathbf{w}_s^t| + \boldsymbol{\vartheta}^t \cdot (\mathbf{p}_W^t - \mathbf{w}_s^t) \right) - \eta \right]^+. \tag{4.24}$$

Next, by introducing auxiliary variables $\{u_s\}_{s=1}^{Ns}$, the non-smooth convex program (4.23) can be equivalently re-written as the following smooth convex minimization:

$$\text{minimize} \quad \sum_{t=1}^{T} \sum_{i=1}^{N_g} C_i(P_{G_i}^t) - \sum_{j=1}^{N_a} \sum_{\substack{r \in \mathcal{R}_j, \\ s \in \mathcal{S}_{jr}}} U_{jrs}(\mathbf{p}_{jrs}) + \mu \left(\eta + \frac{\sum_{s=1}^{Ns} u_s}{N_s(1-\beta)} \right) \tag{4.25a}$$

$$\text{subject to} \quad \text{(4.23b)} - \text{(4.23j)}$$

$$\sum_{t=1}^{T} \left(\boldsymbol{\varpi}^t \cdot |\mathbf{p}_W^t - \mathbf{w}_s^t| + \boldsymbol{\vartheta}^t \cdot (\mathbf{p}_W^t - \mathbf{w}_s^t) \right) \leq u_s + \eta, \ s \in \mathcal{N}_s \tag{4.25b}$$

$$\text{variables} \quad \{\mathbf{p}_j\}_{j=0}^{N_a}, \ \{u_s \in \mathbb{R}_+\}_{s=1}^{N_s}.$$

Under mild conditions, the optimal solution set of (4.25) converges exponentially fast to its counterpart of (4.23), as the sample size N_s increases. The proof is based on the theory of large deviations [85], but is omitted here due to space limitations.

Problem (4.25) can be solved centrally at the ISO in principle. However, with large-scale DR, distributed solvers are well motivated not only for computational efficiency but also for privacy reasons. Specifically, functions $U_{jrs}(\mathbf{p}_{jrs})$ and sets $\{\mathcal{P}_{jrs}\}$ are private, and are not revealed to the ISO; and (ii) the operational sets $\{\mathcal{P}_{jrs}\}_{j,r,s}$ of very large numbers of heterogenous appliances may become prohibitively complicated; e.g., mix-integer constraints can even be involved to model the ON/OFF status and uninterruptible operating time of end-user appliances [32, 80]. This renders the overall

problem intractable for the ISO. To this end, the DR aggregators can play a critical role to split the resulting optimization task as detailed next.

4.4 Distributed Market Clearing via ADMM

Selecting how to decompose the optimization task as well as updating the associated multipliers are crucial for the distributed design. Fewer updates simply imply lower communication overhead between the ISO and the aggregators. One splitting approach is the dual decomposition with which the dual subgradient ascent algorithm is typically very slow. Instead, a fast-convergent solver via the ADMM [25] is adapted in this section for the distributed MC.

4.4.1 The ADMM Algorithm

Consider the following separable convex minimization problem with linear equality constraints:

$$\underset{\mathbf{x}\in\mathcal{X},\mathbf{y}\in\mathcal{Y}}{\text{minimize}} \quad f(\mathbf{x})+g(\mathbf{y}) \tag{4.26a}$$

$$\text{subject to} \quad \mathbf{A}\mathbf{x}+\mathbf{B}\mathbf{y}=\mathbf{c}. \tag{4.26b}$$

For the stochastic MC problem (4.25), the primal variable $\mathbf{x}$ comprises the group $\{u_s\}_{s\in\mathcal{N}_s}$ and $\mathbf{p}_0$, while $\mathbf{y}$ collects $\{\mathbf{p}_j\}_{j\in\mathcal{N}_a}$. Hence, set $\mathcal{X}$ captures constraints (4.23b)–(4.23h) and (4.25b) while $\mathcal{Y}$ represents (4.23j). The linear equality constraint (4.26b) corresponds to (4.23i).

Let $\boldsymbol{\lambda}:=[\lambda_1^1,\ldots,\lambda_{N_a}^T]'\in\mathbb{R}^{TN_a}$ denote the Lagrange multiplier vector associated with the constraint (4.23i). The partially augmented Lagrangian of (4.25) is thus given by

$$\begin{aligned}L_\rho(\mathbf{x},\mathbf{y},\boldsymbol{\lambda})=&\sum_{t=1}^{T}\sum_{i=1}^{N_g}C_i(P_{G_i}^t)-\sum_{j=1}^{N_a}\sum_{\substack{r\in\mathcal{R}_j,\\ s\in\mathcal{S}_{jr}}}U_{jrs}(\mathbf{p}_{jrs})+\mu\left(\eta+\frac{\sum_{s=1}^{Ns}u_s}{N_s(1-\beta)}\right)\\&+\sum_{t=1}^{T}\sum_{j=1}^{N_a}\lambda_j^t\left(P_{\text{DRA}_j}^t-\sum_{r,s}p_{jrs}^t\right)+\frac{\rho}{2}\sum_{t=1}^{T}\sum_{j=1}^{N_a}\left(P_{\text{DRA}_j}^t-\sum_{r,s}p_{jrs}^t\right)^2\end{aligned} \tag{4.30}$$

Algorithm 4.1 ADMM-based Distributed Market Clearing

1: Initialize $\boldsymbol{\lambda}(0) = \mathbf{0}$
2: **repeat** for $k = 0, 1, 2, \ldots$
3: **update primal variables:**

$$\mathbf{x}(k+1) = \arg\min_{\mathbf{x}\in\mathcal{X}} \; L_\rho(\mathbf{x}, \mathbf{y}(k), \boldsymbol{\lambda}(k)) \tag{4.27}$$

$$\mathbf{y}(k+1) = \arg\min_{\mathbf{y}\in\mathcal{Y}} \; L_\rho(\mathbf{x}(k+1), \mathbf{y}, \boldsymbol{\lambda}(k)) \tag{4.28}$$

4: **update dual variables:** for all $j \in \mathcal{N}_a$ and $t \in \mathcal{T}$

$$\lambda_j^t(k+1) = \lambda_j^t(k) + \rho\big(P_{\mathrm{DRA}_j}^t(k+1) - \sum_{r,s} p_{jrs}^t(k+1)\big) \tag{4.29}$$

5: **until** $\xi \le \epsilon^{\mathrm{pri}}$

where the weight $\rho > 0$ is a penalty parameter controlling the violation of primal feasibility, which turns out to be the step size of the dual update.

As the iterative solver of (4.30) proceeds, the primal residual converges to zero that ensures optimality. Large values of ρ give rise to small primal residuals at the expense of larger dual residuals. Judiciously selecting ρ thus strikes a desirable tradeoff between the size of primal vis-à-vis dual residuals. Note also that by varying ρ over a finite number of iterations may improve convergence [25]. In a nutshell, finding the "optimal" value of ρ is generally application-dependent that requires a trial-and-error tuning.

Different from [86] where the power balance and phase consistency constraints are relaxed, in this work only the aggregator-user power balance equation (4.23i) is dualized so that the nodal balance equation (4.23b) is kept in the subproblem of the ISO. Decomposing the problem (4.25) in such a way can reduce the heavy computational burden at the ISO while respect the privacy of end users within each aggregator. The ADMM iteration cycles between primal variable updates using block coordinate descent (a.k.a. Gauss-Seidel), and dual variable updates via gradient ascent. The resulting distributed MC is tabulated as Algorithm 4.1, where k is the iteration index. The last step is a

reasonable termination criterion based on the primal residual [25, Sec. 3.3.1]

$$\xi := \left[\sum_{t=1}^{T} \sum_{j=1}^{N_a} \left(P^t_{\mathrm{DRA}_j} - \sum_{r,s} p^t_{jrs} \right)^2 \right]^{1/2}. \tag{4.31}$$

Specifically, given the Lagrangian multipliers $\boldsymbol{\lambda}(k)$ and the power consumption $\{\mathbf{p}_{jrs}(k)\}_{jrs}$ of the end-user appliances, the ISO solves the convex subproblem (4.27) given as follows:

$$\begin{aligned}
\mathbf{p}_0(k+1) = \underset{\mathbf{p}_0, \{u_s\}}{\arg\min} \; & \sum_{\substack{t\in\mathcal{T},\\ i\in\mathcal{N}_g}} C_i(P^t_{G_i}) + \mu \left(\eta + \frac{\sum_{s=1}^{Ns} u_s}{N_s(1-\beta)} \right) + \sum_{\substack{t\in\mathcal{T},\\ j\in\mathcal{N}_a}} \lambda^t_j(k) P^t_{\mathrm{DRA}_j} \\
& + \frac{\rho}{2} \sum_{\substack{t\in\mathcal{T},\\ j\in\mathcal{N}_a}} \left(P^t_{\mathrm{DRA}_j} - \sum_{r,s} p^t_{jrs}(k) \right)^2 && \text{(4.32a)} \\
\text{subject to} \quad & \mathbf{A}_g \mathbf{p}^t_G + \mathbf{A}_w \mathbf{p}^t_W - \mathbf{A}_a \mathbf{p}^t_{\mathrm{DRA}} - \mathbf{p}^t_{\mathrm{BL}} = \mathbf{B}_n \boldsymbol{\theta}^t, \; t \in \mathcal{T} && \text{(4.32b)} \\
& P^{\min}_{G_i} \le P^t_{G_i} \le P^{\max}_{G_i}, \; i \in \mathcal{N}_g, \; t \in \mathcal{T} && \text{(4.32c)} \\
& -\mathsf{R}^{\mathrm{down}}_i \le P^t_{G_i} - P^{t-1}_{G_i} \le \mathsf{R}^{\mathrm{up}}_i, \; i \in \mathcal{N}_g, \; t \in \mathcal{T} && \text{(4.32d)} \\
& \mathbf{f}^{\min} \preceq \mathbf{B}_f \boldsymbol{\theta}^t \preceq \mathbf{f}^{\max}, \; t \in \mathcal{T} && \text{(4.32e)} \\
& \theta^t_1 = 0, \; t \in \mathcal{T} && \text{(4.32f)} \\
& \mathbf{0} \preceq \mathbf{p}_W \preceq \mathbf{p}^{\max}_W && \text{(4.32g)} \\
& 0 \le P^t_{\mathrm{DRA}_j} \le P^{\max}_{\mathrm{DRA}_j}, \; j \in \mathcal{N}_a, \; t \in \mathcal{T} && \text{(4.32h)} \\
& \sum_{t=1}^{T} \left(\boldsymbol{\varpi}^t \cdot |\mathbf{p}^t_W - \mathbf{w}^t_s| + \boldsymbol{\vartheta}^t \cdot (\mathbf{p}^t_W - \mathbf{w}^t_s) \right) \le u_s + \eta, \; s \in \mathcal{N}_s && \text{(4.32i)} \\
& u_s \ge 0, \; s \in \mathcal{N}_s. && \text{(4.32j)}
\end{aligned}$$

Interestingly, (4.28) is decomposable so that $\{\mathbf{p}_{jrs}(k)\}_{r,s}$ can be separately solved by each aggregator:

$$\begin{aligned}
\{\mathbf{p}_{jrs}(k+1)\}_{r,s} = \underset{\{\mathbf{p}_{jrs}\}_{r,s}}{\arg\min} \; & -\sum_{t=1}^{T} \lambda^t_j(k) \sum_{r,s} p^t_{jrs} - \sum_{\substack{r\in\mathcal{R}_j,\\ s\in\mathcal{S}_{jr}}} U_{jrs}(\mathbf{p}_{jrs}) \\
& + \frac{\rho}{2} \sum_{t=1}^{T} \left(\sum_{r,s} p^t_{jrs} - P^t_{\mathrm{DRA}_j}(k+1) \right)^2 && \text{(4.33a)}
\end{aligned}$$

$$\text{subject to} \quad \{\mathbf{p}_{jrs} \in \mathcal{P}_{jrs}\}_{r,s}. \tag{4.33b}$$

Having found $\mathbf{p}_0(k)$ and $\{\mathbf{p}_{jrs}(k)\}_{jrs}$, the multipliers $\{\mu_j^t\}_{j,t}$ are updated using gradient ascent as in (4.29). To solve the convex problem (4.33), each aggregator must collect the corresponding users' information including U_{jrs} and $\mathcal{P}_{jrs}$. This is implementable via the advanced metering infrastructure [4].

Remark 4.4 *(Distributed demand response).* It must be further pointed out that the quadratic penalty $\left(P^t_{\text{DRA}_j} - \sum_{r,s} p^t_{jrs}\right)^2$ in (4.30) couples load consumptions $\{p^t_{jrs}\}$ over different residential users. Hence, the ADMM-based distributed solver may not be applicable whenever p^t_{jrs} must be updated per end user rather than the aggregator. This may arise either to strictly protect the privacy of end users from DR aggregators, or, to accommodate large-scale DR programs where each aggregator cannot even afford solving the subproblem (4.33). In this case, leveraging the plain Lagrangian function (no coupling term), the dual decomposition based schemes can be utilized by end users to separately update $\{p^t_{jrs}\}$ in parallel; see e.g., [54] and [163].

The convergence of the ADMM solver and its implications for the market price are discussed next.

4.4.2 Pricing Impacts

Suppose two additional conditions hold for the convex problem (4.25): c1) functions $\{C_i(\cdot)\}_i$ and $\{-U_{jrs}(\cdot)\}_{jrs}$ are closed and proper convex; and c2) the plain Lagrangian L_0 has a saddle point. Then, the ADMM iterates of the objective (4.25a) and the dual variables $\{\lambda^t_j\}_{j,t}$ are guaranteed to converge to the optimum [25]. In addition, if the objective is strongly convex, then the primal variable iterates including $\mathbf{p}_G$, $\mathbf{p}_{\text{DRA}}$, $\mathbf{p}_W$ and $\{\mathbf{p}_j\}_{j\in\mathcal{N}_a}$ converge to the globally optimal solutions [47].

The guaranteed convergence of the dual variables also facilitates the calculation of LMPs. Let $\bar{\boldsymbol{\lambda}}^t := [\bar{\lambda}^t_1, \ldots, \bar{\lambda}^t_{N_a}]'$ and $\bar{\boldsymbol{\tau}}^t := [\bar{\tau}^t_1, \ldots, \bar{\tau}^t_{N_b}]'$ denote the optimal Lagrange multipliers associated with the aggregator-user balance constraint (4.23i), and the nodal balance constraint (4.23b), respectively. Note that with the optimal solutions $\bar{\boldsymbol{\lambda}}^t$ and $\{\bar{\mathbf{p}}_{jrs}\}_{jrs}$ obtained by the ADMM solver, the LMPs $\{\bar{\boldsymbol{\tau}}^t\}_t$ can be found by solving the subproblem (4.32) with primal-dual algorithms. In addition, if $0 < P^t_{\text{DRA}_j} < P^{\max}_{\text{DRA}_j}, \forall j, t$ holds at the optimal solution $\bar{P}^t_{\text{DRA}_j}$, then $\bar{\boldsymbol{\lambda}}^t = \mathbf{A}'_a \bar{\boldsymbol{\tau}}^t$; i.e., $\bar{\lambda}^t_j = \bar{\tau}^t_n$ for all aggregators j

attached with bus n (see also [54]). To this end, payments of the market participants can be calculated with the obtained LMPs and optimal DA dispatches. In the RT market of a two-settlement system, if the supplier at bus n delivers $\tilde{P}^t_{G_n}$ with the real-time price $\tilde{\tau}^t_n$, then the supplier gets paid

$$\Pi_{G_n} = \sum_{t=1}^{T} \bar{\tau}^t_n \bar{P}^t_{G_n} + \tilde{\tau}^t_n (\tilde{P}^t_{G_n} - \bar{P}^t_{G_n}).$$

Likewise, the aggregator at bus n needs to pay

$$\Pi_{\mathrm{DRA}_n} = \sum_{t=1}^{T} \bar{\tau}^t_n \bar{P}^t_{\mathrm{DRA}_n} + \tilde{\tau}^t_n (\tilde{P}^t_{\mathrm{DRA}_n} - \bar{P}^t_{\mathrm{DRA}_n}).$$

The revenue of the wind farm at bus n is

$$\Pi_{W_n} = \sum_{t=1}^{T} \big(\bar{\tau}^t_n \bar{p}^t_{W_n} + s^t_n [w^t_n - \bar{p}^t_{W_n}]^+ - b^t_n [\bar{p}^t_{W_n} - w^t_n]^+ \big).$$

Remark 4.5 *(Pricing consistence).* In a perfectly competitive market, any arbitrage opportunities between the DA and RT markets are exploited by market participants. Hence, the DA nodal prices are consistent with the DT nodal prices meaning the expectations of the latter converge to the former. The concepts of price distortions and revenue adequacy have been recently proposed for the stochastic MC in [154]. In the setup of a single snapshot therein, it has been proved that the medians and expectations of RT prices converge to the DA counterparts for the ℓ_1 and ℓ_2 penalties between the RT and DA power schedules, respectively. Building upon this solid result, it is possible to establish bounded price distortions for the proposed model, while its consistent pricing property can also be analyzed in a similar fashion.

4.5 Numerical Tests

In this section, numerical tests are implemented to verify the performance of the proposed distributed MC. The simulation results for the deterministic MC (Sec. 4.1–4.2) will be presented first, followed by the counterpart of the stochastic MC with wind power injection (Sec. 4.3–4.4).

Table 4.1: Generator parameters. The units of a_i and b_i are \$/(MWh)^2 and \$/MWh, respectively. the rest are in MW.

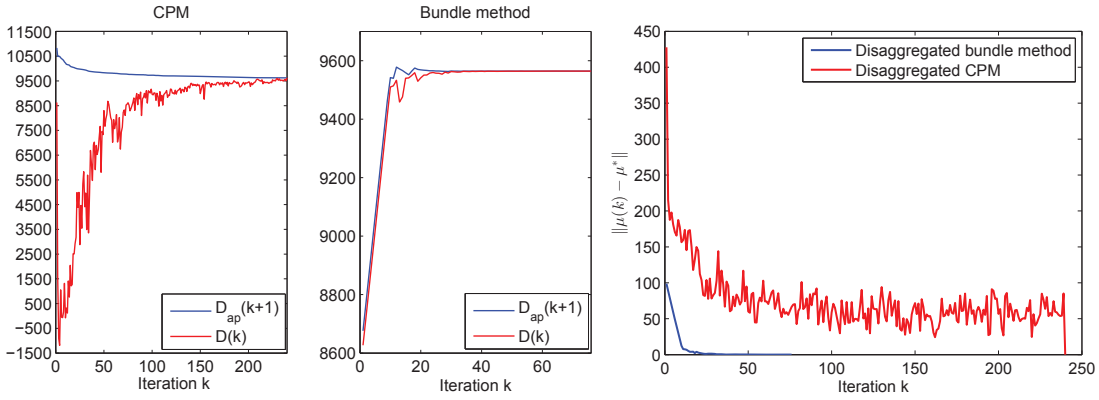
Gen.	a_i	b_i	$P_{G_i}^{\max}$	$P_{G_i}^{\min}$	$\mathsf{R}_i^{\text{up,down}}$
1	0.3	3	60	2.4	50
2	0.15	20	50	0	35
3	0.2	50	50	0	40

Table 4.2: Parameters of residential appliances. All listed hours are the ending ones; w.p. means with probability.

E_{PHEV} (kWh)	Uniform on {10, 11, 12}
$p_{\text{PHEV}}^{\max}$ (kWh)	Uniform on {2.1, 2.3, 2.5}
$p_{\text{PHEV}}^{\min}$ (kWh)	0
T_{jr1}^{st}	1am
T_{jr1}^{end}	6am w.p. 70%, 7am w.p. 30%

The power system tested for the disaggregated bundle method is illustrated in Fig. 4.1, where each of the 4 aggregators serves $1,000$ residential end-users. The scheduling horizon starts from 12am until 11pm, a total of 24 hours. Time-invariant generation cost functions are set to be quadratic as $C_i(P_{G_i}^t) = a_i(P_{G_i}^t)^2 + b_i P_{G_i}^t$ for all i and t. Each end-user has a PHEV to charge overnight. All detailed parameters of the generators and loads are listed in Tables 4.1 and 4.2. The utility functions $\{U_{jrs}(\cdot)\}$ are set to be zero for simplicity. The upper bound on each aggregator's consumption is $P_{\text{DRA}_j}^{\max} = 50$ MW while $\mathbf{p}_{\text{BL}}^t = 5$ MW. At a base of 100 MVA, the values of the network reactances are $\{X_{16}, X_{62}, X_{25}, X_{53}, X_{34}, X_{41}\} = \{0.2, 0.3, 0.25, 0.1, 0.3, 0.4\}$ p.u. Finally, no flow limits are imposed across the network. The resulting optimization problems (4.7a) and (4.12) are modeled via `YALMIP` [90], and solved by `Gurobi` [63].

Figs. 4.4(a) and 4.4(b) illustrate the convergence performance of the proposed disaggregated bundle method vis-à-vis the disaggregated CPM. The pertinent parameters are set as $\epsilon = 10^{-3}$, $\rho(k) \equiv 0$, $\beta = 0.5$, $\boldsymbol{\mu}^{\min} = -50$, and $\boldsymbol{\mu}^{\max} = 50$ [cf. [54]]. Fig. 4.4(a) depicts the evolution of the objective values of the dual $D(\boldsymbol{\mu}(k))$ and the approximate dual $D_{\text{ap}}(\boldsymbol{\mu}(k+1))$. It is clearly seen that the bundle method converges much faster (more than three times) than its CPM counterpart. Note that due to the effect of the proximal penalty [cf. (4.11a)], quantity $D_{\text{ap}}(\boldsymbol{\mu}(k+1))$ for the bundle may not always serve as an upper bound of f^* as the one for the CPM. Finally, convergence of the



(a) (b)

Figure 4.4: (a) Convergence of the objective values. (b) Convergence of the Lagrange multipliers.

Lagrange multiplier sequence $\boldsymbol{\mu}(k)$ is shown in Fig. 4.4(b), which also corroborates the merit of the bundle method for its faster parameter convergence over the CPM. It is interesting to observe that the distance-to-optimal curve of the bundle method is quite smooth compared with the CPM one. This again illustrates the effect of the proximal regulation penalizing large deviations.

To verify the merits of the proposed CVaR-based MC, the tested power system is modified from the WECC system as illustrated in Fig. 4.3. Each of the 4 DR aggregators serves 200 residential customers. The parameters of the conventional generators are listed in Tables 4.3, while other setups are the same as the ones for MC without renewables. The convex programs (4.32) and (4.33) are modeled using the Matlab-based package `CVX` [59] with `SeDuMi` [131].

Variable characteristics of the daily power market are captured via two groups of parameters shown in Fig. 4.5: the fixed base load demand $\{\mathbf{p}_{\text{BL}}^t\}$, and the purchase prices $\{\mathbf{b}^t\}$ at the buses attached with three wind farms. The prices were obtained by scaling the real data from the MISO [92]. Two peaks of $\{\mathbf{b}^t\}$ appear during the morning 7am to 12pm, and early night 6pm to 9pm. The selling prices $\{\mathbf{s}^t\}$ were set to $\mathbf{s}^t = 0.9\mathbf{b}^t$ satisfying the convexity condition in Proposition 4.1. The rated capacity of each wind farm was set to 20 MW, yielding a 23% wind power penetration of the total power generation capacity.

Table 4.3: Conventional generator parameters. The units of a_i and b_i are \$/(MWh)^2 and \$/MWh, respectively. the rest are in MW.

Unit	a_i	b_i	$P_{G_i}^{\max}$	$P_{G_i}^{\min}$	R_i^{up}	$\mathsf{R}_i^{\text{down}}$
1	0.3	50	90	10	50	50
2	0.15	30	50	5	35	40
3	0.2	40	60	8	40	40

Wind power output samples $\{\mathbf{w}_s^t\}_{s,t}$ are needed as inputs of (4.32). These samples can be obtained either from forecasts of wind power generation, or, by using the distributions of wind speed together with the wind-speed-to-wind-power mappings [cf. [164]]. The needed samples for this simulation were obtained from the model $\mathbf{w}_s^t = \bar{\mathbf{w}}^t + \mathbf{n}_s^t, \forall t \in \mathcal{T}$. The DA wind power forecasts $\{\bar{\mathbf{w}}^t\}$ were taken from the MISO market on March 8, 2014. The forecast error $\mathbf{n}_s^t$ was assumed zero-mean white Gaussian. Possible negative-valued elements of the generated samples $\{\mathbf{w}_s^t\}_{s=1}^{N_s}$ were truncated to zero. Finally, the sample size $N_s = 200$, the probability level $\beta = 0.95$, the trade-off weight $\mu = 1$, and the primal-residual tolerance $\epsilon^{\text{pri}} = 10^{-4}$ were set for all simulations, unless otherwise stated.

Fig. 4.6 demonstrates the fast convergence of the proposed ADMM-based solver. The pertinent parameters were set to $\rho = 35$ and $\lambda_j^t(0) = p_{jrs}^t(0) = 0$. Clearly, both the cost and the primal residual converge very fast to the optimum within 10 iterations. Note that due to the infeasibility of the iterates at the beginning, the objective function starts from a value smaller than the optimum, and then monotonically converge to the latter.

Three methods were tested to show the performance of the optimal dispatch and cost: (i) the novel CVaR-based risk-limiting MC; (ii) the no risk-limiting MC with the expected wind power generation $\{\bar{\mathbf{w}}^t\}$; and (iii) the MC without wind power integration. Specifically, $\mathbf{p}_W^t = \bar{\mathbf{w}}^t$ was simply used in the nodal balance (4.32b) for (ii), while $\mathbf{p}_W^t \equiv \mathbf{0}$ for (iii). There are no CVaR-pertinent terms in the objective and constraints for the last two alternatives. For all three approaches, the generation cost $\sum_{t=1}^{T}\sum_{i=1}^{N_g} C_i(P_{G_i}^t)$ is fixed after solving (4.25). Hence, randomness of the optimal total cost stems from the transaction cost due to the stochasticity of the actual wind power generation $\{\mathbf{w}^t\}$ [cf. (4.21)].

Fig. 4.7 presents the cumulative distribution functions (CDFs) of the optimal total

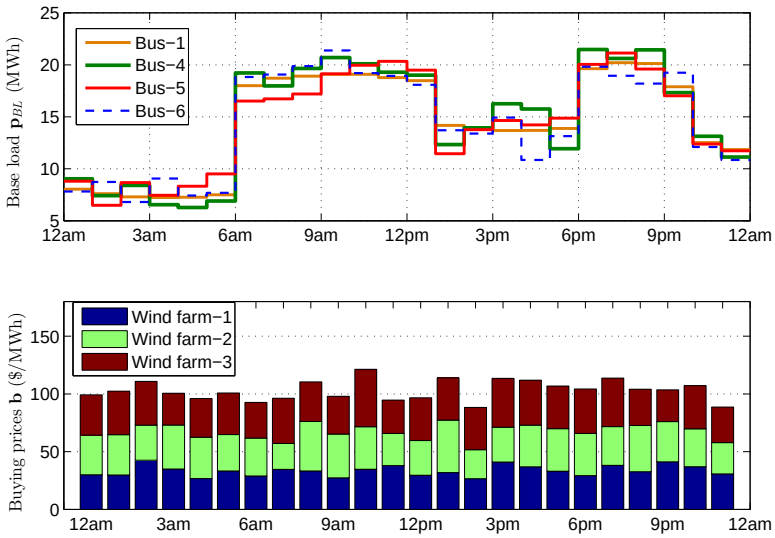


Figure 4.5: Fixed base load demand $\{\mathbf{p}_{\mathrm{BL}}^t\}$ and energy purchase prices $\{\mathbf{b}^t\}$.

costs using 100,000 i.i.d. wind samples with mean $\{\bar{\mathbf{w}}^t\}$. Clearly, the two competing alternatives always incur higher costs than the novel CVaR-based approach. The values of the mean and standard deviation (std) of the optimal total cost are listed in Table 4.4. It can be seen that, compared with the other two methods, the proposed scheme has a markedly reduced expected total cost and small changes in the std.

Figs. 4.8, 4.9, and 4.10 compare the optimal power dispatches $\{\mathbf{p}_G^t, \mathbf{p}_W^t, \mathbf{p}_{\mathrm{DRA}}^t\}_{t\in\mathcal{T}}$ of the proposed scheme with those of the scheme (ii). In Fig. 4.8, it can be clearly seen that over a single day the CVaR-based MC dispatches lower and smoother $\mathbf{p}_G$ than the one with (ii). Furthermore, for the novel method, generators 1 and 3 are dispatched to output their minimum generation $P_{G_i}^{\min}$, while the output of the generator 2 changes within its generation limits across time. Such a dispatch results from the economic incentive since the unit 2 has the lowest generation cost among all three generators [cf.

Table 4.4: Mean and standard deviation of the total cost and the conventional generation cost: Risk-limiting versus no risk-limiting dispatch. The units are all in $.

Dispatch scheme	Mean	Std	Conv. gen. cost
CVaR-based risk-limiting	44363.26	493.15	26047.66
With expected wind power	50095.68	498.13	50194.59
Without wind power	51619.24	476.25	57122.82

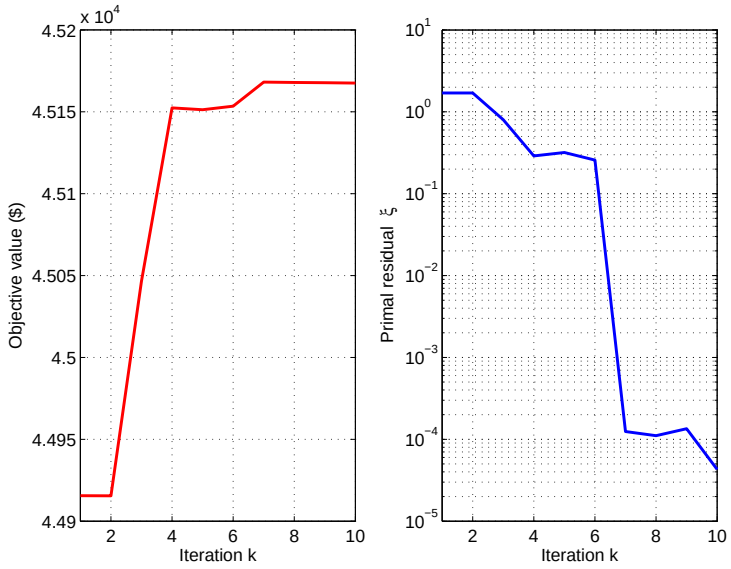


Figure 4.6: Convergence of the objective value (4.25a) and the primal residual (4.31).

Table 4.3]. On the contrary, both generators 2 and 3 fluctuate within a relatively large range in (ii), mainly to meet the variation of base load demand $\mathbf{p}_{\text{BL}}$; see Fig. 4.5.

As shown in Fig. 4.9, the novel CVaR-based approach also dispatches more $\mathbf{p}_W^t$ than that of (ii). This is because the energy purchase prices $\mathbf{b}^t$ are smaller than the conventional generation costs [cf. Table 4.1 and Fig. 4.5]. In addition, $p_{W_1}^t$ and $p_{W_2}^t$ contribute most of the committed wind power at 1pm and 2pm due to the cheaper buying prices during the corresponding slots [cf. Fig. 4.5]. Interestingly, Fig. 4.10 shows that the PHEVs are scheduled to start charging earlier for the CVaR-based MC, where $\mathbf{p}_{\text{DRA}}$ is jointly optimized with $\mathbf{p}_G$ and $\mathbf{p}_W$.

Finally, Fig. 4.11 shows the effect of the weight parameter μ on the optimal costs of the conventional generation and the CVaR-based transaction. As expected, the CVaR-based transaction cost decreases with the increase of μ. For a larger μ, less $\mathbf{p}_W^t$ is scheduled so that more wind power is likely to be sold in the RT market that yields selling revenues rather than purchase costs. Consequently, to keep the supply-demand balance, higher conventional generation cost is incurred by the increase of $\mathbf{p}_G^t$.

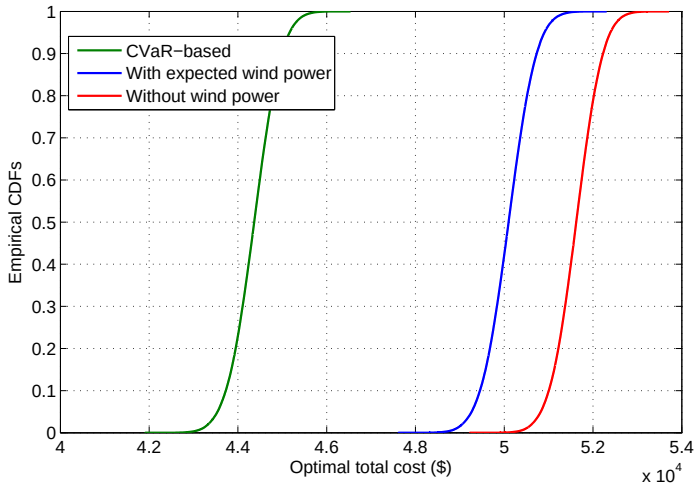


Figure 4.7: Empirical CDFs of the optimal social cost.

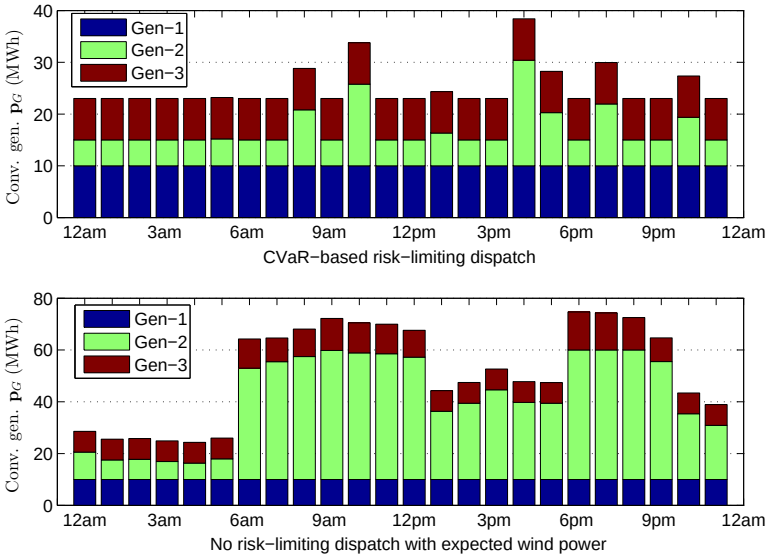


Figure 4.8: Optimal power dispatch of $\mathbf{p}_G$.

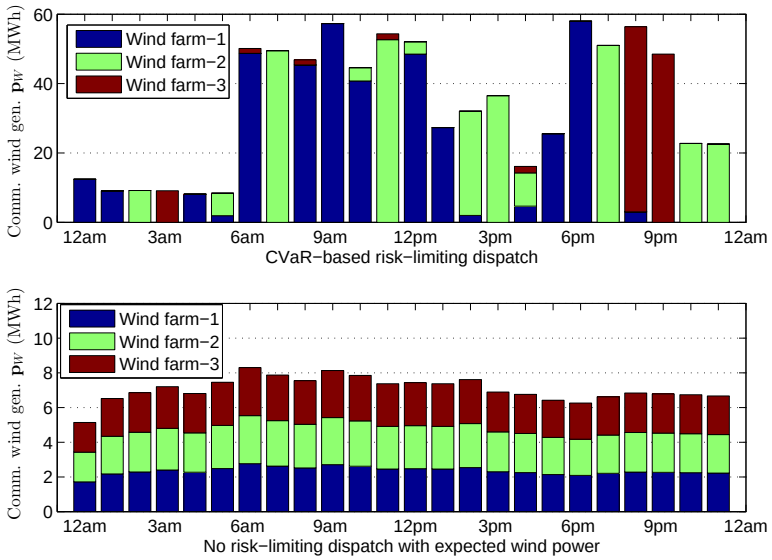


Figure 4.9: Optimal power dispatch of $\mathbf{p}_W$.

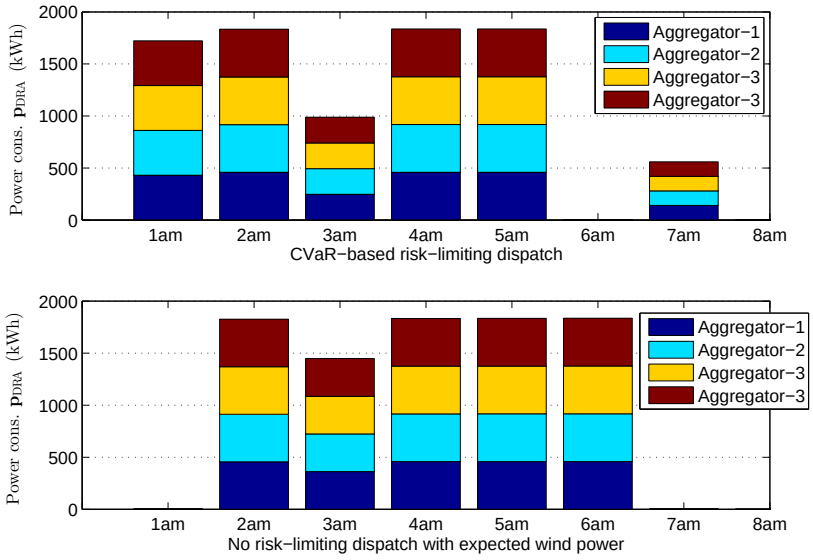


Figure 4.10: Optimal power dispatch of $\mathbf{p}_{\text{DRA}}$.

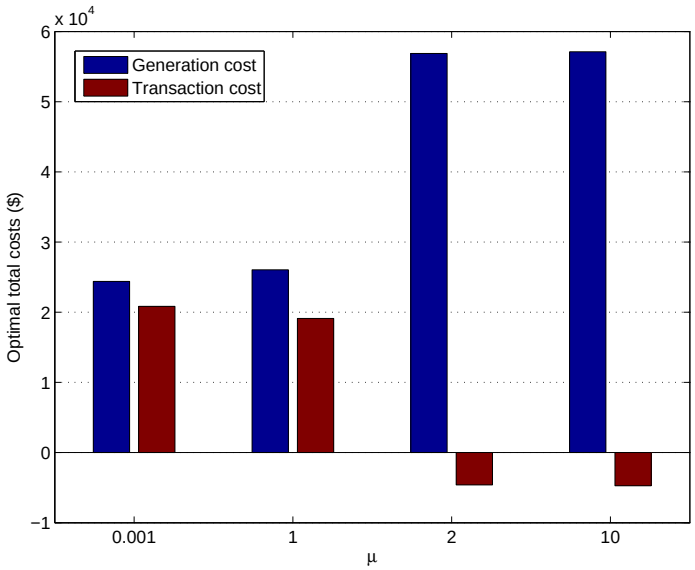


Figure 4.11: Optimal costs of conventional generation and CVaR-based transaction.

Chapter 5

Distributed Robust Resource Allocation for Wireless Networks

5.1 MIMO Cognitive Radio Networks

Consider a wireless MIMO CR network comprising K transmitter-receiver pairs $\{U_k^t, U_k^r\}$. Let M_k and N_k, $k \in \mathcal{K} := \{1, 2, \ldots, K\}$, denote the number of antennas of the k-th transmitter-receiver pair, as shown in Fig. 5.1. Further, let $\mathbf{s}_k$ denote the $M_k \times 1$ information symbol vector transmitted by U_k^t per time slot, with covariance matrix $\mathbb{E}\{\mathbf{s}_k \mathbf{s}_k^{\mathcal{H}}\} = \mathbf{I}_{M_k}$. In order to mitigate self-interference, transmitter U_k^t pre-multiplies $\mathbf{s}_k$ by a transmit-beamforming matrix $\mathbf{F}_k \in \mathbb{C}^{M_k \times M_k}$; that is, U_k^t actually transmits the $M_k \times 1$ symbol vector $\mathbf{x}_k := \mathbf{F}_k \mathbf{s}_k$. With $\mathbf{H}_{k,j} \in \mathbb{C}^{N_k \times M_j}$ denoting the U_j^t to U_k^r channel matrix, the $N_k \times 1$ symbol received at U_k^r can be written as

$$\mathbf{y}_k = \mathbf{H}_{k,k}\mathbf{x}_k + \sum_{j \neq k} \mathbf{H}_{k,j}\mathbf{x}_j + \mathbf{n}_k \tag{5.1}$$

where $\mathbf{n}_k \in \mathbb{C}^{N_k}$ is the zero-mean complex Gaussian distributed receiver noise, which is assumed independent of $\mathbf{s}_k$ and $\{\mathbf{H}_{k,j}\}$, with covariance matrix $\mathbb{E}\{\mathbf{n}_k \mathbf{n}_k^{\mathcal{H}}\} = \sigma_k^2 \mathbf{I}_{N_k}$.

Low-complexity receiver processing motivates the use of a linear filter matrix $\mathbf{W}_k \in \mathbb{C}^{M_k \times N_k}$ at U_k^r to recover $\mathbf{s}_k$ as

$$\hat{\mathbf{s}}_k := \mathbf{W}_k \mathbf{y}_k, \quad k \in \mathcal{K}. \tag{5.2}$$

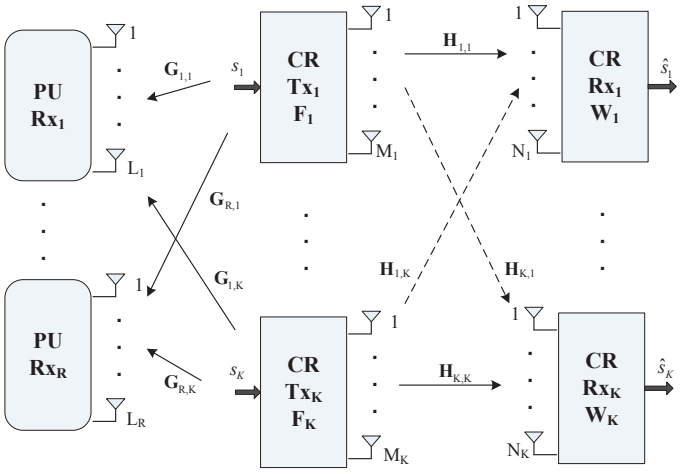


Figure 5.1: The system model for MIMO ad hoc CR networks.

Using $\mathbf{W}_k$ at U_k^r, the MSE matrix $\mathbf{E}_k := \mathbb{E}\{(\hat{\mathbf{s}}_k - \mathbf{s}_k)(\hat{\mathbf{s}}_k - \mathbf{s}_k)^{\mathcal{H}}\}$, which quantifies the reconstruction error, is given by [cf. (5.1)]

$$\mathbf{E}_k = \mathbf{W}_k \mathbf{A}_k \mathbf{W}_k^{\mathcal{H}} - \mathbf{W}_k \mathbf{H}_{k,k} \mathbf{F}_k - \mathbf{F}_k^{\mathcal{H}} \mathbf{H}_{k,k}^{\mathcal{H}} \mathbf{W}_k^{\mathcal{H}} + \mathbf{I}_{M_k} \tag{5.3}$$

where $\mathbf{A}_k := \sum_{j=1}^{K} \mathbf{H}_{k,j} \mathbf{F}_j \mathbf{F}_j^{\mathcal{H}} \mathbf{H}_{k,j}^{\mathcal{H}} + \sigma_k^2 \mathbf{I}_{N_k}$. Entry (i,i) of $\mathbf{E}_k$ represents the MSE of the i-th data stream (i-th entry of $\mathbf{s}_k$) from U_k^t to U_k^r, and $\mathrm{Tr}\{\mathbf{E}_k\}$ corresponds to the MSE of $\hat{\mathbf{s}}_k$.

5.1.1 Sum-MSE Minimization

To complete the formulation, let $\mathbf{G}_k \in \mathbb{C}^{L \times M_k}$ denote the channel between CR U_k^t and a PU receiver, equipped with L antennas, and $\iota^{\max}$ the maximum instantaneous interference that the PU can tolerate. Note that a single PU receiver is considered here. However, extension to multiple receiving PUs is straightforward; see also Remark 5.5. As in e.g., [123, 144], suppose that $\iota^{\max}$ is pre-partitioned in per-CR transmitter portions $\{\iota_k^{\max}\}$, possibly depending on QoS requirements of individual CR pairs. Then, the transmit- and receive-beamforming matrices minimizing the overall MSE can be obtained as (see also [45])

$$\underset{\{\mathbf{F}_k, \mathbf{W}_k\}_{k=1}^K}{\text{minimize}} \quad \sum_{k=1}^{K} \mathrm{Tr}\{\mathbf{E}_k\} \tag{5.4a}$$

$$\text{subject to} \quad \mathrm{Tr}\{\mathbf{F}_k\mathbf{F}_k^{\mathcal{H}}\} \leq p_k^{\max},\ k \in \mathcal{K} \tag{5.4b}$$

$$\mathrm{Tr}\{\mathbf{G}_k\mathbf{F}_k\mathbf{F}_k^{\mathcal{H}}\mathbf{G}_k^{\mathcal{H}}\} \leq \iota_k^{\max},\ k \in \mathcal{K} \tag{5.4c}$$

where $p_k^{\max}$ is the maximum transmit-power of U_k^t.

Remark 5.1 *(Adopted performance metric).* Among candidate performance metrics, the sum of MSEs from different data streams is adopted here, which has been widely employed in the beamforming literature; see e.g., [45, 141] and references therein. The relationships between MSE, bit error rate (BER) and SINR have been thoroughly considered in [110], and further investigated in [45]. Specifically, it has been shown that an improvement in the total MSE naturally translates in a lower BER. Furthermore, the sum of MSEs facilitates derivation of optimal filters, and the equivalence between minimizing the weighted sum of MSEs and maximizing the weighted sum rate has been established in [36, 132].

Unfortunately, due to lack of explicit cooperation between PU and CR nodes, CR-to-PU channels $\{\mathbf{G}_k\}$ are in general difficult to estimate accurately. As PU protection must be enforced though, it is important to take into account the CR-to-PU channel *uncertainty*, and guarantee that the interference power experienced by the PU receiver stays below a prescribed level for *any* possible (random) channel realization [42, 173]. Before developing a beamforming approach robust to inaccuracies associated with channel estimation, problem (5.4) is conveniently re-formulated first in order to reduce the number of variables involved.

For the sum-MSE cost in (5.4a), the optimum $\{\mathbf{W}_k^{\mathrm{opt}}\}$ will turn out to be expressible in closed form. To show this, note first that for fixed $\{\mathbf{F}_k\}$, (5.4) is convex in $\mathbf{W}_k$, and $\{\mathbf{W}_k^{\mathrm{opt}}\}$ can be obtained from the first-order optimality conditions. Express the Lagrangian function associated with problem (5.4) as

$$L(\mathcal{P},\mathcal{D}) = \sum_{k=1}^{K}\mathrm{Tr}\{\mathbf{E}_k\} + \sum_{k=1}^{K}\lambda_k\left(\mathrm{Tr}\{\mathbf{F}_k\mathbf{F}_k^{\mathcal{H}}\} - p_k^{\max}\right) + \sum_{k=1}^{K}\nu_k\left(\mathrm{Tr}\{\mathbf{G}_k\mathbf{F}_k\mathbf{F}_k^{\mathcal{H}}\mathbf{G}_k^{\mathcal{H}}\} - \iota_k^{\max}\right) \tag{5.5}$$

where $\mathcal{P} := \{\mathbf{F}_k, \mathbf{W}_k\}_{k=1}^K$ and $\mathcal{D} := \{\lambda_k, \nu_k\}_{k=1}^K$ collects the primal and dual variables, respectively. Then, by equating the complex gradient $\partial L(\mathcal{P},\mathcal{D})/\partial\mathbf{W}_k^*$ to zero, matrix

$\mathbf{W}_k^{\text{opt}}$ is expressed as

$$\mathbf{W}_k^{\text{opt}} = \mathbf{F}_k^{\mathcal{H}} \mathbf{H}_{k,k}^{\mathcal{H}} \mathbf{A}_k^{-1}, \quad k \in \mathcal{K}. \tag{5.6}$$

Clearly, the optimal set $\{\mathbf{W}_k^{\text{opt}}\}$ does not depend on channels $\{\mathbf{G}_k\}$, but only on $\{\mathbf{H}_{k,j}\}$.

Substituting $\{\mathbf{W}_k^{\text{opt}}\}$ into (5.4a), and using the covariance $\mathbf{Q}_k := \mathbb{E}\{\mathbf{x}_k \mathbf{x}_k^{\mathcal{H}}\} = \mathbf{F}_k \mathbf{F}_k^{\mathcal{H}}$ as a matrix optimization variable, it follows that problem (5.4) can be equivalently re-written as

$$\underset{\{\mathbf{Q}_k \succeq \mathbf{0}\}}{\text{maximize}} \quad \sum_{k=1}^{K} u_k \left(\{\mathbf{Q}_k\}\right) \tag{5.7a}$$

$$\text{subject to} \quad \text{Tr}\{\mathbf{Q}_k\} \leq p_k^{\max}, \ k \in \mathcal{K} \tag{5.7b}$$

$$\text{Tr}\{\mathbf{G}_k \mathbf{Q}_k \mathbf{G}_k^{\mathcal{H}}\} \leq \iota_k^{\max}, \ k \in \mathcal{K} \tag{5.7c}$$

where the per-CR link utility $u_k \left(\{\mathbf{Q}_k\}\right)$ is given by

$$u_k \left(\{\mathbf{Q}_k\}\right) := \text{Tr}\left\{\mathbf{H}_{k,k} \mathbf{Q}_k \mathbf{H}_{k,k}^{\mathcal{H}} \left(\mathbf{H}_{k,k} \mathbf{Q}_k \mathbf{H}_{k,k}^{\mathcal{H}} + \mathbf{R}_{k,k}\right)^{-1}\right\} \tag{5.8}$$

with $\mathbf{R}_{k,k} := \sum_{i \neq k} \mathbf{H}_{k,i} \mathbf{Q}_i \mathbf{H}_{k,i}^{\mathcal{H}} + \sigma_k^2 \mathbf{I}_{N_k}$. One remark is in order regarding the problem (5.7).

Remark 5.2 *(Conventional MIMO networks).* Upon discarding the interference constraints (5.7c), the resulting beamforming problems along with their centralized and distributed solvers can be considered also for non-CR MIMO ad-hoc and cellular networks in downlink or uplink operation.

Channels $\{\mathbf{G}_k\}$ must be perfectly known in order to solve (5.7). A robust version of (5.7), which accounts for imperfect channel knowledge, is dealt with in the next section.

5.1.2 Robust Interference Constraint

In typical CR scenarios, CR-to-PU channels are challenging to estimate accurately. In fact, CR and PU nodes do not generally cooperate [172], thus rendering channel estimation challenging. To model estimation inaccuracies, consider expressing the CR-to-PU channel matrix $\mathbf{G}_k$ as

$$\mathbf{G}_k = \widehat{\mathbf{G}}_k + \Delta \mathbf{G}_k \,, \quad k \in \mathcal{K} \tag{5.9}$$

where $\widehat{\mathbf{G}}_k$ is the estimated channel, which is known at CR transmitter U_k^t, and $\{\Delta\mathbf{G}_k\}$ captures the underlying channel uncertainty [6,173]. Specifically, the error matrix $\Delta\mathbf{G}_k$ is assumed to take values from the bounded set

$$\mathcal{G}_k := \left\{\Delta\mathbf{G}_k \mid \mathrm{Tr}\{\Delta\mathbf{G}_k\Delta\mathbf{G}_k^{\mathcal{H}}\} \leq \epsilon_k^2\right\},\ k \in \mathcal{K} \tag{5.10}$$

where $\epsilon_k > 0$ specifies the radius of $\mathcal{G}_k$, and thus reflects the degree of uncertainty associated with $\widehat{\mathbf{G}}_k$. The set in (5.10) can be readily extended to the general ellipsoidal uncertainty model [26, Ch. 4]. Such an uncertainty model properly resembles the case where a time division duplex (TDD) strategy is adopted by the PU system, and CRs have prior knowledge of the PUs' pilot sequence(s). But even without training symbols, CR-to-PU channel estimates can be formed using the deterministic path loss coefficients, and the size of the uncertainty region can be deduced from fading channel statistics. Compared to [42], the norm-bounded uncertainty model leads to *worst-case* interference constraints that ensure PU protection for any realization of the uncertain portion of the propagation channels.

Based on (5.10), a robust interference constraint can be written as

$$\mathrm{Tr}\{(\widehat{\mathbf{G}}_k + \Delta\mathbf{G}_k)\mathbf{Q}_k(\widehat{\mathbf{G}}_k + \Delta\mathbf{G}_k)^{\mathcal{H}}\} \leq \iota_k^{\max},\ \forall\ \Delta\mathbf{G}_k \in \mathcal{G}_k,\ k \in \mathcal{K} \tag{5.11}$$

and consequently, a robust counterpart of (5.7) can be formulated as follows

$$\underset{\{\mathbf{Q}_k \succeq \mathbf{0}\}}{\text{maximize}} \quad \sum_{k=1}^{K} u_k\left(\{\mathbf{Q}_k\}\right) \tag{5.12a}$$

$$\text{subject to} \quad \mathrm{Tr}\{\mathbf{Q}_k\} \leq p_k^{\max},\ k \in \mathcal{K} \tag{5.12b}$$

$$\mathrm{Tr}\{\mathbf{G}_k\mathbf{Q}_k\mathbf{G}_k^{\mathcal{H}}\} \leq \iota_k^{\max},\ \forall\ \Delta\mathbf{G}_k \in \mathcal{G}_k,\ k \in \mathcal{K}. \tag{5.12c}$$

Clearly, once $\{\mathbf{Q}_k^{\mathrm{opt}}\}$ are found by solving (5.12), the wanted $\{\mathbf{W}_k^{\mathrm{opt}}\}$ can be readily obtained via (5.6), since $\mathbf{A}_k := \sum_{j=1}^{K}\mathbf{H}_{k,j}\mathbf{Q}_j\mathbf{H}_{k,j}^{\mathcal{H}} + \sigma_k^2\mathbf{I}_{N_k}$. However, $\sum_k u_k\left(\{\mathbf{Q}_k\}\right)$ is non-convex in $\{\mathbf{Q}_k\}$, and hence problem (5.12) is hard to solve in general. Additionally, constraints (5.11) are not in a tractable form, which motivates their transformation. These issues are addressed in the next section. But first, two remarks are in order.

Remark 5.3 *(Uncertain MIMO channels).* In an underlay hierarchical spectrum access setup, it is very challenging (if not impossible) for the CRs to obtain accurate estimates

of the CR-to-PU channels. In fact, since the PUs hold the spectrum license, they have no incentive to feed back CR-to-PU channel estimates to the CR system [172]. Hence, in lieu of explicit CR-PU cooperation, CRs have to resort to crude or blind estimates of their channels with PUs. On the other hand, sufficient time for training along with sophisticated estimation algorithms render the CR-to-CR channels easier to estimate. This explains why similar to relevant works [6, 42, 55, 56, 144, 173], CR-to-CR channels are assumed known, while CR-to-PU channels are taken as uncertain in CR-related optimization methods. Limited-rate channel state information that can become available e.g., with quantized CR-to-CR channels [42,45,141], can be considered in future research but goes beyond the scope of the present thesis.

Remark 5.4 *(Radius of the uncertainty region).* In practice, radius and shape of the uncertainty region have to be tailored to the specific channel estimation approach implemented at the CRs, and clearly depend on the second-order channel error statistics. For example, if $\Delta\mathbf{G}_k$ has zero mean and covariance matrix $\mathbf{\Sigma}_{\Delta\mathbf{G}_k} = \hat{\sigma}_k^2\mathbf{I}$, where $\hat{\sigma}_k^2$ depends on the receiver noise power, and the transmit-power of the PU (see, e.g. [18]), then the radius of the uncertainty region can be set to $\epsilon_k^2 = \kappa\xi_k\hat{\sigma}_k^2$, where ξ_k denotes the path loss coefficient, and $\kappa > 0$ a parameter that controls how strict the PU protection is. Alternatively, the model $\epsilon_k^2 = \kappa\hat{\sigma}_k^2\|\hat{\mathbf{G}}_k\|_F^2$ can be utilized [144]. If $\mathbf{\Sigma}_{\Delta\mathbf{G}_k} \neq \hat{\sigma}_k^2\mathbf{I}$, then the uncertainty region can be set to $\mathcal{G}_k := \left\{\Delta\mathbf{G}_k \mid \mathrm{Tr}\{\Delta\mathbf{G}_k\mathbf{\Sigma}_{\Delta\mathbf{G}_k}^{-1}\Delta\mathbf{G}_k^{\mathcal{H}}\} \leq \kappa\right\}$ [158] and the robust constraint (5.11) can be modified accordingly. Similar models are also considered in [125].

5.2 Distributed Robust CR Beamforming

To cope with the non-convexity of the utility function (5.12a), a block-coordinate ascent solver is developed in this section. Define first the sum of all but the k-th utility as $f_k(\mathbf{Q}_k, \mathbf{Q}_{-k}) := \sum_{j\neq k} u_j$, with $\mathbf{Q}_{-k} := \{\mathbf{Q}_j | j \neq k\}$. Notice that $u_k(\cdot)$ is concave and $f_k(\cdot)$ is convex in $\mathbf{Q}_k$ [cf. Lemma 5.2]. Then, (5.12) can be regarded as a difference of convex functions (d.c.) program, whenever only a single variable $\mathbf{Q}_k$ is optimized and $\mathbf{Q}_{-k}$ is kept fixed. This motivates the so-termed concave-convex procedure [153], which belongs to the majorization-minimization class of algorithms [70], to solve problem (5.12) through a sequence of convex problems, one per matrix variable $\mathbf{Q}_k$. Specifically,

the idea is to linearize the convex function $f_k(\cdot)$ around a feasible point $\tilde{\mathbf{Q}}_k$, and thus to (locally) approximate the objective (5.12a) as (see also [79, 116])

$$\begin{aligned}\sum_{k=1}^{K} u_k\left(\{\mathbf{Q}_k\}\right) &= u_k\left(\{\mathbf{Q}_k\}\right) + f_k(\mathbf{Q}_k, \mathbf{Q}_{-k}) \\ &\approx u_k\left(\{\mathbf{Q}_k\}\right) + f_k(\tilde{\mathbf{Q}}_k, \mathbf{Q}_{-k}) + \mathrm{Tr}\left\{\mathbf{D}_k^{\mathcal{H}}(\mathbf{Q}_k - \tilde{\mathbf{Q}}_k)\right\} \end{aligned} \tag{5.13}$$

where

$$\mathbf{D}_k := \nabla_{\mathbf{Q}_k} f_k(\tilde{\mathbf{Q}}_k, \mathbf{Q}_{-k}) := \left.\frac{\partial f_k}{\partial \mathbf{Q}_k^*}\right|_{\mathbf{Q}_k = \tilde{\mathbf{Q}}_k}. \tag{5.14}$$

Therefore, for fixed $\mathbf{Q}_{-k}$, matrix $\mathbf{Q}_k$ can be obtained by solving the following sub-problem

$$\begin{aligned} \underset{\mathbf{Q}_k}{\text{maximize}} \quad & u_k\left(\mathbf{Q}_k, \mathbf{Q}_{-k}\right) + \mathrm{Tr}\left\{\mathbf{D}_k^{\mathcal{H}}\mathbf{Q}_k\right\} && \text{(5.15a)} \\ \text{subject to} \quad & \mathbf{Q}_k \succeq \mathbf{0} && \text{(5.15b)} \\ & \mathrm{Tr}\{\mathbf{Q}_k\} \le p_k^{\max} && \text{(5.15c)} \\ & \mathrm{Tr}\{\mathbf{G}_k\mathbf{Q}_k\mathbf{G}_k^{\mathcal{H}}\} \le \iota_k^{\max},\ \forall\ \Delta\mathbf{G}_k \in \mathcal{G}_k && \text{(5.15d)} \end{aligned}$$

where

$$\mathbf{D}_k := -\sum_{j\neq k} \mathbf{H}_{j,k}^{\mathcal{H}}\mathbf{B}_j^{-1}\mathbf{V}_j\mathbf{B}_j^{-1}\mathbf{H}_{j,k}\Big|_{\mathbf{Q}_k=\tilde{\mathbf{Q}}_k} \tag{5.16}$$

$$\mathbf{B}_j := \sum_{i=1}^{K} \mathbf{H}_{j,i}\mathbf{Q}_i\mathbf{H}_{j,i}^{\mathcal{H}} + \sigma_j^2\mathbf{I}_{N_j} \tag{5.17}$$

$$\mathbf{V}_j := \mathbf{H}_{j,j}\mathbf{Q}_j\mathbf{H}_{j,j}^{\mathcal{H}}. \tag{5.18}$$

Note that the complex gradient matrix (5.16) can be derived as follows. Using that $\frac{\partial \mathrm{Tr}\left\{\mathbf{X}^{\mathcal{H}}\mathbf{A}\right\}}{\partial \mathbf{X}^*} = \mathbf{A}$ [91], and letting $[\mathbf{A}]_{mn}$ denote the (m,n)-th entry of matrix $\mathbf{A}$, it follows that

$$\frac{\partial [\mathbf{B}_j^{\mathcal{H}}]_{st}}{\partial \mathbf{Q}_k^*} = \frac{\partial \mathrm{Tr}\left\{\mathbf{e}_s^{\mathcal{H}}\mathbf{H}_{j,k}\mathbf{Q}_k^{\mathcal{H}}\mathbf{H}_{j,k}^{\mathcal{H}}\mathbf{e}_t\right\}}{\partial \mathbf{Q}_k^*} = \mathbf{H}_{j,k}^{\mathcal{H}}\mathbf{e}_t\mathbf{e}_s^{\mathcal{H}}\mathbf{H}_{j,k} \tag{5.19}$$

$$\frac{\partial [\mathbf{B}_j^{\mathcal{H}}]_{st}}{\partial [\mathbf{Q}_k^*]_{mn}} = \mathbf{e}_m^{\mathcal{H}}\mathbf{H}_{j,k}^{\mathcal{H}}\mathbf{e}_t\mathbf{e}_s^{\mathcal{H}}\mathbf{H}_{j,k}\mathbf{e}_n = \mathbf{e}_s^{\mathcal{H}}\mathbf{H}_{j,k}\mathbf{e}_n\mathbf{e}_m^{\mathcal{H}}\mathbf{H}_{j,k}^{\mathcal{H}}\mathbf{e}_t \tag{5.20}$$

which can be written in a compact form as $\frac{\partial \mathbf{B}_j^{\mathcal{H}}}{\partial [\mathbf{Q}_k^*]_{mn}} = \mathbf{H}_{j,k}\mathbf{e}_n\mathbf{e}_m^{\mathcal{H}}\mathbf{H}_{j,k}^{\mathcal{H}}$. Then, the identity $\frac{\partial f}{\partial \mathbf{X}^*} = -\mathbf{X}^{-1}\left(\frac{\partial f}{\partial (\mathbf{X}^{-1})^*}\right)\mathbf{X}^{-1}$ [91], which holds for any Hermitian positive definite matrix $\mathbf{X}$, is used to obtain

$$\frac{\partial u_j}{\partial \mathbf{B}_j^*} = -\mathbf{B}_j^{-1}\frac{\partial \mathrm{Tr}\left\{\mathbf{V}_j(\mathbf{B}_j^{-1})^{\mathcal{H}}\right\}}{\partial (\mathbf{B}_j^{-1})^*}\mathbf{B}_j^{-1} = -\mathbf{B}_j^{-1}\mathbf{V}_j\mathbf{B}_j^{-1}. \tag{5.21}$$

Using now the chain rule, one arrives at

$$\begin{aligned}\frac{\partial u_j}{\partial [\mathbf{Q}_k^*]_{mn}} &= \mathrm{Tr}\left\{\left(\frac{\partial u_j}{\partial \mathbf{B}_j^{\mathcal{H}}}\right)'\frac{\partial \mathbf{B}_j^{\mathcal{H}}}{\partial [\mathbf{Q}_k^*]_{mn}}\right\} \\ &= \mathrm{Tr}\left\{-\mathbf{e}_m^{\mathcal{H}}\mathbf{H}_{j,k}^{\mathcal{H}}\mathbf{B}_j^{-1}\mathbf{V}_j\mathbf{B}_j^{-1}\mathbf{H}_{j,k}\mathbf{e}_n\right\}\end{aligned} \tag{5.22}$$

which readily leads to the desired result

$$\frac{\partial u_j}{\partial \mathbf{Q}_k^*} = -\mathbf{H}_{j,k}^{\mathcal{H}}\mathbf{B}_j^{-1}\mathbf{V}_j\mathbf{B}_j^{-1}\mathbf{H}_{j,k}. \tag{5.23}$$

At each iteration $n = 1, 2, \ldots$, the block coordinate ascent solver amounts to updating the covariance matrices $\{\mathbf{Q}_k\}$ in a round robin fashion via (5.15), where the solution obtained at the $(n-1)$-st iteration are exploited to compute the complex gradient (5.16). The term $\mathrm{Tr}\left\{\mathbf{D}_k^{\mathcal{H}}\mathbf{Q}_k\right\}$ discourages a *selfish* behavior of the k-th CR-to-CR link, which would otherwise try to simply minimize its own MSE, as in the game-theoretic formulations of [123] and [144]. In the next subsection, the robust interference constraint will be translated to a tractable form, and (5.15) will be re-stated accordingly.

5.2.1 Equivalent Robust Interference Constraint

Constraint (5.15d) renders (5.15) a *semi-infinite* program [cf. [12, Ch. 3]]. An equivalent constraint in the LMI form will be derived next, thus turning (5.15) into an equivalent *semi-definite* program (SDP), which can be efficiently solved in polynomial time by standard interior point methods. To this end, the following lemma is needed.

Lemma 5.1 *(S-Procedure [26, p. 655]). Consider* $\mathbf{A}, \mathbf{D} \in \mathbb{H}^{n\times n}, \mathbf{b} \in \mathbb{C}^n, c, e \in \mathbb{R}$*, and assume the interior condition holds, i.e., there exists an* $\bar{\mathbf{x}}$ *satisfying* $\bar{\mathbf{x}}^{\mathcal{H}}\mathbf{D}\bar{\mathbf{x}} < e$*. Then, the inequality*

$$\mathbf{x}^{\mathcal{H}}\mathbf{A}\mathbf{x} + 2\Re(\mathbf{b}^{\mathcal{H}}\mathbf{x}) + c \geq 0,\ \forall\ \mathbf{x}^{\mathcal{H}}\mathbf{D}\mathbf{x} \leq e \tag{5.24}$$

holds if and only if there exists $\theta \geq 0$ *such that*

$$\begin{bmatrix} \theta\mathbf{D}+\mathbf{A} & \mathbf{b} \\ \mathbf{b}^{\mathcal{H}} & c-e\theta \end{bmatrix} \succeq \mathbf{0}\,. \tag{5.25}$$

Using Lemma 5.1, the robust constraint (5.15d) can be equivalently reformulated as follows.

Proposition 5.1 *There exists* $\theta_k \geq 0$*, so that the robust interference constraint* (5.15d) *is equivalent to the following LMI:*

$$\begin{bmatrix} \theta_k\mathbf{I}_{L\times M_k}-(\mathbf{I}_L\otimes\mathbf{Q}_k) & -\mathrm{vec}(\mathbf{Q}_k^{\mathcal{H}}\widehat{\mathbf{G}}_k^{\mathcal{H}}) \\ -\mathrm{vec}(\mathbf{Q}_k^{\mathcal{H}}\widehat{\mathbf{G}}_k^{\mathcal{H}})^{\mathcal{H}} & \iota_k^{\max}-\epsilon_k^2\theta_k \\ & -\mathrm{Tr}\{\widehat{\mathbf{G}}_k\mathbf{Q}_k\widehat{\mathbf{G}}_k^{\mathcal{H}}\} \end{bmatrix} \succeq \mathbf{0}\,. \tag{5.26}$$

Proof: Using the properties of the trace operator $\mathrm{Tr}(\mathbf{Z}^{\mathcal{H}}\mathbf{A}\mathbf{Z}) = \mathrm{vec}(\mathbf{Z})^{\mathcal{H}}(\mathbf{I}\otimes\mathbf{A})\mathrm{vec}(\mathbf{Z})$ and $\mathrm{Tr}(\mathbf{B}^{\mathcal{H}}\mathbf{Z}) = \mathrm{vec}(\mathbf{B})^{\mathcal{H}}\mathrm{vec}(\mathbf{Z})$, constraint (5.15d) can be re-written as

$$\begin{aligned} -\mathbf{g}_k^{\mathcal{H}}\left(\mathbf{I}_L\otimes\mathbf{Q}_k\right)\mathbf{g}_k - 2\Re\left(\mathrm{vec}(\mathbf{Q}_k^{\mathcal{H}}\widehat{\mathbf{G}}_k^{\mathcal{H}})^{\mathcal{H}}\mathbf{g}_k\right) \\ +\iota_k^{\max}-\mathrm{Tr}\{\widehat{\mathbf{G}}_k\mathbf{Q}_k\widehat{\mathbf{G}}_k^{\mathcal{H}}\}\geq 0,\ \forall\ \|\mathbf{g}_k\|_2\leq\epsilon_k \end{aligned} \tag{5.27}$$

where $\mathbf{g}_k := \mathrm{vec}(\Delta\mathbf{G}_k^{\mathcal{H}})$. Then, applying Lemma 5.1 to (5.27) yields readily (5.26). □

Proposition 5.2 *Problem* (5.15) *can be equivalently re-written as the following SDP form:*

$$\underset{\mathbf{Q}_k\succeq\mathbf{0},\mathbf{T},\theta_k\geq 0}{\text{minimize}} \quad \mathrm{Tr}\left\{\mathbf{T}\right\}-\mathrm{Tr}\left\{\mathbf{D}_k^{\mathcal{H}}\mathbf{Q}_k\right\} \tag{5.28a}$$

$$\text{subject to} \quad \mathrm{Tr}\{\mathbf{Q}_k\}\leq p_k^{\max} \tag{5.28b}$$

$$\begin{bmatrix} \mathbf{H}_{k,k}\mathbf{Q}_k\mathbf{H}_{k,k}^{\mathcal{H}}+\mathbf{R}_{k,k} & \mathbf{R}_{k,k}^{1/2} \\ \mathbf{R}_{k,k}^{1/2} & \mathbf{T} \end{bmatrix} \succeq \mathbf{0} \tag{5.28c}$$

$$\begin{bmatrix} \theta_k\mathbf{I}_{L\times M_k}-(\mathbf{I}_L\otimes\mathbf{Q}_k) & -\mathrm{vec}(\mathbf{Q}_k^{\mathcal{H}}\widehat{\mathbf{G}}_k^{\mathcal{H}}) \\ -\mathrm{vec}(\mathbf{Q}_k^{\mathcal{H}}\widehat{\mathbf{G}}_k^{\mathcal{H}})^{\mathcal{H}} & \iota_k^{\max}-\epsilon_k^2\theta_k \\ & -\mathrm{Tr}\{\widehat{\mathbf{G}}_k\mathbf{Q}_k\widehat{\mathbf{G}}_k^{\mathcal{H}}\} \end{bmatrix} \succeq \mathbf{0}\,. \tag{5.28d}$$

Proof: First, note that [cf. (5.8)]

$$u_k\left(\mathbf{Q}_k,\mathbf{Q}_{-k}\right) = \mathrm{Tr}\left\{\mathbf{I}_{N_k}-\mathbf{R}_{k,k}\left(\mathbf{H}_{k,k}\mathbf{Q}_k\mathbf{H}_{k,k}^{\mathcal{H}}+\mathbf{R}_{k,k}\right)^{-1}\right\}.$$

Algorithm 5.1 Centralized robust sum-MSE minimization

1: Collect all channel matrices $\{\mathbf{H}_{j,k}\}$, and noise powers $\{\sigma_k^2\}$
2: Collect all CR-to-PU channel matrices $\{\widehat{\mathbf{G}}_k\}$, and confidence intervals $\{\epsilon_k\}$
3: Initialize $\mathbf{Q}_k^{(0)} = \mathbf{0}, \forall\ k \in \mathcal{K}$
4: **repeat** $(n = 1, 2, \ldots)$
5: **for** $k = 1, 2, \ldots, K$ **do**
6: Compute $\mathbf{D}_k^{(n)}$ via (5.16)
7: Update $\mathbf{Q}_k^{(n)}$ by solving (5.28) [(5.46) for the proximal point-based method]
8: **end for**
9: **until** $\mathcal{U}\left(\mathbf{Q}^{(n)}\right) - \mathcal{U}\left(\mathbf{Q}^{(n-1)}\right) < \upsilon$
10: Calculate $\{\mathbf{W}_k^{\text{opt}}\}$ via (5.6)
11: Broadcast optimal transmit- and receive-beamformers

Thus, (5.15) is equivalent to

$$\begin{aligned} \underset{\mathbf{Q}_k}{\text{minimize}} \quad & \text{Tr}\left\{\mathbf{R}_{k,k}^{1/2}\left(\mathbf{H}_{k,k}\mathbf{Q}_k\mathbf{H}_{k,k}^{\mathcal{H}} + \mathbf{R}_{k,k}\right)^{-1}\mathbf{R}_{k,k}^{1/2}\right\} - \text{Tr}\left\{\mathbf{D}_k^{\mathcal{H}}\mathbf{Q}_k\right\} \\ \text{subject to} \quad & \text{(5.15b)} - \text{(5.15d)}. \end{aligned}$$

Then, an auxiliary matrix variable $\mathbf{Y}$ is introduced such that

$$\mathbf{Y} \succeq \mathbf{R}_{k,k}^{1/2}\left(\mathbf{H}_{k,k}\mathbf{Q}_k\mathbf{H}_{k,k}^{\mathcal{H}} + \mathbf{R}_{k,k}\right)^{-1}\mathbf{R}_{k,k}^{1/2}$$

which can be equivalently recast as (5.28c) by using the Schur complement [26]. Combining the LMI form of the robust interference constraint (5.26), the formulation of (5.28) follows immediately. □

Problem (5.12) can be solved in a *centralized* fashion upon collecting CR-to-CR channels $\{\mathbf{H}_{j,k}\}$, CR-to-PU estimated channels $\{\widehat{\mathbf{G}}_k\}$, and confidence intervals $\{\epsilon_k\}$ at a CR fusion center. The optimal transmit-covariance matrices can be found at the fusion center by solving (5.28), and sent back to all CRs. This centralized scheme is tabulated as Algorithm 5.1, where $\mathbf{Q}_k^{(n)}$ denotes the transmit-covariance matrix of CR U_k^t at iteration n of the block coordinate ascent algorithm; $\mathbf{Q}^{(n)} := \left(\mathbf{Q}_1^{(n)}, \ldots, \mathbf{Q}_K^{(n)}\right)$ represents the set of transmit-covariance matrices at iteration n; $\mathcal{U}(\cdot)$ is the objective function (5.12a). A simple stopping criterion for terminating the iterations is $\mathcal{U}\left(\mathbf{Q}^{(n)}\right) - \mathcal{U}\left(\mathbf{Q}^{(n-1)}\right) < \upsilon$, where $\upsilon > 0$ denotes a preselected threshold.

Algorithm 5.2 Distributed on-line robust sum-MSE minimization

Initialize $\mathbf{Q}_k^{(0)} = \mathbf{0}, \forall\ k \in \mathcal{K}$
repeat $(n = 1, 2, \ldots)$
  for $k = 1, 2, \ldots, K$ **do**
    U_k^t acquires $\mathbf{H}_{j,k}$ from its neighboring U_j^r
    transmit $\{\mathbf{B}_k^{(n)}, \mathbf{V}_k^{(n)}\}$ to neighboring nodes
    receive $\{\mathbf{B}_j^{(n)}, \mathbf{V}_j^{(n)}\}_{j \neq k}$ from neighboring nodes
    Compute $\mathbf{D}_k^{(n)}$ via (5.16)
    Measure $\mathbf{R}_{k,k}^{(n)}$
    Update $\mathbf{Q}_k^{(n)}$ by solving (5.28) [(5.46) for the proximal point-based method]
    Update $\mathbf{W}_k^{(n)}$ via (5.6)
    Transmit and receive signals using $\mathbf{Q}_k^{(n)}$ and $\mathbf{W}_k^{(n)}$
  end for
until $u_k\left(\mathbf{Q}^{(n)}\right) - u_k\left(\mathbf{Q}^{(n-1)}\right) < \upsilon, \forall\ k \in \mathcal{K}$

To alleviate the high communication cost associated with the centralized setup, and ensure scalability with regards to network size and enhanced robustness to fusion center failure, a distributed optimization algorithm is generally desirable. It can be noticed that the proposed coordinate ascent approach lends itself to a *distributed* optimization procedure that can be implemented in an *on-line* fashion. Specifically, each CR U_k^t can update locally $\mathbf{Q}_k$ via (5.28) based on a measurement of the interference $\mathbf{R}_{k,k}^{(n)}$ [79], and the following information necessary to compute the complex gradient (5.16): *i)* its covariance matrix $\mathbf{Q}_k^{(n-1)}$ obtained at the previous iteration; *ii)* matrices $\{\mathbf{B}_j^{(n)}\}$ and $\{\mathbf{V}_j^{(n)}\}$ obtained from the neighboring CR links via local message passing. Furthermore, it is clear that the terms in (5.16) corresponding to CRs located far away from CR U_k^t are negligible due to the path loss effect in channel $\{\mathbf{H}_{j,k}\}_{j \neq k}$; hence, summation in (5.16) is only limited to the interfering CRs, and consequently, matrices $\{\mathbf{B}_j^{(n)}\}$ and $\{\mathbf{V}_j^{(n)}\}$ need to be exchanged only locally. The overall distributed scheme is tabulated as Algorithm 5.2. The on-line implementation of the iterative optimization allows tracking of slow variations of the channel matrices; in this case, cross-channels $\{\mathbf{H}_{j,k}\}$ in Algorithm 5.2 need to be re-acquired whenever a change is detected. Finally, notice that instead of updating the transmit-covariances in a Gauss-Seidel fashion, Jacobi iterations or asynchronous schemes [14] can be alternatively employed.

Remark 5.5 *(Multiple PU receivers).* For ease of exposition, the formulated robust optimization problems consider a single PU receiver. Clearly, in case of $N_{\mathrm{PU}} > 1$ receiving PU devices, or when a grid of N_{PU} potential PU locations is obtained from the sensing phase [41], a robust interference constraint for each of the KN_{PU} CR-to-PU links must be included in (5.12). As for (5.28), it is still an SDP, but with N_{PU} LMI constraints (5.28d), and one additional optimization variable (θ_k) per PU receiver.

Remark 5.6 *(Network synchronization).* Similar to [6, 36, 45, 55, 56, 79, 110, 116, 123, 132, 141, 144, 157–159, 173], time synchronization is assumed to have been acquired. In practice, accurate time synchronization among the CR transmitters can be attained (and maintained during operation) using e.g., pairwise broadcast synchronization protocols [105], consensus-based methods [155], or mutual network synchronization approaches [117]. To this end, CRs have to exchange synchronization beacons on a regular basis; clearly, the number of time slots occupied by the transmission of these beacons depends on the particular algorithm implemented, the CR network size, and the targeted synchronization accuracy. For example, the algorithm in [105] entails two message exchanges per transmitter pairs, while the message-passing overhead of consensus-based methods generally depends on the wanted synchronization accuracy [155]. Since the CR network operates in an underlay setup, this additional message passing can be performed over the primary channel(s). Alternatively, a CR control channel can be employed to avoid possible synchronization errors due to the interference inflicted by the active PU transmitters.

5.2.2 Convergence of Block Coordinate Ascent

Since the original optimization problem (5.12) is non-convex, convergence of the block coordinate ascent with local convex approximation has to be analytically established. To this end, recall that (5.15) and (5.28) are equivalent; thus, convergence can be asserted by supposing that (5.15) is solved per Gauss-Seidel iteration instead of (5.28). The following lemma is needed first.

Lemma 5.2 *For each $k \in \mathcal{K}$, the feasible set of problem* (5.15)*, namely $\mathcal{Q}_k := \{\mathbf{Q}_k | \mathbf{Q}_k \in (5.15b) - (5.15d)\}$, is convex. The real-valued function $f_k(\mathbf{Q}_k, \mathbf{Q}_{-k})$ is convex in $\mathbf{Q}_k$ over the feasible set $\mathcal{Q}_k$, when the set $\mathbf{Q}_{-k} := \{\mathbf{Q}_j, j \neq k\}$ is fixed.*

Proof: First, convexity of $\mathcal{Q}_k$ can be readily proved by the definition of a convex set [26, Ch. 2]. Re-write the function $u_j(\mathbf{Q}_k, \mathbf{Q}_{-k})$ as [cf. (5.8) and (5.18)]

$$u_j(\mathbf{Q}_k, \mathbf{Q}_{-k}) = \mathrm{Tr}\left\{\mathbf{V}_j^{1/2}\mathbf{P}_j^{-1}(\mathbf{Q}_k)\mathbf{V}_j^{1/2}\right\} \tag{5.29}$$

where

$$\mathbf{P}_j(\mathbf{Q}_k) = \mathbf{H}_{j,k}\mathbf{Q}_k\mathbf{H}_{j,k}^{\mathcal{H}} + \sum_{i\neq k}\mathbf{H}_{j,i}\mathbf{Q}_i\mathbf{H}_{j,i}^{\mathcal{H}} + \sigma_j^2\mathbf{I}_{N_j}$$

is an affine map with respect to $\mathbf{Q}_k$. Since u_j is convex in $\mathbf{P}_j$ [87, Thm. 2], and convexity is preserved under affine mappings and nonnegative weighted-sums [26, Ch. 3], it follows that $f_k(\mathbf{Q}_k, \mathbf{Q}_{-k})$ is convex in $\mathbf{Q}_k$. □

Based on Lemma 5.2, convergence of the block coordinate ascent algorithm is established next.

Proposition 5.3 *The sequence of objective function values* (5.12a) *obtained by the coordinate ascent algorithm with concave-convex procedure converges.*

Proof: It suffices to show that the sequence of objective values (5.12a) is monotonically non-decreasing. Since the objective function value is bounded from above, the function value sequence must be convergent by invoking the monotone convergence theorem. Letting $\widetilde{\mathcal{U}}_k(\cdot)$ denote the objective function (5.15a), which is the concave surrogate of $\mathcal{U}(\cdot)$ as the original objective (5.12a), consider

$$\mathbf{Q}_k^{(n)} := \underset{\mathbf{Q}_k\in\mathcal{Q}_k}{\arg\max}\ \widetilde{\mathcal{U}}_k\left(\mathbf{Q}_k; \mathbf{Q}_1^{(n)}, \ldots, \mathbf{Q}_{k-1}^{(n)}, \mathbf{Q}_{k+1}^{(n-1)}, \ldots, \mathbf{Q}_K^{(n-1)}\right) \tag{5.30}$$

where n stands for the iteration index. Furthermore, define

$$\mathbf{Z}_k^{(n)} := (\mathbf{Q}_1^{(n+1)}, \ldots, \mathbf{Q}_k^{(n+1)}, \mathbf{Q}_{k+1}^{(n)}, \ldots, \mathbf{Q}_K^{(n)}), \tag{5.31}$$

$$\tilde{\mathbf{Q}}_{-k}^{(n)} := (\mathbf{Q}_1^{(n+1)}, \ldots, \mathbf{Q}_{k-1}^{(n+1)}, \mathbf{Q}_{k+1}^{(n)}, \ldots, \mathbf{Q}_K^{(n)}). \tag{5.32}$$

Then, for all $k \in \mathcal{K}$, it holds that

$$\begin{aligned}\mathcal{U}\left(\mathbf{Z}_k^{(n)}\right) &= u_k\left(\mathbf{Q}_k^{(n+1)}, \tilde{\mathbf{Q}}_{-k}^{(n)}\right) + f_k\left(\mathbf{Q}_k^{(n+1)}, \tilde{\mathbf{Q}}_{-k}^{(n)}\right) \\ &\geq u_k\left(\mathbf{Q}_k^{(n+1)}, \tilde{\mathbf{Q}}_{-k}^{(n)}\right) + f_k\left(\mathbf{Q}_k^{(n)}, \tilde{\mathbf{Q}}_{-k}^{(n)}\right)\end{aligned} \tag{5.33a}$$

$$+ \operatorname{Tr}\left\{\mathbf{D}_k^{\mathcal{H}}\left(\mathbf{Q}_k^{(n+1)} - \mathbf{Q}_k^{(n)}\right)\right\} \tag{5.33b}$$

$$\geq u_k\left(\mathbf{Q}_k^{(n)}, \tilde{\mathbf{Q}}_{-k}^{(n)}\right) + f_k\left(\mathbf{Q}_k^{(n)}, \tilde{\mathbf{Q}}_{-k}^{(n)}\right) + \operatorname{Tr}\left\{\mathbf{D}_k^{\mathcal{H}}\left(\mathbf{Q}_k^{(n)} - \mathbf{Q}_k^{(n)}\right)\right\} \tag{5.33c}$$

$$= \mathcal{U}\left(\mathbf{Z}_{k-1}^{(n)}\right) \tag{5.33d}$$

where (5.33b) follows from the convexity of $f_k(\cdot)$ established in Lemma 5.2; (5.33c) holds because $\mathbf{Q}_k^{(n+1)}$ is the optimal solution of (5.15) for fixed $\tilde{\mathbf{Q}}_{-k}^{(n)}$.

To complete the proof, it suffices to show that $\mathcal{U}\left(\mathbf{Q}^{(n+1)}\right)$ is monotonically non-decreasing, namely that

$$\mathcal{U}\left(\mathbf{Q}^{(n+1)}\right) \geq \mathcal{U}\left(\mathbf{Z}_{K-1}^{(n)}\right) \geq \ldots \geq \mathcal{U}\left(\mathbf{Z}_1^{(n)}\right) \geq \mathcal{U}\left(\mathbf{Q}^{(n)}\right) \tag{5.34}$$

□

Interestingly, by inspecting the structure of $\{\mathbf{H}_{k,k}, k \in \mathcal{K}\}$, it is also possible to show that every limit point generated by the coordinate ascent algorithm with local convex approximation satisfies the first-order optimality conditions. Conditions on $\{\mathbf{H}_{k,k}, k \in \mathcal{K}\}$ that guarantee stationarity of the limit points are provided next. First, it is useful to establish strict concavity of the objective (5.15a) in the following lemma.

Lemma 5.3 *If the channel matrices* $\{\mathbf{H}_{k,k}, k \in \mathcal{K}\}$ *of the CR links* $\{U_k^t \to U_k^r\}$ *have full column rank, then the objective function* (5.15a) *is strictly concave in* $\mathbf{Q}_k$.

Proof: First, notice that the objective function (5.15a) can be re-written as

$$\widetilde{\mathcal{U}}_k(\mathbf{Q}_k) = N_k + \operatorname{Tr}\left\{\mathbf{D}_k^{\mathcal{H}}\mathbf{Q}_k\right\} - \operatorname{Tr}\left\{\mathbf{R}_{k,k}\left(\mathbf{H}_{k,k}\mathbf{Q}_k\mathbf{H}_{k,k}^{\mathcal{H}} + \mathbf{R}_{k,k}\right)^{-1}\right\}. \tag{5.35}$$

Then, it suffices to prove strict convexity in $\mathbf{Q}_k$ of the third term on the right hand side of (5.35). This is equivalent to showing that (subscripts are dropped for brevity)

$$J(t) := \operatorname{Tr}\left\{\mathbf{R}\left(\mathbf{H}\mathbf{Q}\mathbf{H}^{\mathcal{H}} + \mathbf{R}\right)^{-1}\right\} \tag{5.36}$$

is strictly convex in $t \in \{t \mid \mathbf{Q} := \mathbf{X} + t\mathbf{Y} \in \mathcal{Q}\}$ for any given $\mathbf{X} \in \mathbb{H}_+^{n\times n}$ and nonzero $\mathbf{Y} \in \mathbb{H}^{n\times n}$.

To this end, consider the second-order derivative of $J(t)$, which is given by

$$\ddot{J}(t) = 2\operatorname{Tr}\{\mathbf{CRCLCL}\} \tag{5.37}$$

where $\mathbf{C} := \left(\mathbf{R} + \mathbf{H}(\mathbf{X} + t\mathbf{Y})\mathbf{H}^{\mathcal{H}}\right)^{-1}$ and $\mathbf{L} := \mathbf{H}\mathbf{Y}\mathbf{H}^{\mathcal{H}}$. Note that matrix $\mathbf{CRC}$ is Hermitian positive definite, since $\mathbf{C}$ and $\mathbf{R}$ are Hermitian positive definite too. With $\mathbf{H}$ full column rank, it readily follows that $\mathbf{L} \neq \mathbf{0}$ for any $\mathbf{Y} \neq \mathbf{0}$. This ensures that the Hermitian positive semi-definite matrix $\mathbf{LCL}$ is not an all-zero matrix, i.e., $\mathbf{LCL} \neq \mathbf{0}$.

Let $\nu_1 \geq \nu_2 \geq \cdots \geq \nu_N > 0$ and $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_N \geq 0$ denote the eigenvalues of matrices $\mathbf{CRC}$ and $\mathbf{LCL}$, respectively. Since matrix $\mathbf{LCL} \neq \mathbf{0}$, μ_1 is strictly positive, and thus

$$\ddot{J}(t) \geq 2\sum_{i=1}^{N} \nu_i \mu_{N-i+1} \tag{5.38a}$$

$$\geq 2\nu_N \mu_1 > 0 \tag{5.38b}$$

where (5.38a) follows from von Neumann's trace inequality [94]. Finally, (5.38b) shows the strong convexity (and hence strict convexity) of $J(t)$.

For completeness, we provide an alternative proof of the lemma. With some manipulations, function $h(\mathbf{Q}) := \mathrm{Tr}\left\{\mathbf{R}\left(\mathbf{H}\mathbf{Q}\mathbf{H}^{\mathcal{H}} + \mathbf{R}\right)^{-1}\right\}$ can be re-expressed as

$$\begin{aligned} h(\mathbf{Q}) &= g(\mathbf{R}^{-1/2}\mathbf{H}\mathbf{Q}\mathbf{H}^{\mathcal{H}}\mathbf{R}^{-1/2}) \\ &= \mathrm{Tr}\left\{\left(\mathbf{I} + \mathbf{R}^{-1/2}\mathbf{H}\mathbf{Q}\mathbf{H}^{\mathcal{H}}\mathbf{R}^{-1/2}\right)^{-1}\right\} \end{aligned} \tag{5.39}$$

where $g(\mathbf{X}) := \mathrm{Tr}\left\{(\mathbf{I} + \mathbf{X})^{-1}\right\}$. Let $\lambda_1(\mathbf{X}), \ldots, \lambda_n(\mathbf{X})$ denote again the eigenvalues of a matrix $\mathbf{X}$. Note that the spectral function $g(\mathbf{X}) = s(\lambda(\mathbf{X})) := \sum_i \left(\frac{1}{1+\lambda_i(\mathbf{X})}\right)$ is strictly convex if and only if the corresponding symmetric function $s(\cdot)$ is strictly convex [43]. To this end, the strict convexity of $\frac{1}{1+x}$ for $x \geq 0$ implies the strict convexity of $s(\cdot)$, and thus of $g(\mathbf{X})$. Under the condition of full column rank of $\mathbf{H}$, we will show that *strict* convexity is preserved under the linear mapping in (5.39). Specifically, define

$$\check{\mathbf{Q}}_i := \mathbf{R}^{-1/2}\mathbf{H}\mathbf{Q}_i\mathbf{H}^{\mathcal{H}}\mathbf{R}^{-1/2},\ i = 1, 2.$$

Then, for any $\mathbf{Q}_1 \neq \mathbf{Q}_2 \in \mathcal{Q}$ and $0 < \lambda < 1$, we have that

$$h(\lambda\mathbf{Q}_1 + (1-\lambda)\mathbf{Q}_2) = g(\lambda\check{\mathbf{Q}}_1 + (1-\lambda)\check{\mathbf{Q}}_2) \tag{5.40a}$$

$$< \lambda g(\check{\mathbf{Q}}_1) + (1-\lambda) g(\check{\mathbf{Q}}_2) \tag{5.40b}$$

$$= \lambda h(\mathbf{Q}_1) + (1-\lambda) h(\mathbf{Q}_2) \tag{5.40c}$$

where (5.40b) follows from the strict convexity of $g(\cdot)$, and the fact that $\check{\mathbf{Q}}_1 \neq \check{\mathbf{Q}}_2$ holds for any $\mathbf{Q}_1 \neq \mathbf{Q}_2$, since $\mathbf{H}$ is full column rank. □

We are now ready to establish stationarity of the limit points.

Theorem 5.1 *If matrices* $\{\mathbf{H}_{k,k}, k \in \mathcal{K}\}$ *have full column rank, then every limit point of the coordinate ascent algorithm with concave-convex procedure is a stationary point of* (5.12).

Proof: The proof of Theorem 5.1 relies on the basic convergence claim of the block coordinate descent method in [12, Ch. 2] and [116]. What must be shown is that every limit point of the algorithm satisfies the first-order optimality conditions over the Cartesian product of the closed convex sets. Let $\bar{\mathbf{Q}} := \left(\bar{\mathbf{Q}}_1, \ldots, \bar{\mathbf{Q}}_K\right)$ be a limit point of the sequence $\{\mathbf{Q}^{(n)}\}$, and $\{\mathbf{Q}^{(n_j)} | j = 1, 2, \ldots\}$ a subsequence that converges to $\bar{\mathbf{Q}}$. First, we will show that $\lim_{j\to\infty} \mathbf{Q}_1^{(n_j+1)} = \bar{\mathbf{Q}}_1$. Argue by contradiction, i.e., assume that $\{\mathbf{Q}_1^{(n_j+1)} - \mathbf{Q}_1^{(n_j)}\}$ does not converge to zero. Define $\gamma^{(n_j)} := \|\mathbf{Q}_1^{(n_j+1)} - \mathbf{Q}_1^{(n_j)}\|_F$. By possibly restricting to a subsequence of $\{n_j\}$, it follows that there exists some $\bar{\gamma} > 0$ such that $\bar{\gamma} \leq \gamma^{(n_j)}$ for all j. Let $\mathbf{S}_1^{(n_j)} := (\mathbf{Q}_1^{(n_j+1)} - \mathbf{Q}_1^{(n_j)})/\gamma^{(n_j)}$. Thus, we have that $\mathbf{Q}_1^{(n_j+1)} = \mathbf{Q}_1^{(n_j)} + \gamma^{(n_j)}\mathbf{S}_1^{(n_j)}$ and $\|\mathbf{S}_1^{(n_j)}\|_F = 1$. Because $\mathbf{S}_1^{(n_j)}$ belongs to a compact set, it can be assumed convergent to a limit point $\bar{\mathbf{S}}_1$ along with a subsequence of $\{n_j\}$.

Since it holds that $0 \leq \epsilon\bar{\gamma} \leq \gamma^{(n_j)}$ for all $\epsilon \in [0, 1]$, the point $\mathbf{Q}_1^{(n_j)} + \epsilon\bar{\gamma}\mathbf{S}_1^{(n_j)}$ lies on the segment connecting two feasible points $\mathbf{Q}_1^{(n_j)}$ and $\mathbf{Q}_1^{(n_j+1)}$. Thus, $\mathbf{Q}_1^{(n_j)} + \epsilon\bar{\gamma}\mathbf{S}_1^{(n_j)}$ is also feasible due to the convexity of $\mathcal{Q}_1$ [cf. Lemma 5.2]. Moreover, concavity of $\widetilde{\mathcal{U}}_1(\cdot; \mathbf{Q}_{-1}^{(n_j)})$ implies that $\widetilde{\mathcal{U}}_1$ is monotonically non-decreasing in the interval connecting point $\mathbf{Q}_1^{(n_j)}$ to $\mathbf{Q}_1^{(n_j+1)}$ over the set $\mathcal{Q}_1$. Hence, it readily follows that

$$\widetilde{\mathcal{U}}_1(\mathbf{Q}_1^{(n_j+1)}; \mathbf{Q}_{-1}^{(n_j)}) \geq \widetilde{\mathcal{U}}_1(\mathbf{Q}_1^{(n_j)} + \epsilon\bar{\gamma}\mathbf{S}_1^{(n_j)}; \mathbf{Q}_{-1}^{(n_j)}) \geq \widetilde{\mathcal{U}}_1(\mathbf{Q}_1^{(n_j)}; \mathbf{Q}_{-1}^{(n_j)}). \tag{5.41}$$

Note that $\widetilde{\mathcal{U}}_1(\cdot)$ is a tight lower bound of $\mathcal{U}(\cdot)$ at each current feasible point. Also, from (5.34), $\widetilde{\mathcal{U}}_1(\mathbf{Q}_1^{(n_j+1)}; \mathbf{Q}_{-1}^{(n_j)})$ is guaranteed to converge to $\widetilde{\mathcal{U}}_1(\bar{\mathbf{Q}})$ as $j \to \infty$. Thus, upon taking the limit as $j \to \infty$ in (5.41), it follows that

$$\widetilde{\mathcal{U}}_1(\bar{\mathbf{Q}}_1 + \epsilon\bar{\gamma}\bar{\mathbf{S}}_1; \bar{\mathbf{Q}}_{-1}) = \widetilde{\mathcal{U}}_1(\bar{\mathbf{Q}}), \quad \forall\, \epsilon \in [0, 1]\,. \tag{5.42}$$

However, since $\bar{\gamma}\bar{\mathbf{S}}_1 \neq \mathbf{0}$, (5.42) contradicts the unique maximum condition implied by the strict concavity of $\widetilde{\mathcal{U}}_1(\cdot;\cdot)$ in $\mathbf{Q}_1$ [cf. Lemma 5.3]. Therefore, $\mathbf{Q}_1^{(n_j+1)}$ converges to $\bar{\mathbf{Q}}_1$ as well.

Consider now checking the optimality condition for $\bar{\mathbf{Q}}_1$. Since $\mathbf{Q}_1^{(n_j+1)}$ is the local (and also global) maximum of $\widetilde{\mathcal{U}}_1(\cdot;\mathbf{Q}_{-1}^{(n_j)})$, we have that

$$\Re\left\{\mathrm{Tr}\left\{\nabla_1\widetilde{\mathcal{U}}_1\left(\mathbf{Q}_1^{(n_j+1)};\mathbf{Q}_{-1}^{(n_j)}\right)^{\mathcal{H}}\left(\mathbf{Q}_1-\mathbf{Q}_1^{(n_j+1)}\right)\right\}\right\}\leq 0,\ \forall\ \mathbf{Q}_1\in\mathcal{Q}_1 \tag{5.43}$$

where $\nabla_1\widetilde{\mathcal{U}}_1(\cdot)$ denotes the gradient of $\widetilde{\mathcal{U}}_1(\cdot)$ with respect to $\mathbf{Q}_1$. Taking the limit as $j\to\infty$, and using the fact that $\nabla_1\widetilde{\mathcal{U}}_1(\bar{\mathbf{Q}})=\nabla_1\mathcal{U}_1(\bar{\mathbf{Q}})$, it is easy to show that

$$\Re\left\{\mathrm{Tr}\left\{\nabla_1\mathcal{U}(\bar{\mathbf{Q}})^{\mathcal{H}}(\mathbf{Q}_1-\bar{\mathbf{Q}}_1)\right\}\right\}\leq 0,\quad \forall\ \mathbf{Q}_1\in\mathcal{Q}_1. \tag{5.44}$$

Using similar arguments, it holds that

$$\Re\left\{\mathrm{Tr}\left\{\nabla_i\mathcal{U}(\bar{\mathbf{Q}})^{\mathcal{H}}(\mathbf{Q}_i-\bar{\mathbf{Q}}_i)\right\}\right\}\leq 0,\ \forall\ \mathbf{Q}_i\in\mathcal{Q}_i,\ i=1,2,\ldots,K \tag{5.45}$$

which establishes the stationarity of $\bar{\mathbf{Q}}$ and completes the proof. □

5.2.3 Proximal Point-based Robust Algorithm

The full column rank requirement can be quite restrictive in practice; e.g., if $M_k > N_k$ for at least one CR link, or in the presence of spatially correlated MIMO channels [77]. Furthermore, computing the rank of channel matrices increases the computational burden to an extent that may not be affordable by the CRs. In this section, an alternative approach based on proximal-point regularization [118] is pursued to ensure convergence, without requiring restrictions on the antenna configuration and the channel rank.

The idea consists in penalizing the objective of (5.15) using a quadratic regularization term $\frac{1}{2\tau_k}\|\mathbf{Q}_k-\mathbf{Q}_k^{(n-1)}\|_F^2$, with a given sequence of numbers $\tau_k>0$. Then, (5.28) is modified as

$$\underset{\substack{\mathbf{Q}_k\succeq\mathbf{0},\theta_k\geq 0\\ \mathbf{T},\mathbf{Y}}}{\text{minimize}}\quad \mathrm{Tr}\left\{\mathbf{T}\right\}-\mathrm{Tr}\left\{\mathbf{D}_k^{\mathcal{H}}\mathbf{Q}_k\right\}+\frac{1}{2\tau_k}\mathrm{Tr}\{\mathbf{Y}\} \tag{5.46a}$$

$$\text{subject to}\quad \begin{bmatrix}\mathbf{I}_{M_k} & \mathbf{Q}_k-\mathbf{Q}_k^{(n-1)}\\ \mathbf{Q}_k-\mathbf{Q}_k^{(n-1)} & \mathbf{Y}\end{bmatrix}\succeq\mathbf{0} \tag{5.46b}$$

$$(5.28\text{b}),(5.28\text{c}),(5.28\text{d})$$

where (5.46b) is derived by using the Schur complement through the variable $\mathbf{Y}$.

The role of $\frac{1}{2\tau_k}\|\mathbf{Q}_k - \mathbf{Q}_k^{(n-1)}\|_F^2$ is to render the cost in (5.46a) strictly convex and coercive. Moreover, for small values of τ_k, the optimization variable $\mathbf{Q}_k$ is forced to stay "close" to $\mathbf{Q}_k^{(n-1)}$ obtained at the previous iteration, thereby improving the stability of the iterates [13, Ch. 6]. Centralized and distributed schemes with the proximal point regularization are given by Algorithms 5.1 and 5.2, respectively, with problem (5.46) replacing (5.28). Convergence of the resulting schemes is established in the following theorem. To avoid ambiguity, these proximal point-based algorithms will be hereafter referred as Algorithms 5.1(P) and 5.2(P), respectively.

Theorem 5.2 *Suppose that the sequence* $\{\mathbf{Q}^{(n)}\}$ *generated by Algorithm 5.1(P) (Algorithm 5.2(P)) has a limit point. Then, every limit point is a stationary point of* (5.12).

Proof: The Gauss-Seidel method with a proximal point regularization converges without any underlying convexity assumptions [61]. A modified version of the proof is reported here, where the local convex approximation (5.13) and the peculiarities of the problem at hand are leveraged to establish not only convergence of the algorithm, but also optimality of the obtained solution.

Assume there exists a subsequence $\{\mathbf{Q}^{(n_j)}|j = 1, 2, \ldots\}$ converging to a limit point $\bar{\mathbf{Q}} := \left(\bar{\mathbf{Q}}_1, \ldots, \bar{\mathbf{Q}}_K\right)$. Let $\mathbf{Q}_k^{(n+1)}$ be obtained as

$$\mathbf{Q}_k^{(n+1)} := \underset{\mathbf{Q}_k \in \mathcal{Q}_k}{\arg\max}\ \widetilde{\mathcal{U}}_k\left(\mathbf{Q}_k; \mathbf{Q}_1^{(n+1)}, \ldots, \mathbf{Q}_{k-1}^{(n+1)}, \mathbf{Q}_{k+1}^{(n)}, \ldots, \mathbf{Q}_K^{(n)}\right) - \frac{1}{2\tau_k}\|\mathbf{Q}_k - \mathbf{Q}_k^{(n)}\|_F^2 . \tag{5.47}$$

Thus, it follows that [cf. (5.32)]

$$\widetilde{\mathcal{U}}_1(\mathbf{Q}_1^{(n_j+1)}; \tilde{\mathbf{Q}}_{-1}^{(n_j)}) \geq \widetilde{\mathcal{U}}_1(\mathbf{Q}_1^{(n_j)}; \tilde{\mathbf{Q}}_{-1}^{(n_j)}) + \frac{1}{2\tau_k}\|\mathbf{Q}_1^{(n_j+1)} - \mathbf{Q}_1^{(n_j)}\|_F^2 . \tag{5.48}$$

Going along the lines of the proof of Theorem 5.1, it holds that

$$\lim_{j\to\infty} \widetilde{\mathcal{U}}_1(\mathbf{Q}_1^{(n_j+1)}; \tilde{\mathbf{Q}}_{-1}^{(n_j)}) = \lim_{j\to\infty} \widetilde{\mathcal{U}}_1(\mathbf{Q}_1^{(n_j)}; \tilde{\mathbf{Q}}_{-1}^{(n_j)}) = \widetilde{\mathcal{U}}_1(\bar{\mathbf{Q}}). \tag{5.49}$$

Therefore, taking the limit as $j \to \infty$ in (5.48), one arrives at

$$\lim_{j\to\infty} \|\mathbf{Q}_1^{(n_j+1)} - \mathbf{Q}_1^{(n_j)}\|_F^2 = 0 \tag{5.50}$$

which implies that $\mathbf{Q}_1^{(n_j+1)}$ also converges to $\bar{\mathbf{Q}}_1$.

Since $\mathbf{Q}_1^{(n_j+1)}$ is generated as in (5.47), it satisfies the optimality condition

$$\Re\left\{\mathrm{Tr}\left\{\left[\nabla_1\widetilde{\mathcal{U}}_1(\mathbf{Q}_1^{(n_j+1)};\mathbf{Q}_{-1}^{(n_j)})-\frac{1}{\tau_1}(\mathbf{Q}_1^{(n_j+1)}-\mathbf{Q}_1^{(n_j)})\right]^{\mathcal{H}}(\mathbf{Q}_1-\mathbf{Q}_1^{(n_j+1)})\right\}\right\}\leq 0,$$
$$\forall\ \mathbf{Q}_1\in\mathcal{Q}_1. \tag{5.51}$$

Taking the limit as $j\to\infty$ in (5.51), and using again the fact that $\nabla_1\widetilde{\mathcal{U}}_1(\bar{\mathbf{Q}})=\nabla_1\mathcal{U}_1(\bar{\mathbf{Q}})$, we obtain

$$\Re\left\{\mathrm{Tr}\left\{\nabla_1\mathcal{U}(\bar{\mathbf{Q}})^{\mathcal{H}}(\mathbf{Q}_1-\bar{\mathbf{Q}}_1)\right\}\right\}\leq 0,\quad \forall\ \mathbf{Q}_1\in\mathcal{Q}_1\,. \tag{5.52}$$

Then, repeating the same argument for all $k\in\mathcal{K}$, leads to

$$\Re\left\{\mathrm{Tr}\left\{\nabla_k\mathcal{U}(\bar{\mathbf{Q}})^{\mathcal{H}}(\mathbf{Q}_k-\bar{\mathbf{Q}}_k)\right\}\right\}\leq 0,\quad \forall\ \mathbf{Q}_k\in\mathcal{Q}_k \tag{5.53}$$

which shows that the limit point $\bar{\mathbf{Q}}$ is also a stationary point. □

As asserted in Theorem 5.2, Algorithms 5.1(P) and 5.2(P) converge to a stationary point of (5.12) for any possible antenna configuration. The price to pay however, is a possibly slower convergence rate that is common to proximal point-based methods [13, Ch. 6] (see also the numerical tests in Section 5.5.1). For this reason, the proximal point-based method should be used in either a centralized or a distributed setup whenever the number of transmit-antennas exceeds that of receive-antennas in at least one transmitter-receiver pair. In this case, Algorithms 5.1(P) and 5.2(P) ensure first-order optimality of the solution obtained. When $M_k\leq N_k$, for all $k\in\mathcal{K}$, the two solvers have complementary strengths in convergence rate and computational complexity. Specifically, Algorithms 5.1 and 5.2 require the rank of all CR direct channel matrices $\{\mathbf{H}_{k,k}\}$ beforehand, which can be computationally burdensome, especially for a high number of antenna elements. If the rank determination can be afforded, and the convergence rate is at a premium, then Algorithms 5.1 and 5.2 should be utilized.

5.2.4 Aggregate Interference Constraints

Suppose now that the individual interference budgets $\{\iota_k^{\max}\}$ are not available a priori. Then, the aggregate interference power $\{\iota^{\max}\}$ has to be divided among transmit-CRs by the resource allocation scheme in order for the overall system performance to be

optimized. Accordingly, (5.12) is modified as follows to incorporate a robust constraint on the total interference power inflicted to the PU node:

$$\underset{\{\mathbf{Q}_k\}_{k=1}^K}{\text{maximize}} \quad \sum_{k=1}^{K} u_k(\{\mathbf{Q}_k\}) \tag{5.54a}$$

$$\text{subject to} \quad \text{Tr}\{\mathbf{Q}_k\} \leq p_k^{\max},\ k \in \mathcal{K} \tag{5.54b}$$

$$\sum_{k=1}^{K} \text{Tr}\{\mathbf{G}_k \mathbf{Q}_k \mathbf{G}_k^{\mathcal{H}}\} \leq \iota^{\max},\ \forall\ \Delta\mathbf{G}_k \in \mathcal{G}_k,\ k \in \mathcal{K}. \tag{5.54c}$$

The new interference constraint (5.54c) couples the CR nodes (or, more precisely, the subset of transmit-CR nodes in the proximity of the PU receiver). Thus, the overhead of message passing increases since cooperation among coupled CR nodes is needed.

A common technique for dealing with coupled constraints is the dual decomposition method [12], which facilitates evaluation of the dual function by dualizing the coupled constraints. However, since problem (5.54) is non-convex and non-separable, the duality gap is generally non-zero. Thus, the primal variables obtained during the intermediate iterates may not be feasible, i.e., transmit-covariances can possibly lead to violation of the interference constraint. Since the ultimate goal is to design an on-line algorithm where (5.54c) must be satisfied during network operation, the primal decomposition technique is well motivated to cope with the coupled interference constraints [79]. To this end, consider introducing two sets of auxiliary variables $\{\iota_k\}$ and $\{t_k\}$ in problem (5.54), which is equivalently re-formulated as

$$\underset{\{\mathbf{Q}_k, \iota_k, t_k\}}{\text{maximize}} \quad \sum_{k=1}^{K} \text{Tr}\left\{\mathbf{H}_{k,k}\mathbf{Q}_k\mathbf{H}_{k,k}^{\mathcal{H}}\left(\mathbf{H}_{k,k}\mathbf{Q}_k\mathbf{H}_{k,k}^{\mathcal{H}} + \mathbf{R}_{k,k}\right)^{-1}\right\} \tag{5.55a}$$

$$\text{subject to} \quad \text{Tr}\{\mathbf{Q}_k\} \leq p_k^{\max}, \quad k \in \mathcal{K} \tag{5.55b}$$

$$\text{Tr}\{\mathbf{G}_k \mathbf{Q}_k \mathbf{G}_k^{\mathcal{H}}\} \leq t_k,\ \forall\ \Delta\mathbf{G}_k \in \mathcal{G}_k, k \in \mathcal{K} \tag{5.55c}$$

$$0 \leq t_k \leq \iota_k, \quad k \in \mathcal{K} \tag{5.55d}$$

$$\sum_{k=1}^{K} \iota_k \leq \iota^{\max}. \tag{5.55e}$$

For fixed $\{\iota_k\}$, the inner maximization subproblem turns out to be

$$p(\{\iota_k\}) := \underset{\{\mathbf{Q}_k, t_k\}}{\text{maximize}} \quad \sum_{k=1}^{K} \text{Tr}\left\{\mathbf{H}_{k,k}\mathbf{Q}_k\mathbf{H}_{k,k}^{\mathcal{H}}\left(\mathbf{H}_{k,k}\mathbf{Q}_k\mathbf{H}_{k,k}^{\mathcal{H}} + \mathbf{R}_{k,k}\right)^{-1}\right\} \tag{5.56a}$$

$$\text{subject to} \quad \mathrm{Tr}\{\mathbf{Q}_k\} \le p_k^{\max}, \quad k \in \mathcal{K} \tag{5.56b}$$

$$\mathrm{Tr}\{\mathbf{G}_k \mathbf{Q}_k \mathbf{G}_k^{\mathcal{H}}\} \le t_k, \ \forall\ \Delta\mathbf{G}_k \in \mathcal{G}_k, k \in \mathcal{K} \tag{5.56c}$$

$$0 \le t_k \le \iota_k, \quad k \in \mathcal{K} \tag{5.56d}$$

which, as discussed in preceding sections, can be solved using the block coordinate ascent algorithm (or its proximal point version) in either a centralized or a distributed fashion. After solving (5.56) for a given set $\{\iota_k\}$, the per-CR interference budgets $\{\iota_k\}$ are updated by the following master problem:

$$\underset{\{\iota_k\}}{\text{maximize}} \quad p(\{\iota_k\}) \tag{5.57a}$$

$$\text{subject to} \quad \{\iota_k\} \in \mathcal{I} \tag{5.57b}$$

with the simplex set $\mathcal{I}$ given by

$$\mathcal{I} := \left\{ \{\iota_k\} \mid \iota_k \ge 0, \sum_{k=1}^{K} \iota_k \le \iota^{\max} \right\}. \tag{5.58}$$

Overall, the primal decomposition method solves (5.55) by iteratively solving (5.56) and (5.57). Notice that the master problem (5.57) dynamically divides the total interference budget $\iota^{\max}$ among CR transmitters, so as to find the best allocation of resources that maximizes the overall system performance. Using the block coordinate ascent algorithm, the k-th transmit-covariance matrix $\mathbf{Q}_k$ is obtained by solving the following problem [cf. Algorithm 5.2]

$$\tilde{p}_k(\iota_k) := \underset{\mathbf{Q}_k \succeq \mathbf{0}, t_k \ge 0}{\text{maximize}} \quad u_k\left(\mathbf{Q}_k, \mathbf{Q}_{-k}\right) + \mathrm{Tr}\left\{\mathbf{D}_k^{\mathcal{H}} \mathbf{Q}_k\right\} \tag{5.59a}$$

$$\text{subject to} \quad \mathrm{Tr}\{\mathbf{Q}_k\} \le p_k^{\max} \tag{5.59b}$$

$$\mathrm{Tr}\{\mathbf{G}_k \mathbf{Q}_k \mathbf{G}_k^{\mathcal{H}}\} \le t_k, \forall\ \Delta\mathbf{G}_k \in \mathcal{G}_k \tag{5.59c}$$

$$t_k \le \iota_k \tag{5.59d}$$

where the proximal point-based regularization term is added if Algorithm 5.2(P) is implemented. Since (5.59) is a convex problem, it can be seen that the subgradient of $\tilde{p}_k(\iota_k)$ with respect to ι_k is the optimal Lagrange multiplier λ_k corresponding to the constraint (5.59d) [12, Ch. 5]. Thus, it becomes possible to utilize the subgradient

Algorithm 5.3 Distributed on-line robust sum-MSE minimization with aggregate interference constraint

1: Initialize $\mathbf{Q}_k^{(0)}(0) = \mathbf{0}$, and $\iota_k(0) = \iota^{\max}/K$, $\forall\ k \in \mathcal{K}$
2: **repeat** $(\ell = 1, 2, \ldots)$
3: [CRs]: Solve (5.56) via Algorithm 5.2 or 5.2(P)
4: [CRs]: Transmit $\{\lambda_k(\ell)\}$ to the cluster-head node
5: [Cluster-head node]: Update $\{\iota_k(\ell+1)\}$ via (5.60)
6: [CRs]: Receive $\{\iota_k(\ell+1)\}$ from the cluster-head node
7: **until** $u_k\left(\mathbf{Q}^{(n)}(\ell)\right) - u_k\left(\mathbf{Q}^{(n')}(\ell-1)\right) < \upsilon,\ \forall\ k \in \mathcal{K}$

projection method to solve the master problem. Strictly speaking, due to the non-convexity of the original objective (5.55a), primal decomposition method leveraging the subgradient algorithm is not an exact, but rather an approximate (and simple) approach to solve (5.55). However, because (5.59a) is a *tight* concave lower bound of (5.55a) around the approximating feasible point, $p(\{\iota_k\})$ is well-approximated by $\tilde{p}_k(\iota_k)$ as $\{\mathbf{Q}_k^{(n)}\}$ approaches the optimal value $\{\mathbf{Q}_k^{\text{opt}}\}$. Hence, λ_k also comes "very close" to the true subgradient of $p_k(\{\iota_k\})$ with respect to ι_k. Therefore, at iteration ℓ of the primal decomposition method, the subgradient projection updating the interference budgets $\boldsymbol{\iota} := [\iota_1, \iota_2, \ldots, \iota_K]'$ becomes

$$\boldsymbol{\iota}(\ell+1) = \text{Proj}_{\mathcal{I}}\left[\boldsymbol{\iota}(\ell) + s(\ell)\boldsymbol{\lambda}(\ell)\right] \tag{5.60}$$

where $\boldsymbol{\lambda} := [\lambda_1, \lambda_2, \ldots, \lambda_K]'$; $s(\ell)$ is a positive step size; $\text{Proj}_{\mathcal{I}}[\cdot]$ denotes projection onto the convex feasible set $\mathcal{I}$. Projection onto the simplex set in (5.58) is a computationally-affordable operation that can be efficiently implemented as in e.g., [109].

Once (5.56) is solved distributedly, each CR that is coupled by the interference constraint has to transmit the local scalar Lagrange multiplier $\lambda_k(\ell)$ to a cluster-head CR node. This node, in turn, will update $\{\iota_k(\ell+1)\}$ and will feed these quantities back to the CRs. The resulting on-line distributed scheme is tabulated as Algorithm 5.3. Notice however that in order for the overall algorithm to adapt to possibly slowly varying channels, operation (5.60) can be computed at the end of each cycle of the block coordinate ascent algorithm, rather than wait for its convergence.

5.3 Smart-Grid Powered Cooperative Multipoint Systems

Consider a smart-grid powered cluster-based CoMP downlink. A set $\mathcal{I} := \{1, \ldots, I\}$ of distributed BSs (e.g., macro/micro/pico BSs) provides service to a set $\mathcal{K} := \{1, \ldots, K\}$ of mobile users; see Fig. 5.2. Each BS has $M \geq 1$ transmit antennas and each user has a single receive antenna. Each BS is equipped with one or more energy harvesting devices (solar panels and/or wind turbines), and is also connected to the power grid with a two-way energy trading facility. Different from [27, 151, 152], each BS has an energy storage device (possibly consisting of several large-capacity batteries) so that it does not have to consume or sell all the harvested energy on the spot, but can save it for later use.

For each CoMP cluster, there is a low-latency backhaul network connecting the set of BSs to a central controller [156], which coordinates energy trading as well as cooperative communication. The central entity can collect both the communication data (transmit messages) from each of the BSs through the cellular backhaul links, as well as the energy information (energy buying/selling prices) from these BSs via smart meters installed at BSs, and the grid-deployed communication/control links connecting them.

Assume slot-based transmissions from the BSs to the users. While the actual harvested energy amounts and wireless channels cannot be accurately predicted, uncertainty regions for the wireless channels and renewable energy arrivals can be obtained, based on historical measurements and/or forecasting techniques. The slot duration is selected equal to the minimum time interval between the changes of the (channel or energy) uncertainty regions. Consider a finite scheduling horizon consisting of T slots, indexed by the set $\mathcal{T} := \{1, \ldots, T\}$. For convenience, the slot duration is normalized to unity unless otherwise specified; thus, the terms "energy" and "power" will hereafter be used interchangeably throughout the section.

5.3.1 Downlink CoMP Transmission Model

Per slot t, let $\mathbf{h}_{ik}^t \in \mathbb{C}^M$ denote the vector channel from BS i to user k, $\forall\ i \in \mathcal{I},\ k \in \mathcal{K}$; and let $\mathbf{h}_k^t := [\mathbf{h}_{1k}^t{}', \ldots, \mathbf{h}_{Ik}^t{}']'$ collect the channel vectors from all BSs to user k. With linear transmit beamforming performed across BSs, the vector signal transmitted to

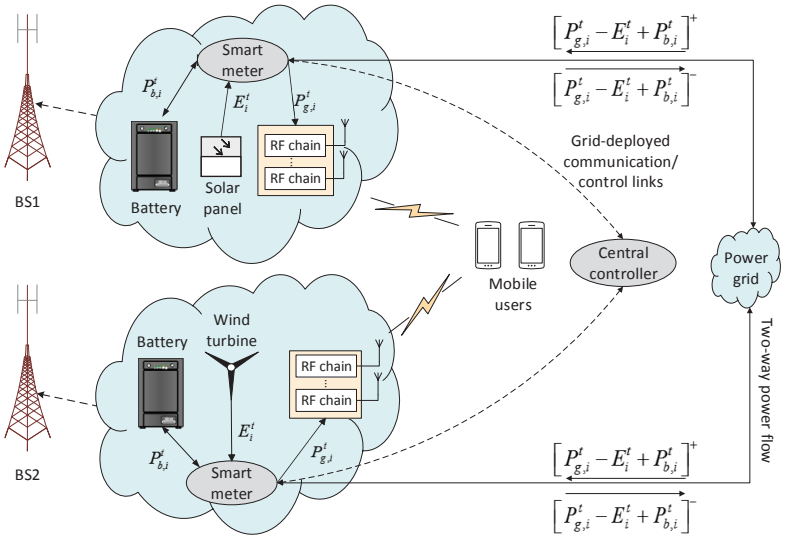


Figure 5.2: A two-BS CoMP system powered by smart grids, where BSs with local renewable energy harvesting and storage devices implement two-way energy trading with the main grid.

user k is

$$\mathbf{q}_k^t = \mathbf{w}_k^t s_k^t, \quad \forall\ k$$

where s_k^t denotes the information-bearing symbol, and $\mathbf{w}_k^t \in \mathbb{C}^{MI}$ denotes the beamforming vector across the BSs for user k. For convenience, s_k^t is assumed to be a complex random variable with zero mean and unit variance. The received vector at user k is therefore

$$y_k^t = \mathbf{h}_k^{t\,\mathcal{H}} \mathbf{q}_k^t + \sum_{l \neq k} \mathbf{h}_k^{t\,\mathcal{H}} \mathbf{q}_l^t + n_k^t \tag{5.61}$$

where $\mathbf{h}_k^{t\,\mathcal{H}} \mathbf{q}_k^t$ is the desired signal for user k, $\sum_{l \neq k} \mathbf{h}_k^{t\,\mathcal{H}} \mathbf{q}_l^t$ is the inter-user interference from the same cluster, and n_k^t denotes additive noise, which may also include the downlink interference from other BSs outside this cluster. It is assumed that n_k^t is a circularly symmetric complex Gaussian random variable with zero mean and variance σ_k^2.

The SINR at user k can be expressed as

$$\mathrm{SINR}_k(\{\mathbf{w}_k^t\}) = \frac{|\mathbf{h}_k^{t\,\mathcal{H}} \mathbf{w}_k^t|^2}{\sum_{l \neq k} (|\mathbf{h}_k^{t\,\mathcal{H}} \mathbf{w}_l^t|^2) + \sigma_k^2}\ . \tag{5.62}$$

The transmit power at each BS i clearly is given by $P_{x,i}^t = \sum_{k\in\mathcal{K}} \mathbf{w}_k^{t\,\mathcal{H}} \mathbf{B}_i \mathbf{w}_k^t$, where the matrix $\mathbf{B}_i \in \mathbb{R}^{MI\times MI}$ is defined as

$$\mathbf{B}_i := \text{diag}\left(\underbrace{0,\ldots,0}_{(i-1)M}, \underbrace{1,\ldots,1}_{M}, \underbrace{0,\ldots,0}_{(I-i)M}\right)$$

that selects the corresponding rows out of $\{\mathbf{w}_k^t\}_{k\in\mathcal{K}}$ to form the i-th BS's transmit-beamforming vector.

The channel state information $\mathbf{h}_k^t$ is seldom precisely available *a priori* in practice. Relying on past channel measurements and/or reliable channel predictions, we adopt the following additive error model: $\mathbf{h}_k^t = \hat{\mathbf{h}}_k^t + \boldsymbol{\delta}_k^t$, where $\hat{\mathbf{h}}_k^t$ is the predicted channel. The uncertainty of this estimate is bounded by a spherical region [140]:

$$\mathcal{H}_k^t := \left\{\hat{\mathbf{h}}_k^t + \boldsymbol{\delta}_k^t \mid \left\|\boldsymbol{\delta}_k^t\right\| \le \epsilon_k^t\right\}, \quad \forall\, k,t \tag{5.63}$$

where $\epsilon_k^t > 0$ specifies the radius of $\mathcal{H}_k^t$. This leads to the worst-case SINR per user k as [cf. (5.62)]

$$\widetilde{\text{SINR}}_k(\{\mathbf{w}_k^t\}) := \min_{\mathbf{h}_k^t\in\mathcal{H}_k^t} \frac{|\mathbf{h}_k^{t\,\mathcal{H}}\mathbf{w}_k^t|^2}{\sum_{l\neq k}(|\mathbf{h}_k^{t\,\mathcal{H}}\mathbf{w}_l^t|^2)+\sigma_k^2}\ . \tag{5.64}$$

To guarantee QoS per slot, it is required that

$$\widetilde{\text{SINR}}_k(\{\mathbf{w}_k^t\}) \ge \gamma_k, \quad \forall\, k \tag{5.65}$$

where γ_k denotes the target SINR value per user k.

5.3.2 Energy Storage Model

Let C_i^0 denote the initial energy, and C_i^t the amount of stored energy in the batteries of the i-th BS at the beginning of time slot t. With $C_i^{\max}$ bounding the capacity of batteries, it is clear that $0 \le C_i^t \le C_i^{\max}$, $\forall\, i \in \mathcal{I}$. Let $P_{b,i}^t$ denote the power delivered to or drawn from the batteries at slot t, which amounts to either charging ($P_{b,i}^t > 0$) or discharging ($P_{b,i}^t < 0$). Hence, the stored energy obeys the dynamic equation

$$C_i^t = C_i^{t-1} + P_{b,i}^t, \quad t \in \mathcal{T},\ i \in \mathcal{I}. \tag{5.66}$$

The amount of power (dis-)charged is also bounded by

$$\max -\varpi_i C_i^{t-1}, \quad P_{b,i}^{\min} \le P_{b,i}^t \le P_{b,i}^{\max} \tag{5.67}$$

where $P_{b,i}^{\min} < 0$ and $P_{b,i}^{\max} > 0$, while $\varpi_i \in (0,1]$ is the battery efficiency at BS i. The constraint $-\varpi_i C_i^{t-1} \leq P_{b,i}^t$ means that at most a fraction ϖ_i of the stored energy C_i^{t-1} is available for discharge.

5.3.3 Energy Harvesting and Transaction Cost Models

Let E_i^t denote the energy harvested during the last slot that is available at the beginning of slot t at each BS $i \in \mathcal{I}$; and let $\mathbf{e}_i := [E_i^1, \ldots, E_i^T]'$. Due to the unpredictable and intermittent nature of RES, $\mathbf{e}_i$ is unknown a priori. In general, uncertain quantities can be modeled by postulating either an underlying probability distribution or an uncertainty region. Probability distributions (possibly mixed discrete/continuous) of the RES generation are seldom available in practice. Although (non-)parametric approaches can be used to learn the distributions, the processes can be very complicated due to the spatio-temporal correlations incurred by various meteorological factors. On the other hand, the approach of postulating an uncertainty region provides the decision maker range forecasts instead of point forecasts, which are essentially distribution-free. Suppose that $\mathbf{e}_i$ lies in an uncertainty set $\mathcal{E}_i$, which can be obtained from historical measurements and/or fine forecast techniques. From the perspective of computational tractability, two practical paradigms for $\mathcal{E}_i$ are considered here.

i) The first model amounts to a polyhedral set [165]:

$$\mathcal{E}_i^{\mathrm{p}} := \left\{ \mathbf{e}_i \mid \underline{E}_i^t \leq E_i^t \leq \overline{E}_i^t, E_{i,s}^{\min} \leq \sum_{t \in \mathcal{T}_{i,s}} E_i^t \leq E_{i,s}^{\max},\ \mathcal{T} = \bigcup_{s=1}^{S} \mathcal{T}_{i,s} \right\} \tag{5.68}$$

where $\underline{E}_i^t$ ($\overline{E}_i^t$) denotes a lower (upper) bound on E_i^t; time horizon $\mathcal{T}$ is partitioned into consecutive but non-overlapping sub-horizons $\mathcal{T}_{i,s}$, $s = 1, \ldots, S$. Each sub-horizon can consist of multiple time slots, and the total energy harvested by BS i over the sth sub-horizon is bounded by $E_{i,s}^{\min}$ and $E_{i,s}^{\max}$.

ii) The second model relies on an ellipsoidal uncertainty set (see e.g. [34])

$$\mathcal{E}_i^{\mathrm{e}} := \left\{ \mathbf{e}_i = \hat{\mathbf{e}}_i + \boldsymbol{\varsigma}_i \mid \boldsymbol{\varsigma}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{\varsigma}_i \leq 1 \right\} \tag{5.69}$$

where $\hat{\mathbf{e}}_i := [\hat{E}_i^1, \ldots, \hat{E}_i^T]'$ denotes the nominal energy harvested at BS i, which

can be the forecasted energy, or simply its expected value. Vector $\boldsymbol{\varsigma}_i$ is the corresponding error in forecasting. The known matrix $\boldsymbol{\Sigma} \succ \mathbf{0}$ quantifies the shape of the ellipsoid $\mathcal{E}_i^{\mathrm{e}}$, and hence determines the quality of the forecast.

Remark 5.7 *(Spatio-temporally correlated energy harvesting models).* The aforementioned two practical models capture RES uncertainty across the scheduling (sub-)horizons per BS. The parameters required for constructing the sets $\{\mathcal{E}_i^{\mathrm{p}}, \mathcal{E}_i^{\mathrm{e}}\}$ can be obtained offline via statistical learning techniques using historical data. In general, green energy harvested at different BSs can be spatially correlated if some BSs are geographically close. In this case, joint spatio-temporal uncertainty models can be postulated whenever the underlying correlations are known *a priori*; see details in [165]. In general, a refined uncertainty model quantifying the actual harvested energy in a smaller region with a higher confidence level can be less conservative in the robust optimization formulation. However, the complexity of solving the resulting optimization problems directly depends on the choice of the uncertainty set.

Note that the coherence time of RES arrivals can be much longer than that of wireless channels in practice [150, 151]. Yet, coherence times corresponding to the uncertainty regions of wireless channels can be much larger than those of the channel itself. In Section II-A, we implicitly assume that channel $\mathbf{h}_k^t$ remains unchanged per slot. However, $\widetilde{\mathrm{SINR}}_k$ in (5.64) can be easily redefined as the worst-case SINR over multiple channel coherence times with the same uncertainty region (and possibly different channel realizations). This way the issue of different time scales becomes less critical. In addition, the aforementioned models in fact accommodate cases where uncertainty regions for RES arrivals remain unchanged over multiple time slots. The proposed approach presented next readily applies to obtain the robust ahead-of-time schedule in this setup.

For the i-th BS per slot t, the total energy consumption $P_{g,i}^t$ includes the transmission-related power $P_{x,i}^t$, and the rest that is due to other components such as air conditioning, data processor, and circuits, which can be generally modeled as a constant power, $P_{c,i} > 0$ [126, 152]; namely,

$$P_{g,i}^t = P_{c,i} + P_{x,i}^t/\xi = P_{c,i} + \sum_{k\in\mathcal{K}} \mathbf{w}_k^{t\,\mathcal{H}} \mathbf{B}_i \mathbf{w}_k^t/\xi$$

where $\xi > 0$ denotes the power amplifier efficiency. For notational convenience, we absorb ξ into $\mathbf{B}_i$ by redefining $\mathbf{B}_i := \mathbf{B}_i/\xi$; and further assume that $P_{g,i}^t$ is bounded by

$P_{g,i}^{\max}$.

In addition to the harvested RES, the power grid can also supply the needed $P_{g,i}^t$ per BS i. With a two-way energy trading facility, the BS can also sell its surplus energy to the grid at a fair price in order to reduce operational costs. Given the required energy $P_{g,i}^t$, the harvested energy E_i^t, and the battery charging energy $P_{b,i}^t$, the shortage energy that needs to be purchased from the grid for BS i is clearly $[P_{g,i}^t - E_i^t + P_{b,i}^t]^+$; or, the surplus energy (when the harvested energy is abundant) that can be sold to the grid is $[E_i^t - P_{g,i}^t - P_{b,i}^t]^+$ Note that both the shortage energy and surplus energy are non-negative, and we have at most one of them be positive at any time t for BS i.

Suppose that the energy can be purchased from the grid at price α^t, while the energy is sold to the grid at price β^t per slot t. Notwithstanding, we shall always set $\alpha^t > \beta^t$ for all t, to avoid meaningless buy-and-sell activities of the BSs for profit. Assuming that the prices α^t and β^t are known *a priori*, the *worst-case transaction cost* for BS i for the whole scheduling horizon is therefore given by

$$G(\{P_{g,i}^t\},\{P_{b,i}^t\}) := \max_{\mathbf{e}_i \in \mathcal{E}_i} \sum_{t=1}^{T} \Big(\alpha^t [P_{g,i}^t - E_i^t + P_{b,i}^t]^+ - \beta^t [E_i^t - P_{g,i}^t - P_{b,i}^t]^+ \Big). \quad (5.70)$$

5.4 Energy Management for CoMP Beamforming

Based on the models in Section 5.3, we consider here robust energy management for transmit beamforming in a CoMP cluster. Knowing only the uncertainty regions of the wireless channels and renewable energy arrivals, the central controller per cluster performs an (e.g. hour-) ahead-of-time schedule to optimize cooperative transmit beamforming vectors $\{\mathbf{w}_k^t\}$ and battery charging energy $\{P_{b,i}^t\}$, in order to minimize the worst-case total cost $\sum_{i\in\mathcal{I}} G(\{P_{g,i}^t\},\{P_{b,i}^t\})$, while satisfying the QoS guarantees $\widetilde{\mathrm{SINR}}_k(\{\mathbf{w}_k^t\}) \geq \gamma_k$, $\forall\ k$, over the scheduling horizon $\mathcal{T}$. For convenience, we introduce the auxiliary variables $P_i^t := P_{g,i}^t + P_{b,i}^t$, and formulate the problem as

$$\begin{aligned}
\underset{\{\mathbf{w}_k^t, P_{b,i}^t, C_i^t, P_i^t\}}{\text{minimize}} \quad & \sum_{i\in\mathcal{I}} G(\{P_i^t\}) && (5.71a)\\
\text{subject to} \quad & P_i^t = P_{c,i} + \sum_{k\in\mathcal{K}} \mathbf{w}_k^{t\,\mathcal{H}} \mathbf{B}_i \mathbf{w}_k^t + P_{b,i}^t, \quad i\in\mathcal{I},\ t\in\mathcal{T} && (5.71b)\\
& 0 \leq P_{c,i} + \sum_{k\in\mathcal{K}} \mathbf{w}_k^{t\,\mathcal{H}} \mathbf{B}_i \mathbf{w}_k^t \leq P_{g,i}^{\max}, \quad i\in\mathcal{I},\ t\in\mathcal{T} && (5.71c)
\end{aligned}$$

$$C_i^t = C_i^{t-1} + P_{b,i}^t, \quad i \in \mathcal{I},\ t \in \mathcal{T} \tag{5.71d}$$

$$0 \leq C_i^t \leq C_i^{\max}, \quad i \in \mathcal{I},\ t \in \mathcal{T} \tag{5.71e}$$

$$P_{b,i}^{\min} \leq P_{b,i}^t \leq P_{b,i}^{\max}, \quad i \in \mathcal{I},\ t \in \mathcal{T} \tag{5.71f}$$

$$-\varpi_i C_i^{t-1} \leq P_{b,i}^t, \quad i \in \mathcal{I},\ t \in \mathcal{T} \tag{5.71g}$$

$$\widetilde{\mathrm{SINR}}_k(\{\mathbf{w}_k^t\}) \geq \gamma_k, \quad k \in \mathcal{K},\ t \in \mathcal{T}. \tag{5.71h}$$

Consider for simplicity that (5.71) is feasible. The problem can in fact become infeasible if the SINR thresholds γ_k are high and the wireless channel qualities are not good enough. In this case, the subsequent problems (5.81a) are always infeasible. Recall that the proposed scheme is proposed to determine the ahead-of-time beamformers and energy schedules (offline). Such an infeasibility, once detected, can naturally lead to an admission control policy, i.e., to a criterion for dropping users or SINR requirements that render the problem infeasible [145].

Solving (5.71) can provide a robust solution for the smart-grid powered CoMP downlink with worst-case performance guarantees. It is worth mentioning here that thanks to the worst-case cost $G(\{P_i^t\})$, randomness introduced due to the wireless fading propagation and also due to the RES uncertainty has been eliminated; thus (5.71) contains only deterministic variables. Because of (5.71b), (5.71c), and (5.71h), the problem is nonconvex, which motivates the reformulation pursued next.

5.4.1 Convex Reformulation

First, let us consider convexity issues of the objective function $G(\{P_i^t\})$. Define $\psi^t := (\alpha^t - \beta^t)/2$ and $\phi^t := (\alpha^t + \beta^t)/2$, and then rewrite:

$$G(\{P_i^t\}) = \max_{\mathbf{e}_i \in \mathcal{E}_i} \sum_{t=1}^{T} \left(\psi^t |P_i^t - E_i^t| + \phi^t (P_i^t - E_i^t) \right) \ .$$

Since $\alpha^t > \beta^t \geq 0$, we have $\phi^t > \psi^t > 0$. It is then clear that $\psi^t |P_i^t - E_i^t| + \phi^t (P_i^t - E_i^t)$ is a convex function of P_i^t for any given E_i^t. As a pointwise maximization of these convex functions, $G(\{P_i^t\})$ is also a convex function of $\{P_i^t\}$ (even when the set $\mathcal{E}_i$ is non-convex) [13].

Except for (5.71b), (5.71c), and (5.71h), all other constraints are convex. We next rely on the popular SDP relaxation technique to convexify (5.71h). By the definitions

of $\mathcal{H}_k^t$ and $\widetilde{\text{SINR}}_k(\{\mathbf{w}_k^t\})$, the constraint $\widetilde{\text{SINR}}_k(\{\mathbf{w}_k^t\}) \geq \gamma_k$ can be rewritten as:

$$F_k(\boldsymbol{\delta}_k^t) \geq 0 \text{ for all } \boldsymbol{\delta}_k^t \text{ such that } \boldsymbol{\delta}_k^{t\,\mathcal{H}} \boldsymbol{\delta}_k^t \leq (\epsilon_k^t)^2 \tag{5.72}$$

where

$$F_k(\boldsymbol{\delta}_k^t) := (\hat{\mathbf{h}}_k^t + \boldsymbol{\delta}_k^t)^{\mathcal{H}} \left(\frac{\mathbf{w}_k^t \mathbf{w}_k^{t\,\mathcal{H}}}{\gamma_k^t} - \sum_{l \neq k} \mathbf{w}_l^t \mathbf{w}_l^{t\,\mathcal{H}} \right) (\hat{\mathbf{h}}_k^t + \boldsymbol{\delta}_k^t) - \sigma_k^2.$$

To this end, it clearly holds that $\mathbf{X}_k^t := \mathbf{w}_k^t \mathbf{w}_k^{t\,\mathcal{H}} \in \mathbb{C}^{MI \times MI}$ is positive semidefinite, and $\text{rank}(\mathbf{X}_k^t) = 1$. By using the S-procedure in Lemma 5.1, constraint (5.72) can be transformed to

$$\boldsymbol{\Gamma}_k^t := \begin{pmatrix} \mathbf{Y}_k^t + \tau_k^t \mathbf{I} & \mathbf{Y}_k^t \hat{\mathbf{h}}_k^t \\ \hat{\mathbf{h}}_k^{tH} \mathbf{Y}_k^{tH} & \hat{\mathbf{h}}_k^{tH} \mathbf{Y}_k^t \hat{\mathbf{h}}_k^t - \sigma_k^2 - \tau_k^t (\epsilon_k^t)^2 \end{pmatrix} \succeq \mathbf{0} \tag{5.73}$$

where $\tau_k^t \geq 0$ and

$$\mathbf{Y}_k^t := \frac{1}{\gamma_k} \mathbf{X}_k^t - \sum_{l \neq k} \mathbf{X}_l^t. \tag{5.74}$$

Introducing auxiliary variables τ_k^t and dropping the rank constraints $\text{rank}(\mathbf{X}_k^t) = 1$, $\forall\ k, t$, we can then relax (5.71) to:

$$\begin{aligned}
\underset{\{\mathbf{X}_k^t, \tau_k^t, P_i^t, P_{b,i}^t, C_i^t\}}{\text{minimize}} \quad & \sum_{i \in \mathcal{I}} G(\{P_i^t\}) && \text{(5.75a)} \\
\text{subject to} \quad & (5.71\text{d}) - (5.71\text{g}) \\
& P_i^t = P_{c,i} + \sum_{k \in \mathcal{K}} \text{Tr}(\mathbf{B}_i \mathbf{X}_k^t) + P_{b,i}^t,\ \forall\ i, t && \text{(5.75b)} \\
& 0 \leq \sum_{k \in \mathcal{K}} \text{Tr}(\mathbf{B}_i \mathbf{X}_k^t) \leq P_{g,i}^{\max} - P_{c,i},\ \forall\ i, t && \text{(5.75c)} \\
& \boldsymbol{\Gamma}_k^t \succeq \mathbf{0},\ \ \tau_k^t \geq 0,\ \ \mathbf{X}_k^t \succeq \mathbf{0},\ \forall\ k, t. && \text{(5.75d)}
\end{aligned}$$

In addition to the linear constraints (5.71d)–(5.71g), the quadratic power constraints (5.71b) and (5.71c) have now become linear, and SINR constraints (5.71h) become a set of convex SDP constraints in (5.75). Since the objective function is convex, problem (5.75) is a convex program that can be tackled by a centralized solver; e.g., using the projected subgradient descent approach. However, the feasible set (5.75b)–(5.75d) is the intersection of a semi-definite cone and a polytope, for which the iterative projection is complicated. To reduce computational complexity and enhance resilience to

failures, we next develop an efficient algorithm to solve (5.75) in a distributed fashion coordinated by different agents.

5.4.2 Lagrangian Dual Approach

Since (5.75) is convex, a Lagrange dual approach can be developed to efficiently find its solution. Let $\lambda_i^t,\ \forall\ i,t$ denote the Lagrange multipliers associated with the constraints (5.75b). With the convenient notation $\mathbf{Z} := \{\mathbf{X}_k^t, \tau_k^t, P_{b,i}^t, C_i^t, P_i^t\}$ and $\mathbf{\Lambda} := \{\lambda_i^t\}$, the partial Lagrangian function of (5.75) is

$$L(\mathbf{Z}, \mathbf{\Lambda}) := \sum_{i\in\mathcal{I}}\Big[G(\{P_i^t\}) + \sum_{t=1}^{T}\lambda_i^t\big(P_{c,i} + \sum_{k\in\mathcal{K}}\mathrm{Tr}(\mathbf{B}_i\mathbf{X}_k^t) + P_{b,i}^t - P_i^t\big)\Big]. \tag{5.76}$$

The Lagrange dual function is then given by

$$\begin{aligned} D(\mathbf{\Lambda}) := \underset{\mathbf{Z}}{\text{minimize}} \quad & L(\mathbf{Z}, \mathbf{\Lambda}) \\ \text{subject to} \quad & (5.71\text{d}) - (5.71\text{g}),\ (5.75\text{c}) - (5.75\text{d}) \end{aligned} \tag{5.77}$$

and the dual problem of (5.75) is:

$$\underset{\mathbf{\Lambda}}{\text{maximize}}\ D(\mathbf{\Lambda}). \tag{5.78}$$

Subgradient iterations: Let j denote the iteration index. To obtain the optimal solution $\mathbf{\Lambda}^*$ to the dual problem (5.78), we resort to the dual subgradient ascent method, which amounts to the following update

$$\lambda_i^t(j+1) = \lambda_i^t(j) + \mu(j)g_{\lambda_i^t}(j),\ \forall\ i,t \tag{5.79}$$

where $\mu(j) > 0$ is an appropriate stepsize. The subgradient $\mathbf{g}(j) := [g_{\lambda_i^t}(j),\ \forall\ i,t]$ can then be expressed as

$$g_{\lambda_i^t}(j) = P_{c,i} + \sum_{k\in\mathcal{K}}\mathrm{Tr}(\mathbf{B}_i\mathbf{X}_k^t(j)) + P_{b,i}^t(j) - P_i^t(j),\ \forall\ i,t \tag{5.80}$$

where $\mathbf{X}_k^t(j)$, $P_{b,i}^t(j)$, and $P_i^t(j)$ are given by

$$\begin{aligned} \{\mathbf{X}_k^t(j)\}_{k=1}^K \in \underset{\{\mathbf{X}_k^t, \tau_k^t\}}{\arg\min} & \sum_{i\in\mathcal{I},\, k\in\mathcal{K}}\lambda_i^t(j)\mathrm{Tr}(\mathbf{B}_i\mathbf{X}_k^t) \\ \text{s.t.} \quad & (5.75\text{c}) - (5.75\text{d}),\ \ \forall\ t \end{aligned} \tag{5.81a}$$

$$\{P_{b,i}^t(j)\}_{t=1}^T \in \underset{\{P_{b,i}^t, C_{i,t}\}}{\arg\min} \sum_{t=1}^T \lambda_i^t(j) P_{b,i}^t$$

$$\text{s.t.} \quad (5.71\text{d}) - (5.71\text{g}), \ \ \forall\ i \tag{5.81b}$$

$$\{P_i^t(j)\}_{t=1}^T \in \underset{\{P_i^t\}}{\arg\min} \left\{ G(\{P_i^t\}) - \sum_{t=1}^T \lambda_i^t(j) P_i^t \right\} . \tag{5.81c}$$

Solving the subproblems: Subproblems (5.81a) are standard SDPs per $t \in \mathcal{T}$; hence, $\{\mathbf{X}_k^t(j)\}_{k=1}^K$ for all t can be efficiently solved by general interior-point methods [13].

The subproblems (5.81b) are simple linear programs (LPs) over $\{P_{b,i}^t, C_i^t\}_{t=1}^T$ per $i \in \mathcal{I}$; hence, $\{P_{b,i}^t(j)\}_{t=1}^T$, $\forall\ i$, can be obtained by existing efficient LP solvers.

Due to the convexity of $G(\{P_i^t\})$, the subproblems (5.81c) are convex per $i \in \mathcal{I}$. Yet, because of the non-differentiability of $G(\{P_i^t\})$ due to the absolute value operator and the maximization over $\mathbf{e}_i \in \mathcal{E}_i$, the problem is challenging to be handled by existing general solvers. For this reason, we resort to the proximal bundle method to obtain $\{P_i^t(j)\}_{t=1}^T$. Upon defining $\tilde{G}(\mathbf{p}_i) := G(\mathbf{p}_i) - \sum_{t=1}^T \lambda_i^t(j) P_i^t$, where $\mathbf{p}_i := [P_i^1, \ldots, P_i^T]'$, the partial subgradient of $\tilde{G}(\mathbf{p}_i)$ with respect to P_i^t can be obtained as

$$\frac{\partial \tilde{G}(\mathbf{p}_i)}{\partial P_i^t} = \begin{cases} \alpha_i^t - \lambda_i^t(j), & \text{if } P_i^t \geq E_i^{t*} \\ \beta_i^t - \lambda_i^t(j), & \text{if } P_i^t < E_i^{t*} \end{cases} \tag{5.82}$$

where $\mathbf{e}_i^* := [E_i^{1*}, \ldots, E_i^{T*}]'$ for the given $\mathbf{p}_i$ is obtained as

$$\mathbf{e}_i^* \in \arg \max_{\mathbf{e}_i \in \mathcal{E}_i} \sum_{t=1}^T \left(\psi^t |P_i^t - E_i^t| + \phi^t (P_i^t - E_i^t) \right). \tag{5.83}$$

It can be readily checked that the objective function in (5.83) is convex in $\mathbf{e}_i$ under the condition $\alpha^t > \beta^t$, $\forall\ t \in \mathcal{T}$. Hence, the globally optimal solution is attainable at the extreme points of $\mathcal{E}_i$. Leveraging the special structure of $\mathcal{E}_i$, we utilize an efficient vertex enumerating algorithms to obtain $\mathbf{e}_i^*$ directly, as detailed next.

5.4.3 Proximal Bundle Method

Given the partial subgradient in (5.82), nonsmooth convex optimization algorithms can be employed to solve the subproblem (5.81c). A state-of-the-art bundle method [13,

Ch. 6], [51] will be developed here with guaranteed convergence to the optimal $\{P_i^t(j)\}_{t=1}^T$; see also [165].

Similar to cutting plane methods, the idea of bundle method is to approximate the epigraph of an objective by the intersection of a number of halfspaces. The latter are generated through the supporting hyperplanes, referred to as cuts, by using the subgradients. Specifically, letting ℓ denote the iteration index of the bundle method, the iterate $\mathbf{p}_{i,\ell+1}$ is obtained by minimizing a polyhedral (piecewise linear) approximation of $\tilde{G}(\mathbf{p}_i)$ with a quadratic proximal regularizer

$$\mathbf{p}_{i,\ell+1} := \arg\min_{\mathbf{p}_i \in \mathbb{R}^T} \left\{ \hat{G}_\ell(\mathbf{p}_i) + \frac{\rho_\ell}{2} \|\mathbf{p}_i - \mathbf{y}_\ell\|^2 \right\} \tag{5.84}$$

where $\hat{G}_\ell(\mathbf{p}_i) := \max\{\tilde{G}(\mathbf{p}_{i,0}) + \mathbf{g}_0'(\mathbf{p}_i - \mathbf{p}_{i,0}), \ldots, \tilde{G}(\mathbf{p}_{i,\ell}) + \mathbf{g}_\ell'(\mathbf{p}_i - \mathbf{p}_{i,\ell})\}$; $\mathbf{g}_{i,\ell}$ is the subgradient of $\tilde{G}(\mathbf{p}_i)$ evaluated at the point $\mathbf{p} = \mathbf{p}_{i,\ell}$ [cf. (5.82)]; and the proximity weight ρ_ℓ controls the stability of iterates.

Different from the proximal CPM, the bundle method updates its proximal center $\mathbf{y}_\ell$ according to a descent query

$$\mathbf{y}_{\ell+1} = \begin{cases} \mathbf{p}_{i,\ell+1}, & \text{if } \tilde{G}(\mathbf{y}_\ell) - \tilde{G}(\mathbf{p}_{i,\ell+1}) \geq \theta \eta_\ell \\ \mathbf{y}_\ell, & \text{if } \tilde{G}(\mathbf{y}_\ell) - \tilde{G}(\mathbf{p}_{i,\ell+1}) < \theta \eta_\ell \end{cases} \tag{5.85}$$

where $\theta \in (0,1)$ is a pre-selected constant, and $\eta_\ell := \tilde{G}(\mathbf{y}_\ell) - \left(\hat{G}_\ell(\mathbf{p}_{\ell+1}) + \frac{\rho_\ell}{2} \|\mathbf{p}_{\ell+1} - \mathbf{y}_\ell\|^2 \right)$ is the predicted descent of the objective in (5.84). Essentially, if the actual descent amount $\tilde{G}(\mathbf{y}_\ell) - \tilde{G}(\mathbf{p}_{\ell+1})$ is no less than a θ fraction of the predicted counterpart η_ℓ, then the iterate takes a "serious" step updating its proximal center $\mathbf{y}_{\ell+1}$ to the latest point $\mathbf{p}_{\ell+1}$; otherwise it is just a "null" step with the center unchanged. The intelligent query (5.85) enables the bundle method to find "good" proximal centers along the iterates, and hence it converges faster than the proximal CPM. In addition, depending on whether a serious or a null step is taken, the proximity weight ρ_ℓ can be updated accordingly to further accelerate convergence [83]; that is,

$$\rho_{\ell+1} = \begin{cases} \max(\rho_\ell/10, \rho_{\min}), & \text{if } \tilde{G}(\mathbf{y}_\ell) - \tilde{G}(\mathbf{p}_{i,\ell+1}) \geq \theta \eta_\ell \\ \min(10\rho_\ell, \rho_{\max}), & \text{if } \tilde{G}(\mathbf{y}_\ell) - \tilde{G}(\mathbf{p}_{i,\ell+1}) < \theta \eta_\ell. \end{cases}$$

The algorithm terminates if $\mathbf{y}_\ell = \mathbf{p}_{i,\ell+1}$, while finite termination is achievable if both the objective and the feasible set are polyhedral [13, Sec. 6.5.3].

Now, to complete the proximal bundle method for solving (5.81c), we only need solve problem (5.84). Using an auxiliary variable u, (5.84) can be re-written as

$$\underset{\mathbf{p}_i,u}{\text{minimize}} \quad u + \frac{\rho_\ell}{2}\|\mathbf{p}_i - \mathbf{y}_\ell\|^2 \tag{5.86a}$$

$$\text{subject to} \quad \tilde{G}(\mathbf{p}_{i,n}) + \mathbf{g}'_{i,n}(\mathbf{p}_i - \mathbf{p}_{i,n}) \leq u,\ n = 0, 1, \ldots, \ell. \tag{5.86b}$$

Introducing multipliers $\boldsymbol{\xi} \in \mathbb{R}_+^{\ell+1}$ and letting $\mathbf{1}$ denote the all-ones vector, the Lagrangian function is given as

$$L(u, \mathbf{p}_i, \boldsymbol{\xi}) = (1 - \mathbf{1}'\boldsymbol{\xi})u + \frac{\rho_\ell}{2}\|\mathbf{p}_i - \mathbf{y}_\ell\|^2 + \sum_{n=0}^{\ell} \xi_n\big(\tilde{G}(\mathbf{p}_{i,n}) + \mathbf{g}'_{i,n}(\mathbf{p}_i - \mathbf{p}_{i,n})\big). \tag{5.87}$$

The optimality condition $\nabla_{\mathbf{p}_i} L(u, \mathbf{p}_i, \boldsymbol{\xi}) = \mathbf{0}$ yields

$$\mathbf{p}_i^* = \mathbf{y}_\ell - \frac{1}{\rho_\ell}\sum_{n=0}^{\ell} \xi_n \mathbf{g}_{i,n}. \tag{5.88}$$

Substituting (5.88) into (5.87), the dual of (5.86) is

$$\underset{\boldsymbol{\xi}}{\text{maximize}} \quad -\frac{1}{2\rho_\ell}\left\|\sum_{n=0}^{\ell} \xi_n \mathbf{g}_{i,n}\right\|^2 + \sum_{n=0}^{\ell} \xi_n\big(\tilde{G}(\mathbf{p}_{i,n}) + \mathbf{g}'_{i,n}(\mathbf{y}_\ell - \mathbf{p}_{i,n})\big) \tag{5.89a}$$

$$\text{subject to} \quad \boldsymbol{\xi} \succeq \mathbf{0},\ \mathbf{1}'\boldsymbol{\xi} = 1\ . \tag{5.89b}$$

It can be readily seen that the dual problem (5.89) is essentially a QP over the probability simplex, which can be solved efficiently; see e.g., [11].

5.4.4 Optimality and Distributed Implementation

When a constant stepsize $\mu(j) = \mu$ is adopted, the subgradient iterations (5.79) are guaranteed to converge to a neighborhood of the optimal $\boldsymbol{\Lambda}^*$ for the dual problem (5.78) from any initial $\boldsymbol{\Lambda}(0)$. The size of the neighborhood is proportional to the stepsize μ. If we adopt a sequence of non-summable diminishing stepsizes satisfying $\lim_{j\to\infty} \mu(j) = 0$ and $\sum_{j=0}^{\infty} \mu(j) = \infty$, then the iterations (5.79) will asymptotically converge to the exact $\boldsymbol{\Lambda}^*$ as $j \to \infty$ [13].

The objective function (5.75a) is not strictly convex because it does not involve all optimization variables. Hence, when it comes to primal convergence, extra care is

necessary [53, 148]. Specifically, the optimal primal can be attained either by adding a strictly convex regularization term, or, by utilizing the augmented Lagrangian. Here, we will simply implement the *Cesáro averaging* method [84] to obtain the optimal power schedules. With $\mu_{\text{sum}}^m := \sum_{j=1}^m \mu(j)$, the running average is

$$\bar{\mathbf{Z}}^m := \frac{1}{\mu_{\text{sum}}^m} \sum_{j=1}^m \mu(j)\mathbf{Z}(j) \tag{5.90}$$

which can be efficiently computed in a recursive way

$$\bar{\mathbf{Z}}^m := \frac{\mu(m)}{\mu_{\text{sum}}^m}\mathbf{Z}(m) + \frac{\mu_{\text{sum}}^{m-1}}{\mu_{\text{sum}}^m}\bar{\mathbf{Z}}^{m-1} \ . \tag{5.91}$$

Note that if a constant stepsize $\mu(j) \equiv \mu$ is adopted, (5.90) and (5.91) boil down to the ordinary running average

$$\bar{\mathbf{Z}}^m = \frac{1}{m}\sum_{j=1}^m \mathbf{Z}(j) = \frac{1}{m}\mathbf{Z}(m) + \frac{m-1}{m}\bar{\mathbf{Z}}^{m-1} \, .$$

If the obtained solution to problem (5.75) satisfies the condition $\text{rank}(\mathbf{X}_k^{t\,*}) = 1 \, \forall \ k, t$, then it clearly yields the optimal beamforming vectors $\mathbf{w}_k^{t\,*}$ as the (scaled) eigenvector with respect to the only positive eigenvalue of $\mathbf{X}_k^{t\,*}$ for the original problem (5.71). Fortunately, it was shown in [128, Thm. 1] that the S-procedure based SDP (5.75) always returns a rank-one optimal solution $\mathbf{X}_k^{t\,*}, \, \forall \ k, t$, when the uncertainty bounds ϵ_k^t are sufficiently small. If ϵ_k is large, the existence of rank-one optimal solutions of (5.75) cannot be provably guaranteed. In this case, a randomized rounding strategy [137] needs to be adopted to obtain vectors $\mathbf{w}_k^{t\,*}$ from $\mathbf{X}_k^{t\,*}$ to nicely approximate the solution of the original problem (5.71). Even though no proof is available to ensure a rank-one solution when ϵ_k is large, it has been extensively observed in simulations that the SDP relaxation always returns a rank-one optimal solution [128]. This confirms the view that the optimal beamforming vectors for the original problem (5.71) will be obtained by the proposed approach with high probability.

The subgradient iterations can be run in a distributed fashion. Specifically, the central controller can maintain the Lagrange multipliers $\mathbf{\Lambda}(j)$ and broadcast them to all BSs via backhaul links. Given the current $\mathbf{\Lambda}(j)$, the central controller solves the subproblems (5.81a) to obtain the beamforming vectors for all BSs. On the other hand,

each BS solves its own subproblems (5.81b)–(5.81c), which are decoupled across BSs. The BSs send back to the central controller $\{P_{g,i}^t(j)\}_{t=1}^T$, $\{P_{b,i}^t(j)\}_{t=1}^T$, and $\{P_i^t(j)\}_{t=1}^T$, which are in turn used to update $\mathbf{\Lambda}(j+1)$ through the subgradient iterations (5.79).

5.5 Numerical Tests

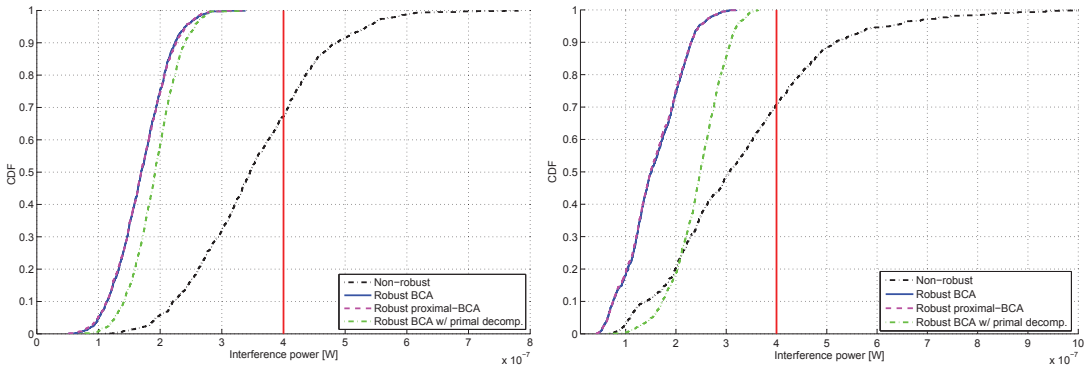
Numerical test are implemented in this section to demonstrate the performance of the novel approaches. Section 5.5.1 presents the simulation results for the MIMO CR network (Sec. 5.1–5.2), while Section 5.5.2 is for the green CoMP system (Sec. 5.3–5.4).

5.5.1 Simulations for MIMO CR Networks

In this section, numerical tests are performed to verify the performance merits of the novel design. The path loss obeys the model $d^{-\eta}$, with d the distance between nodes, and $\eta = 3.5$. A flat Rayleigh fading model is employed. For simplicity, the distances of links $U_k^t \to U_k^r$ are all set to $d_{k,k} = 30$ m; for the interfering links $\{U_k^t \to U_j^r, j \neq k\}$ distances are uniformly distributed over the interval $30 - 100$ m. As for the distances between CR transmitters and PU receivers, two different cases are considered: (c1) the PU receivers are located at a distance from the CRs that is uniformly distributed over $70-100$ m; and, (c2) the CR-to-PU distances are uniformly distributed over $30-100$ m. Finally, the maximum transmit-power and the noise power are identical for all CRs. For the proximal point-based algorithm, the penalty factors $\{\tau_k\}$ are selected equal to 0.1.

To validate the effect of the robust interference constraint, the cumulative distribution functions (CDF) of the interference power at the PU are depicted in Fig. 5.3. Four CR pairs and one PU receiver are considered, all equipped with two antennas. The maximum transmit-powers and noise powers are set so that the (maximum) signal-to-noise ratio (SNR) defined as $\text{SNR} := p_k^{\max}(d_{k,k}^{-\eta})/\sigma_k^2$ equals 15 dB. The total interference threshold is set to $\iota^{\max} = 4 \cdot 10^{-7}$ W, and it is equally split among the CR transmitters. The channel uncertainty is set to $\epsilon_k^2 = \rho \cdot \|\hat{\mathbf{G}}_k\|_F^2$ [144], with $\rho = 0.05$. CDF curves are obtained using $2,000$ Monte Carlo runs. In each run, independent channel realizations are generated. The Matlab-based package `CVX` [59] along with `SeDuMi` [131] are used to solve the proposed robust beamforming problems.

The trajectories provided in Fig. 5.3 refer to the block coordinate ascent (BCA)

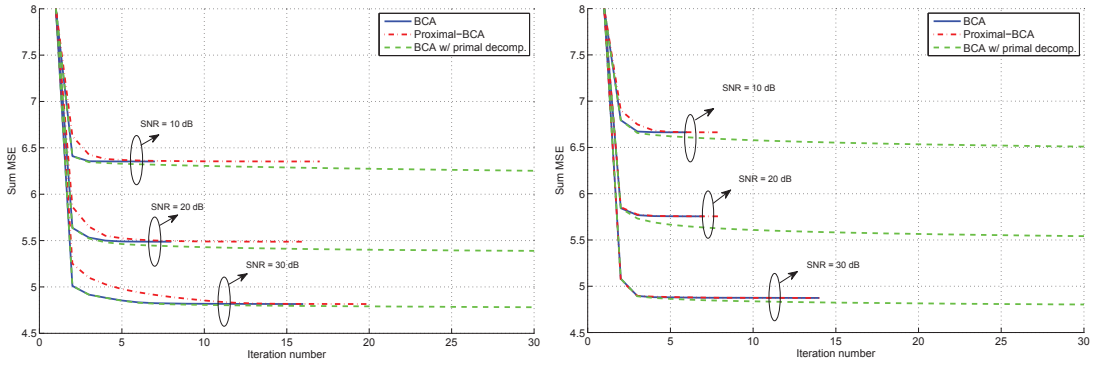


(a) Case (c1) (b) Case (c2)

Figure 5.3: Interference cumulative distribution function (CDF).

algorithm described in Section 5.2; the one with the proximal point-based regularization (proximal-BCA) explained in Section 5.2.3; and the non-robust solver of (5.7), where the estimates $\{\hat{\mathbf{G}}_k\}$ are used in place of the true channels $\{\mathbf{G}_k\}$. Furthermore, the green trajectory corresponds to (5.55), where the subgradient projection (5.60) is implemented at the end of each BCA cycle, which includes K updates of $\mathbf{Q}_k$ for $k = 1, \ldots, K$. As expected, the proposed robust schemes enforce the interference constraint strictly in both scenarios (c1) and (c2). In fact, the interference never exceeds the tolerable limit shown as the vertical red solid line in Fig. 5.3. The CDFs corresponding to the proposed BCA and its proximal counterpart nearly coincide. In fact, the two algorithms frequently converge to identical stationary points in this particular simulation setup. Notice that with the primal decomposition approach the beamforming strategy is less conservative. On the contrary, the non-robust approach frequently violates the interference limit (more than 30% of the time). Finally, comparing Fig. 5.3(a) with Fig. 5.3(b), one notices that the interference inflicted to the PU under (c1) and the one under (c2) are approximately of the same order. Since in the second case the CR-to-PU distances are smaller, the CR transmitters lower their transmit-powers to protect the PU robustly.

Convergence of the proposed algorithms with given channel realizations and over variable SNRs is illustrated in Fig. 5.4. It is clearly seen that the total MSEs decrease monotonically across fast-converging iterations, and speed is roughly identical in (c1) and (c2). As expected, the proximal point-based algorithm exhibits a slightly slower



(a) Case (c1) (b) Case (c2)

Figure 5.4: Convergence of proposed algorithms (SNR = 10, 20, and 30 dB).

convergence rate. Notice also that the primal decomposition method returns improved operational points, especially for medium and low SNR values. Furthermore, the gap between the sum-MSEs obtained with and without the primal decomposition scheme is more evident under (c2). Clearly, the sum-MSEs at convergence in (c2) are higher than the counterparts of (c1). This is because CRs are constrained to use a relatively lower transmit-power in order to enforce the robust interference constraints; this, in turn, leads to higher sum-MSEs and may reduce the quality of the CR-to-CR communications.

In Fig. 5.5, the achieved sum-MSE at convergence is reported as a function of the total interference threshold. Two sizes of the uncertainty region are considered with $\rho = 0.05$ and $\rho = 0.1$. Focusing on the first case, it can be seen that the two achieved sum-MSEs first monotonically decrease as the interference threshold increases, and subsequently they remain approximately constant. Specifically, for smaller $\iota^{\max}$, the transmit-CRs are confined to relatively low transmit-powers in order to satisfy the interference constraint. On the other hand, for high values of $\iota^{\max}$, the interference constraint is no longer a concern, and the attainable sum-MSEs are mainly due to CR self-interference. Notice also that for $\rho = 0.1$ the sum-MSEs are clearly higher, although they present a trend similar to the previous case. This is because the uncertainty region in (5.12c) becomes larger, which results in a higher sum-MSE.

In order to compare performance of the proposed algorithms, the total MSE obtained at convergence is depicted in Fig. 5.6(a) for 50 different experiments. In each experiment, independent channel realizations are generated. The SNR is set to 15 dB. It is

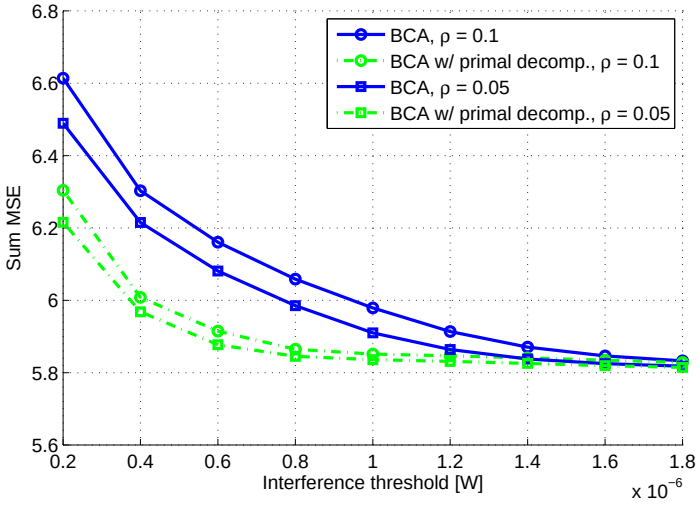
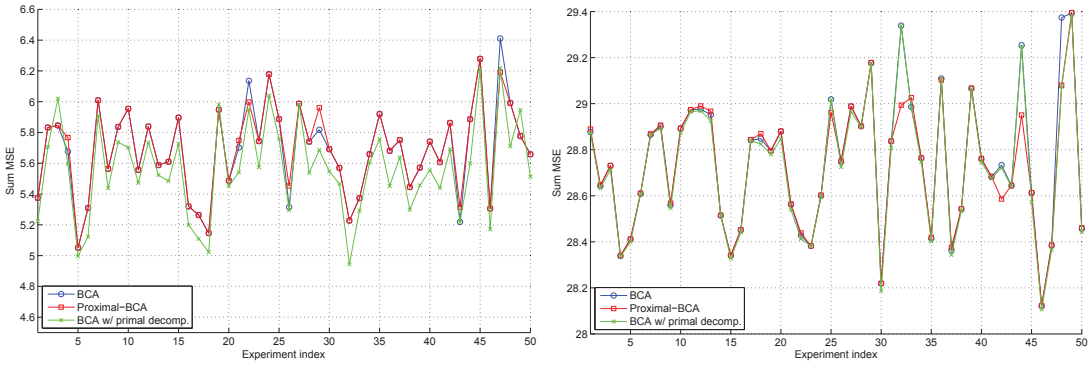


Figure 5.5: Achieved sum-MSE as a function of $\iota^{\max}$ (SNR = 10 dB).

clearly seen that the objectives values of the two proposed methods often coincide. The differences presented in a few experiments are caused by convergence to two different stationary points. In this case, it is certainly convenient to employ the first algorithm, as it ensures faster convergence (see Fig. 5.4(a)) without appreciable variations in the overall MSE. Notice that a smaller mean-square error can be obtained by resorting to the primal decomposition technique.

In Fig. 5.6(b), the simulation setup involves eight CR pairs and one PU receiver. The CR transmitters have four antennas, while the receiving CRs and the PU are equipped with two antennas. The distances $d_{k,k}$ are set to 50 m, while $\{d_{k,j}\}_{k\neq j}$ distances are uniformly distributed in the interval between 30 and 250 m. Finally, CR-to-PU distances are uniformly distributed between 100 and 200 m. Clearly, matrices $\{\mathbf{H}_{k,k}\}$ here do not have full column rank. It is observed that about 10% of the times the proximal point based algorithm yields smaller values of the sum-MSE than Algorithm 5.1. This demonstrates that Algorithm 5.1 may not converge to a stationary point, or, it returns an MSE that is likely to be worse than that of the proximal point-based scheme.

Fig. 5.7 depicts the CDFs of the difference between the sum-MSE obtained with BCA, along with the ones obtained with proximal-BCA and with the primal decomposition method. The simulation setups of Figs. 5.6(a) and 5.6(b) are considered. In the



(a) $K = 4$, $M_k = N_k = 2$, $L = 2$ (b) $K = 8$, $M_k = 4$, $N_k = 2$, $L = 2$

Figure 5.6: Achieved sum-MSE (SNR = 15 dB).

first case, it can be seen that for over 80% of the trials the BCA and proximal-BCA methods yield exactly the same solution. Moreover, BCA with primal decomposition performs better than the BCA method about 90% of the time. Specifically, the gain can be up to 0.765, which corresponds to approximately 14% of the average sum MSE of the BCA. In the second case, the proximal-BCA returns a smaller sum-MSE with higher frequency.

5.5.2 Simulations for CoMP Systems

In this section, simulated tests are presented to verify the performance of the proposed approach. The Matlab-based modeling package `CVX 2.1` [59] along with the solvers `MOSEK 7.0` [5] and `Sedumi 1.02` [131] are used to specify and solve the resulting optimization problems. All numerical tests are implemented on a computer workstation with Intel cores 3.40 GHz and 32 GB RAM.

The scheduling horizons of the considered CoMP network is $T = 8$. Two configurations are tested: (C1) $I = 2$ BSs and $K = 10$ end users (small size); and (C2) $I = 6$ BSs and $K = 30$ end users (large size). All wireless channels are assumed to be flat Rayleigh fading, and normalized to unit power. The noises are modeled as circularly-symmetric Gaussian random vectors. Without loss of generality, the effects of path loss, shadowing, and Doppler fading are ignored. Parameters including the limits of P_{G_i}, C_i, P_{b_i} and the discharging efficiency ϖ_i are listed in Table 5.1. A polyhedral uncertainty set (5.68)

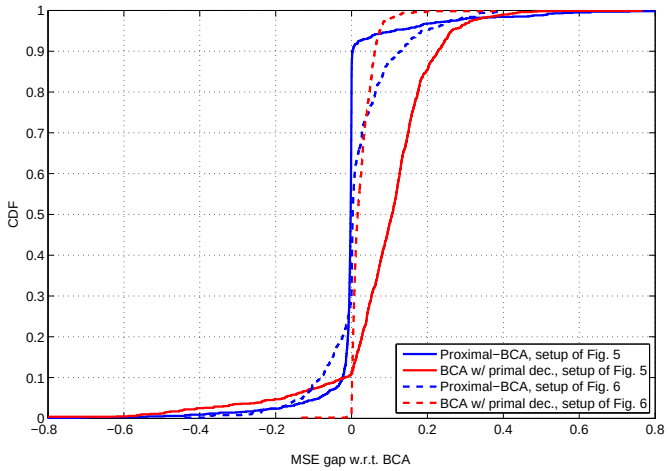


Figure 5.7: CDF of sum-MSE gaps (relative to the BCA) using proximal-BCA (blue) and primal decomposition (red): MSE(BCA) − MSE(proximal-BCA) and MSE(BCA) − MSE(BCA-primal decomp.).

with a single sub-horizon (no time partition) is considered for the RES. In Table 5.2, the energy purchase price α^t is given across the entire time horizon. The selling price is set as $\beta^t = r\alpha^t$ with $r \in [0, 1]$. In addition, the lower limits $\{\underline{E}_i^t\}_{i\in\mathcal{I}}$ are listed therein, which were rescaled from real data we obtained from the MISO [93]. The upper limits were set to $\overline{E}_i^t = 10\underline{E}_i^t$, while the total horizon bounds are $E_i^{\max} = 0.9\sum_t \overline{E}_i^t$.

First, convergence of the objective value (5.75a), and the ℓ_2-norm of the subgradient of the running-average Lagrange multiplier (5.80) is verified in Fig. 5.8(a). It can be seen that both metrics converge within a few hundred iterations, which confirms the validity of the dual decomposition approach along with the subgradient ascent algorithm. With the Cesàro averages, convergence of the dual and primal variables was also confirmed, but it is omitted due to limited space.

Fig. 5.8(b) depicts the effectiveness of the proposed bundle method minimizing the convex nonsmooth objective (5.81c). Clearly, incorporating the scheme of dynamically changing the proximity weight ρ_ℓ, the bundle algorithm converges very fast; typically, within 10 iterations.

The optimal power schedules of $\bar{P}_i^t$ are depicted in Figs. 5.9(a) and 5.9(b). For both configurations, the stairstep curves show that the lowest levels of $\bar{P}_i^t$ occur from slot 4

Table 5.1: Generating capacities, battery initial energy and capacity, charging limits and efficiency.

Unit	$P_{G_i}^{\min}$	$P_{G_i}^{\max}$	C_i^0	$C_i^{\max}$	$P_{b_i}^{\min}$	$P_{b_i}^{\max}$	ϖ_i
1	0	50	5	30	-10	10	0.95
2	0	45	5	30	-10	10	0.95
3	0	45	5	30	-10	10	0.95
4	0	45	5	30	-10	10	0.95
5	0	50	5	30	-10	10	0.95
6	0	45	5	30	-10	10	0.95

to 6. This is because the energy selling and purchase prices are relatively high during these horizons [cf. Table 5.2], which drives the BSs's power consumption low in order to minimize the transaction cost.

Figs. 5.10(a) and 5.10(b) show the optimal battery (dis-)charging amount $\bar{P}_{B_i}^t$, while Figs. 5.11(a) and 5.11(b) depict the state of charge $\bar{C}_i^t$. As a component part of $\bar{P}_i^t$, $\bar{P}_{B_i}^t$ exhibits a similar trend in response to the price fluctuation; that is, there is a relative large amount of battery discharging ($\bar{P}_{B_i}^t < 0$) when the corresponding price is high. Note that both $\bar{P}_{B_i}^t$ and $\bar{C}_i^t$ never exceed their lower and upper limits [cf. Table 5.1].

Robustness of the worst-case design to the uncertainty of channel estimates [cf. (5.63)] is confirmed in Figs. 5.12(a) and 5.12(b). The red solid line indicates the SINR threshold $\gamma_k = 0.1$ that is set common to all users for simplicity. The non-robust scheme simply considers the estimated channel $\hat{\mathbf{h}}_k^t$ as the actual one, and plugs it into the worst-case SINR control design (5.65). This constraint can be relaxed to a linear matrix inequality: $\hat{\mathbf{h}}_k^{tH} \mathbf{Y}_k^t \hat{\mathbf{h}}_k^t - \sigma_k^2 \geq 0$, which is a relaxed version of the proposed counterpart in (5.73). For both the robust and non-robust approaches, each transmit beamformer $\mathbf{w}_k^t$ is obtained as the principal eigenvector associated with the largest eigenvalue of the resulting $\bar{\mathbf{X}}_k^t$. The CDF of the actual SINR is obtained by evaluating (5.62) with $5,000$ independent and identically distributed (i.i.d.) channel realizations. The channel perturbations $\{\boldsymbol{\delta}_k^t\}$ are first generated as complex Gaussian, and then rescaled to the boundary of the spherical region $\mathcal{H}_k^t$ [cf. (5.63)]. It can be seen that 20% of the realizations of the non-robust scheme does not satisfy the SINR constraint, while only about 2% for the proposed approach. Note that the SDP relaxation is not always exact, which results in violating the SINR threshold for a few channel realizations.

Table 5.2: Limits of forecasted wind power and energy purchase prices

Slot	1	2	3	4	5	6	7	8
$\underline{E}_1^t$	2.47	2.27	2.18	1.97	2.28	2.66	3.10	3.38
$\underline{E}_2^t$	2.57	1.88	2.16	1.56	1.95	3.07	3.44	3.11
$\underline{E}_3^t$	2.32	2.43	1.27	1.39	2.14	1.98	2.68	4.04
$\underline{E}_4^t$	2.04	1.92	2.33	2.07	2.13	2.36	3.13	4.16
$\underline{E}_5^t$	2.11	1.19	2.26	2.19	1.55	2.71	3.37	2.45
$\underline{E}_6^t$	2.01	2.29	2.20	0.98	2.43	3.22	2.74	3.93
α^t	0.402	0.44	0.724	1.32	1.166	0.798	0.506	0.468

Finally, CDFs of the transaction cost are depicted in Fig. 5.13. The proposed robust approach is compared with a heuristic scheme that assumes the expected renewable generation $\hat{E}_i^t = \frac{1}{2}(\underline{E}_i^t + \overline{E}_i^t)$ is available and prefixed for problem (5.75). The CDF curves were plotted by evaluating the transaction cost (5.75a) with 10^5 realizations of $\{E_i^t\}_{i,t}$ and the obtained optimal solutions $\{\bar{P}_i^t\}_{i,t}$. The RES realizations were generated as $\{\tilde{E}_i^t\}_{i,t} = \underline{E}_i^t + \kappa U(\overline{E}_i^t - \underline{E}_i^t)$, where U is a uniform random variable on $[0, 1]$. Three cases with $\kappa = 0.01, 0.1, 0.5$ were tested. Clearly, transaction costs of both the proposed and the heuristic methods decrease with the increase of κ. Since a larger value of κ implies more harvested renewables energy yielding a reduced transaction cost. Note that negative transaction costs means net profits are obtained by selling surplus renewables back to the smart grid. Interestingly, for all cases, the proposed approach always outperforms the heuristic scheme with less transaction costs. This is because the proposed schedules of $\{P_{g,i}^t, P_{b,i}^t\}_{i,t}$ are robust to the worst-case renewable generation $\{E_i^t\}_{i,t}$. However, in practice more RES is often available than the worst case. Hence, the proposed method has a larger profit-making capability than the heuristic alternative.

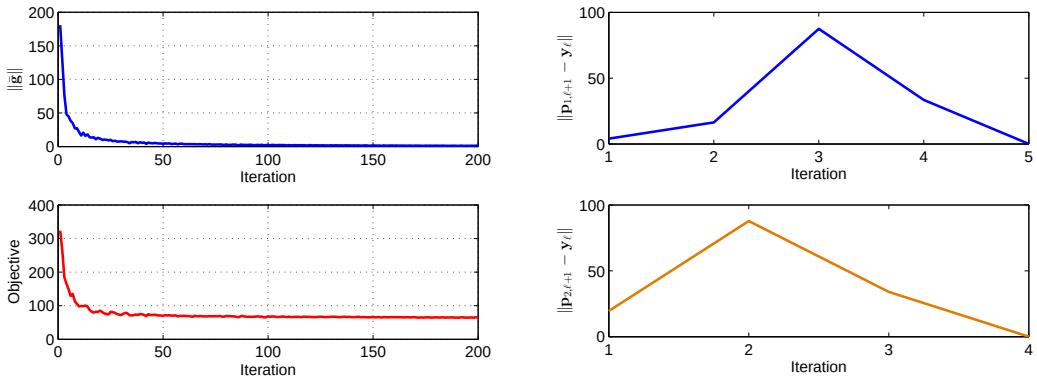


Figure 5.8: (a) Convergence of the objective value and the subgradient norm of Lagrange multipliers ($M = 2$, $K = 10$, $r = 1$). (b) Convergence of the bundle method solving the subproblem (5.81c) ($M = 2$, $K = 10$, $r = 1$, $j = 2$).

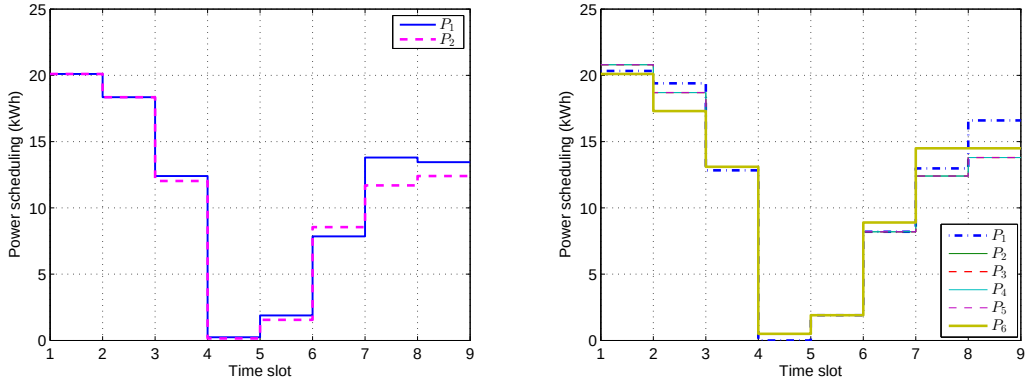


Figure 5.9: (a) Optimal power schedule of $\bar{P}_i^t$ ($I = 2$, $M = 2$, $K = 10$, $r = 1$). (b) Optimal power schedule of $\bar{P}_i^t$ ($I = 6$, $M = 2$, $K = 20$, $r = 1$).

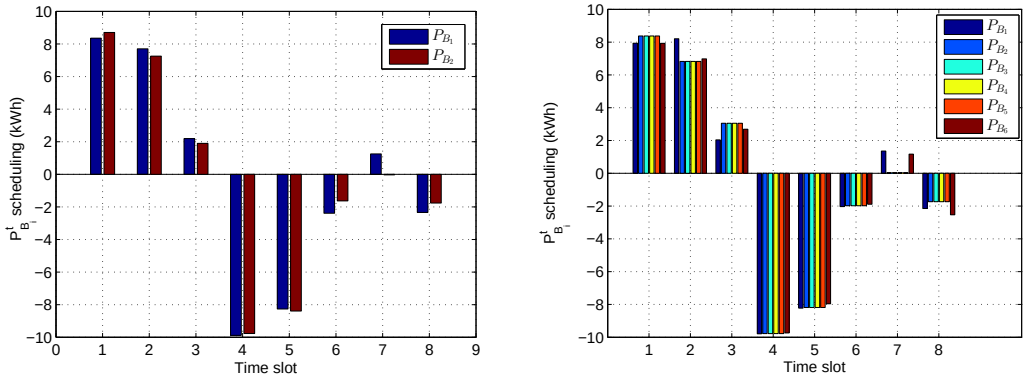


Figure 5.10: (a) Optimal power schedule of $\bar{P}^t_{B_i}$ ($I = 2$, $M = 2$, $K = 10$, $r = 1$). (b) Optimal power schedule of $\bar{P}^t_{B_i}$ ($I = 6$, $M = 2$, $K = 20$, $r = 1$).

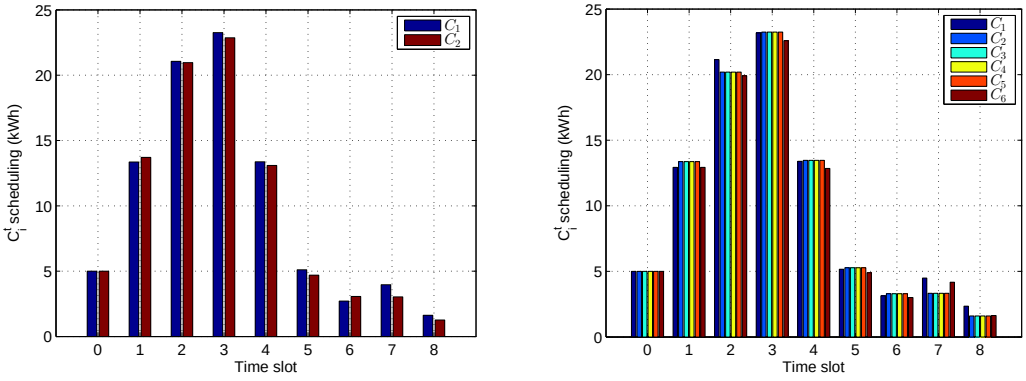


Figure 5.11: (a) Optimal power schedule for $\bar{C}^t_i$ ($I = 2$, $M = 2$, $K = 10$, $r = 1$). (b) Optimal power schedule for $\bar{C}^t_i$ ($I = 6$, $M = 2$, $K = 20$, $r = 1$).

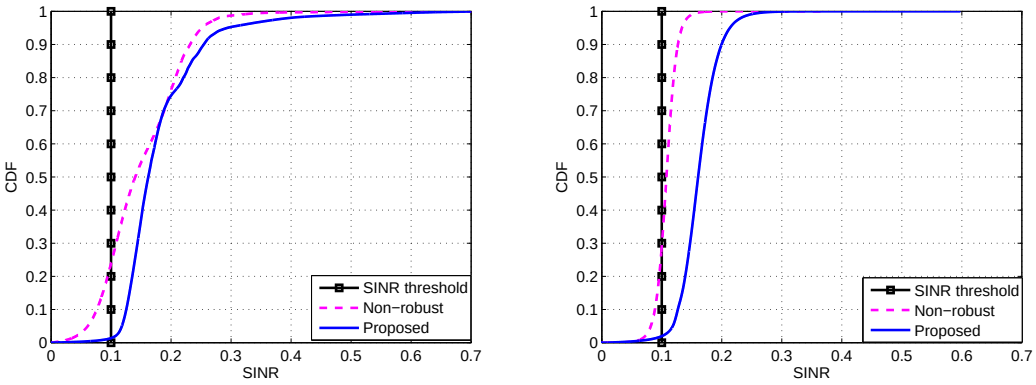


Figure 5.12: (a) SINR CDF ($I = 2$, $M = 2$, $K = 10$, $r = 1$). (b) SINR CDF ($I = 6$, $M = 2$, $K = 20$, $r = 1$).

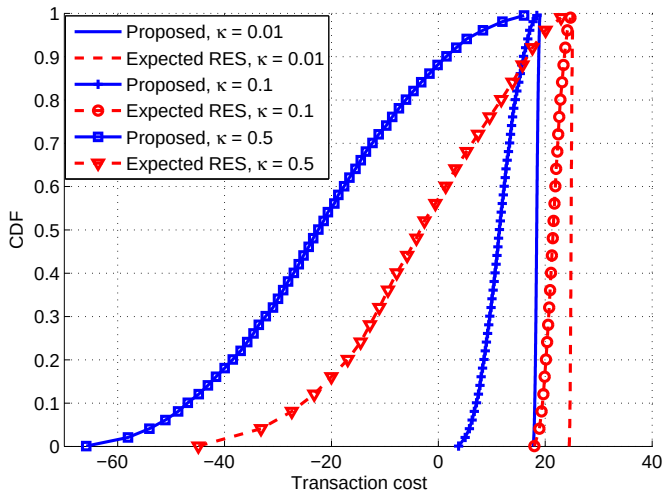


Figure 5.13: Cumulative distribution function (CDF) of the transaction cost ($I = 2$, $M = 2$, $K = 10$, $r = 0.3$).

Chapter 6

Conclusions and Future Directions

Leveraging modern optimization and signal processing tools, robust and distributed resource management techniques were developed in this thesis for potential use by the smart power grid and wireless communication networks. To conclude the present thesis, a summary of the main results and interesting directions for future research are provided in this final chapter.

6.1 Thesis Summary

In Chapter 2, a distributed energy management approach is developed tailored for microgrids with high penetration of renewable energy sources. By introducing the notion of committed renewable energy, a novel model is introduced to deal with the challenging constraint of the supply-demand balance raised by the intermittent nature of renewable energy sources. Not only the conventional generation costs, utilities of the adjustable loads, and distributed storage costs are accounted for, but also the worst-case transaction cost is included in the objective. To schedule power in a distributed fashion, the dual decomposition method is utilized to decompose the original problem into smaller subproblems solved by the LCs of conventional generators, dispatchable loads, DS units and the RES. The material in this chapter draws from [162, 165, 166].

Chapter 3 deals with multi-period economic dispatch and DC-OPF with multiple

wind farms. Risk-constrained optimization problems are formulated based on the LOLP. To circumvent the lack of knowledge of the spatio-temporal joint distribution of the wind power outputs, a scenario approximation technique via Monte Carlo sampling is developed. The attractive features and practical impact of this work are is three directions: i) the risk-constrained formulation is applicable to many practical power systems, including microgrids in island mode, where energy import from the main grid is not possible; ii) the scenario approach enables economic and risk-limited scheduling of smart grids with increasingly higher renewable energy penetration, without relying on specific probabilistic assumptions about the renewable generation; and iii) the special problem structure renders the scenario approach applicable to large-scale problems and very computationally efficient. The material in this chapter draws from [164, 167].

In Chapter 4, the day-ahead market clearing problem is investigated with controllable loads and/or renewable energy. A fast convergent and scalable distributed solver is first developed for MC with large-scale residential DR. Leveraging the dual decomposition technique, only the aggregator-users balance constraint is dualized in order to separate problems for the ISO and each aggregator, while respecting end-user privacy concerns. Simulated tests highlight the merits of the proposed approach for multiplier updates based on the disaggregated bundle method. MC with high-penetration wind power is further analyzed by formulating a stochastic optimization problem that aims at minimizing the market social cost consisting of the generation cost, the utility of dispatchable loads, as well as the CVaR-based transaction cost. The SAA method is introduced to bypass the inherent high-dimensional integral, while an ADMM-based solver is developed to clear the market in a distributed fashion. Extensive tests on a modified WECC system corroborated the effectiveness of the novel approach, which offers risk-limiting dispatch with considerably reduced conventional generation. The material in this chapter draws from [163, 168].

Chapter 5 turns attention to robust resource allocation tasks for wireless networks. Two beamforming schemes are first introduced for underlay MIMO CR systems in the presence of uncertain CR-to-PU propagation channels. Robust interference constraints are derived by employing a norm-bounded channel uncertainty model, which captures errors in the channel estimation phase, or, random fading effects around the deterministic path loss. Accordingly, a robust beamforming design approach is formulated to

minimize the total MSE in the information symbol reconstruction, while ensuring protection of the primary system. In order to solve the formulated non-convex optimization problem, a cyclic block coordinate ascent algorithm is developed, and its convergence to a stationary point is established when all CR-to-CR direct channel matrices have full column rank. A second algorithm based on a proximal point regularization technique is also developed. Although slower than the first, the proximal point-based scheme is shown capable of converging to a stationary point even for rank-deficient channel matrices. The two solutions offer complementary strengths as far as convergence rate, computational complexity, and MSE optimality are concerned. They can both afford on-line distributed implementations. In addition, the primal decomposition is employed to approximately solve the robust beamforming problem with coupled interference constraints. The developed centralized and distributed algorithms are also suitable for non-CR MIMO ad-hoc networks as well as for conventional downlink or uplink multi-antenna cellular systems.

Finally, robust ahead-of-time energy management and transmit-beamforming designs are developed for the CoMP downlink powered by a grid with smart-meter based dynamic pricing and RES available at the BSs. The task goal is to minimize the worst-case energy transaction cost subject to the worst-case user QoS guarantees. Relying on semidefinite relaxation and dual decomposition techniques, efficient decentralized algorithms are developed to obtain the optimal solutions. The proposed scheme provides the offline ahead-of-time beamformer and energy schedules over a finite time horizon. The material in this chapter draws from [146, 160, 161, 170].

6.2 Future Directions

A number of interesting research directions open up towards extending the frameworks presented in this thesis, which will be briefly discussed in the following subsections.

6.2.1 Stochastic and Robust Power System Operations

• **Unit Commitment and AC-OPF.** The robust and risk-limiting frameworks and approaches for the energy management tasks can be naturally extended to the problems of UC and AC-OPF that are worth re-investigating given the growth of RES usage in

microgrids. Accounting for the uncertainties of non-dispatchable loads and transaction prices will make these tasks more challenging to cope with. Lagrangian relaxation and Benders decomposition will be adapted for solving the mix-integer UC problem efficiently. The intrinsic non-convexity of the OPF task can be dealt with by resorting to the recent progress in conic relaxations. Conditions for achieving global optimality have to be established for the new problem structures. Moreover, it is desired to develop efficient distributed algorithms for stochastic UC and AC-OPF in large-scale power systems.

Regarding the risk-constrained energy management, the distributed implementation should be re-investigated since the local control problems and message exchange mechanism will arise from the sampling and constraint-combining procedure. Furthermore, the distribution-free sample bound may end up improving the wind power output by utilizing the information of the wind speed distribution and the mapping relationship from wind speed to wind power. For the islanded microgrid, changes in power demand cause changes in the frequency and voltage levels. Therefore, frequency regulation is important to maintain system stability. The economic dispatch model proposed in this thesis could be improved by incorporating a frequency control method.

Yet another important direction is the real-time power dispatch. The various options include online convex optimization, Markov decision processes, and economic model predictive control; see e.g. [101,135]. Meanwhile, efforts will be made in pursuit of scalable and privacy-preserving approaches that are operational even in possibly adversarial environments.

• **Nonconvex market clearing and transmission switching.** In Chapter 4, the electricity market clearing is based on the DC-OPF model that many ISOs currently adopt. However, certain advantages have been identified by clearing the market using the UC model with start-up (-down) and no-load costs [60,143]. This nonconvex MC results in the so-termed convex hull pricing (CHP) or extended LMP that essentially minimizes the uplift payment. In addition, the system enjoys another degree of freedom from controlling the topology of the transmission network. More economical dispatch can thus be obtained by switching certain transmission lines. Clearly, both the CHP and topology control problems involve combinatorial complexity, which means that they are intrinsically hard to solve. Extensive research efforts are needed here, especially with

the vision of high-penetration renewables and large-scale dispensable loads. Additional topics worth further investigation include congestion management, reserve procurement, as well as security assessment issues.

6.2.2 Big Data Analytics for Future Power Grids

Current grid monitoring systems have to deal continuously with a large pool of data samples at different time scales. There is no denying that the data size of the future grid will be even more massive with the advent of phasor measurement units and DR rapidly gaining popularity worldwide. The value of global utility data analytics is pegged at a cumulative $20 billion from 2013 to 2020. It now becomes imperative to delineate theoretical underpinnings and develop efficient approaches to extract "wisdom" from massive and streaming energy data.

• **Low-rank and sparse inference of market prices, renewable energy, and load demand.** Accurate inference of market prices, renewable generation, and load demand is of paramount importance for grid economic and reliable operations. The tasks involved here are basically interpolation and extrapolation of the missing and future values. Statical inference via low-rank and sparse models is particularly attractive because the data of interest typically have periodic and repetitive patterns, while deviations from nominal operation introduce sparsely appearing components. Initial efforts have been made towards this direction, e.g., multi-kernel learning approaches for predicting day-ahead LMPs [76], and nonnegative sparse coding for wind power inference [169]. New research effort should aim to construct appropriate kernels with good feature selection, and develop efficient online algorithms with streaming data. Graph-based learning and deep learning are also promising approaches towards the goal of achieving improved forecast accuracy.

• **Online energy disaggregation.** As a task of decomposing a whole-home energy consumption into device-level signals, energy disaggregation techniques are useful to support DR programs, make optimal incentive strategies, and detect energy fraud and theft. Online disaggregation via online convex optimization and switched linear systems is a promising direction. Leveraging the model of low rank plus sparse matrix or tensor factorization, new methods are expected to outperform existing approaches in terms of

prediction accuracy and computational efficiency.

• **Anomaly detection.** Economic operations and robust control of power systems critically hinge on accurate data. Since phasor measurement units are prone to false data injection attacks, anomaly detection becomes imperative for enhancing smart grid cybersecurity. Robust statistical learning methods, e.g., robust regression and outlier-aware principal component analysis, are suited to the task.

Applications of energy data analytics also include predictive modeling for building automation and wide-area monitoring, dynamic pricing of charging plug-in vehicles, as well as strategic asset management. More importantly, the privacy issue and unique features and structures in energy data have to be taken into account.

6.2.3 Stochastic Resource Allocation of Wireless Networks

Infrastructure of both communication networks and electricity grids is now on the verge of major paradigm changes. These changes include joint base station selection and coordinated beamforming designs to achieve energy efficiency in a robust manner for multi-cell heterogeneous networks with renewable-sourced power. Joint optimization of UC, DR and storage sizing will also be investigated for green wireless communication systems. Distributed, robust, and real-time designs with guaranteed convergence will be interesting directions to pursue in future works.

Research thrusts will be also directed towards cross-fertilizing the framework and methodology from wireless networks to data centers, where there is also a growing demand for integrating renewables. Such a trend is driven by the considerably high energy consumption of data centers, and the corporate social responsibility for carbon neutrality. It is clear that substantial cost savings and major impact can be achieved through data center efficiency initiatives.

References

[1] Global energy forecasting competition 2012 – wind forecasting. [Online]. Available: http://www.kaggle.com/c/GEF2012-wind-forecasting

[2] "20% wind energy by 2030: Increasing wind energy's contribution to U.S. electricity supply," U.S. Dept. Energy, Tech. Rep., July 2008, [Online]. Available: http://www1.eere.energy.gov/wind/pdfs/41869.pdf.

[3] "EU energy policy to 2050 – achieving 80-95% emissions reductions," European Wind Energy Association, Tech. Rep., Mar. 2011.

[4] "Assessment of demand response and advanced metering," Federal Energy Regulatory Commission, Tech. Rep., Dec. 2012, [Online]. Available: http://www.ferc.gov/legal/staff-reports/12-20-12-demand-response.pdf.

[5] "MOSEK ApS," 2014. [Online]. Available: http://www.mosek.com/.

[6] T. Al-Khasib, M. Shenouda, and L. Lampee, "Dynamic spectrum management for multiple-antenna cognitive radio systems: Designs with imperfect CSI," *IEEE Trans. Wireless Commun.*, vol. 10, no. 9, pp. 2850–2859, Sept. 2011.

[7] V. Alimisis and N. Hatziargyriou, "Evaluation of a hybrid power plant comprising used EV-batteries to complement wind power," *IEEE Trans. Sustain. Energy*, vol. 4, no. 2, pp. 286–293, Apr. 2013.

[8] N. Amjady, J. Aghaei, and H. A. Shayanfar, "Stochastic multiobjective market clearing of joint energy and reserves auctions ensuring power system security," *IEEE Trans. Power Syst.*, vol. 24, no. 4, pp. 1841–1854, Nov. 2009.

[9] N. Amjady, A. A. Rashidi, and H. Zareipour, "Stochastic security-constrained joint market clearing for energy and reserves auctions considering uncertainties of wind power producers and unreliable equipment," *Int. Trans. Electr. Energ. Syst.*, vol. 23, pp. 451–472, May 2013.

[10] G. N. Bathurst, J. Weatherhill, and G. Strbac, "Trading wind generation in short term energy markets," *IEEE Trans. Power Syst.*, vol. 17, no. 3, pp. 782–789, Aug. 2002.

[11] D. P. Bertsekas, "Projected Newton methods for optimization problems with simple constraints," *SIAM J. Control and Optim.*, vol. 20, no. 2, pp. 221–246, Mar. 1982.

[12] ——, *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 1999.

[13] ——, *Convex Optimization Theory.* Belmont, MA: Athena Scientific, 2009.

[14] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods.* Belmont, MA: Athena Scientific, 1997.

[15] D. Bertsimas, E. Litvinov, X. Sun, J. Zhao, and T. Zheng, "Adaptive robust optimization for the security constrained unit commitment problem," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 52–63, Feb. 2013.

[16] D. Bertsimas and J. N. Tsitsiklis, *Introduction to Linear Optimization.* Belmont, MA: Athena Scientific, 1997.

[17] D. Bienstock, M. Chertkov, and S. Harnett, "Chance constrained optimal power flow: Risk-aware network control under uncertainty," Feb. 2013, [Online]. Avaialble: http://arxiv.org/pdf/1209.5779.pdf.

[18] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Trans. Sig. Proc.*, vol. 54, no. 3, pp. 884–893, Mar. 2006.

[19] J. R. Birge and F. Louveaux, *Introduction to Stochastic Programming.* New York: Springer, 1997.

[20] E. Bitar, R. Rajagopal, P. P. Khargonekar, K. Poolla, and P. Varaiya, "Bringing wind energy to market," *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1225–1235, Aug. 2012.

[21] A. Botterud, J. Wang, C. Monteiro, and V. Miranda, "Wind power forecasting and electricity market operations," in *Proc. of IAEE Intl. Conf.*, San Francisco, CA, June 2009.

[22] A. Botterud, Z. Zhou, J. Wang, R. Bessa, H. Keko, J. Mendes, J. Sumaili, and V. Miranda, "Use of wind power forecasting in operational decisions," Argonne National Laboratory, Tech. Rep., Sept. 2011, [Online]. Available: http://www.dis.anl.gov/pubs/71389.pdf.

[23] A. Botterud, Z. Zhou, J. Wang, R. J. Bessa, H. Keko, J. Sumaili, and V. Miranda, "Wind power trading under uncertainty in LMP markets," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 894–903, May 2012.

[24] F. Bouffard, F. D. Galiana, and A. J. Conejo, "Market-clearing with stochastic security – Part I: Formulation," *IEEE Trans. Power Syst.*, vol. 20, no. 4, pp. 1818–1826, Nov. 2005.

[25] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[26] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[27] S. Bu, F. Yu, Y. Cai, and X. Liu, "When the smart grid meets energy-efficient communications: Green wireless cellular networks powered by the smart grid," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 3014–3024, Aug. 2012.

[28] G. Calafiore and M. Campi, "The scenario approach to robust control design," *IEEE Trans. Automat. Contr.*, vol. 51, pp. 742–753, 2006.

[29] C. A. Canizares and S. Kodsi, "Power system security in market clearing and dispatch mechanisms," in *Proc. of IEEE PES General Meeting*, Montreal, Canada, June 2006.

[30] J. A. Carta, P. Ramírez, and S. Velázquez, "A review of wind speed probability distributions used in wind energy analysis: Case studies in the Canary islands," *Renew. Sust. Energ. Rev.*, vol. 13, pp. 933–955, 2009.

[31] K. M. Chandy, S. H. Low, U. Topcu, and H. Xu, "A simple optimal power flow model with energy storage," in *Proc. of the 49th IEEE Conf. on Decision and Control*, Atlanta, GA, Dec. 2010.

[32] T.-H. Chang, M. Alizadeh, and A. Scaglione, "Coordinated home energy management for real-time power balancing," in *Proc. of IEEE PES General Meeting*, San Diego, CA, July 2012.

[33] L. Chen, N. Li, S. H. Low, and J. C. Doyle, "Two market models for demand response in power networks," in *Proc. of Intl. Conf. Smart Grid Commun.*, Gaithersburg, MD, Oct. 2010, pp. 397–402.

[34] Y. Chen and A. D. Domínguez-García, "A method to study the effect of renewable resource variability on power system dynamics," *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 1978–1989, Nov. 2012.

[35] M. Chiang, S. H. Low, R. Calderbank, and J. Doyle, "Layering as optimization decomposition: A mathematical theory of network architectures," *Proc. IEEE*, vol. 95, no. 1, pp. 255–312, Jan. 2007.

[36] S. S. Christensen, R. Agarwal, E. de Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.

[37] R. D. Christie, B. F. Wollenberg, and I. Wangensteen, "Transmission management in the deregulated environment," *Proc. of the IEEE*, vol. 88, no. 2, pp. 170–195, Feb. 2000.

[38] A. J. Conejo, M. Carrión, and J. M. Morales, *Decision Making Under Uncertainty in Electricity Markets*. New York, Dordrecht, Heidelberg, London: Springer, 2010.

[39] A. J. Conejo, E. Castillo, R. Mínguez, and R. García-Bertrand, *Decomposition Techniques in Mathematical Programming: Engineering and Science Applications*. Springer, 2006.

[40] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1748–1759, May 2010.

[41] E. Dall'Anese, S.-J. Kim, and G. B. Giannakis, "Channel gain map tracking via distributed kriging," *IEEE Trans. Veh. Technol.*, vol. 60, no. 3, pp. 1205–1211, Mar. 2011.

[42] E. Dall'Anese, S.-J. Kim, G. B. Giannakis, and S. Pupolin, "Power control for cognitive radio networks under channel uncertainty," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3541–3551, Oct. 2011.

[43] C. Davis, "All convex invariant functions of Hermitian matrices," *Arch. Math.*, vol. 8, no. 4, pp. 276–278, Feb. 1957.

[44] X. de Luna and M. G. Genton, "Predictive spatio-temporal models for spatially sparse environmental data," *Stat. Sinica*, vol. 15, pp. 547–568, 2005.

[45] M. Ding and S. D. Blostein, "MIMO minimum total MSE transceiver design with imperfect CSI at both ends," *IEEE Trans. Sig. Proc.*, vol. 57, no. 3, pp. 1141–1150, Mar. 2009.

[46] A. D. Domínguez-García and C. N. Hadjicostis, "Distributed algorithms for control of demand response and distributed energy resources," in *Proc. of IEEE Conf. on Decision and Control and Euro. Control Conf.*, Orlando, FL, Dec. 2011.

[47] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, no. 1–3, pp. 293–318, 1992.

[48] F. J. Fabozzi, P. N. Kolm, D. A. Pachamanova, and S. M. Focardi, *Robust Portfolio Optimization and Management*. Hoboken, NJ: Wiley, 2007.

[49] U. Faigle, W. Kern, and G. Still, *Algorithmic Principles of Mathematical Programming*. Norwell, MA: Kluwer Academic Publishers, 2002.

[50] Federal Energy Regulatory Commission, "MISO daily report," Feb. 2012. [Online]. Available: http://www.ferc.gov/market-oversight/mkt-electric/midwest/miso-rto-dly-rpt.pdf

[51] S. Feltenmark and K. C. Kiwiel, "Dual application of proximal bundle methods, including Lagrange relaxation of nonconvex problems," *SIAM J. Optim.*, vol. 10, no. 3, pp. 697–721, Feb./Mar. 2000.

[52] J. Font-Segura and X. Wang, "GLRT-based spectrum sensing for cognitive radio with prior information," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 2137–2146, July 2010.

[53] N. Gatsis and G. B. Giannakis, "Residential load control: Distributed scheduling and convergence with lost AMI messages," *IEEE Trans. Smart Grid*, vol. 3, no. 2, pp. 770–786, June 2012.

[54] ——, "Decomposition algorithms for market clearing with large-scale demand response," *IEEE Trans. Smart Grid*, vol. 4, no. 4, pp. 1976–1987, Dec. 2013.

[55] E. A. Gharavol, Y.-C. Liang, and K. Mouthaan, "Robust downlink beamforming in multiuser MISO cognitive radio networks with imperfect channel-state information," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 2852–2860, July 2010.

[56] ——, "Robust linear transceiver design in MIMO ad hoc cognitive radio networks with imperfect channel state information," *IEEE Trans. Wireless Commun.*, vol. 10, no. 5, pp. 1448–1457, May 2011.

[57] G. B. Giannakis, V. Kekatos, N. Gatsis, S. Kim, H. Zhu, and B. Wollenberg, "Monitoring and optimization for power grids: A signal processing perspective," *IEEE Signal Process. Mag.*, vol. 30, no. 5, pp. 107–128, Sept. 2013.

[58] A. Gómez-Expósito, A. J. Conejo, and C. Canizares, Eds., *Electric Energy Systems, Analysis and Operation*. Boca Raton, FL: CRC Press, 2009.

[59] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[60] P. R. Gribik, W. W. Hogan, and S. L. Pope, "Market-clearing electricity prices and energy uplift," Dec. 2007, [Online]. Available: http://www.hks.harvard.edu/fs/whogan/Gribik_Hogan_Pope_Price_Uplift_123107.pdf.

[61] L. Grippo and M. Sciandrone, "On the convergence of the block nonlinear Gauss–Seidel method under convex constraints," *Operations Research Letters*, vol. 26, no. 3, pp. 127–136, Apr. 2000.

[62] X. Guan, Z. Xu, and Q.-S. Jia, "Energy-efficient buildings facilitated by microgrid," *IEEE Trans. Smart Grid*, vol. 1, no. 3, pp. 243–252, Dec. 2010.

[63] Gurobi Optimization, Inc., "Gurobi optimizer reference manual," 2013. [Online]. Available: http://www.gurobi.com

[64] GWEC, "Global wind statistics 2014," Feb. 2015, [Online]. Available: http://www.gwec.net/wp-content/uploads/2015/02/GWEC_GlobalWindStats2014_FINAL_10.2.2015.pdf.

[65] E. Hasan, F. D. Galiana, and A. J. Conejo, "Electricity markets cleared by merit order – Part I: Finding the market outcomes supported by pure strategy Nash equilibria," *IEEE Trans. Power Syst.*, vol. 23, no. 2, pp. 361–371, May 2008.

[66] N. Hatziargyriou, H. Asano, R. Iravani, and C. Marnay, "Microgrids: An overview of ongoing research, development, and demonstration projects," *IEEE Power & Energy Mag.*, vol. 5, no. 4, pp. 78–94, July/Aug. 2007.

[67] J. Hetzer, C. Yu, and K. Bhattarai, "An economic dispatch model incorporating wind power," *IEEE Trans. Energy Convers.*, vol. 23, no. 2, pp. 603–611, June 2008.

[68] J. B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*. Berlin · Heidelberg · New York: Springer-Verlag, 1993, vol. II.

[69] M. Hong, R. Sun, H. Baligh, and Z. Luo, "Joint base station clustering and beamformer design for partial coordinated transmission in heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 226–240, Feb. 2013.

[70] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *Am. Stat.*, vol. 58, no. 1, pp. 30–37, Feb. 2004.

[71] I. Hwang, B. Song, and S. Soliman, "A holistic view on hyper-dense heterogeneous and small cell networks," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 20–27, June 2013.

[72] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, Feb. 2011.

[73] N. Jaleeli, L. S. VanSlyck, D. H. Ewart, and L. H. Fink, "Understanding automatic generation control," *IEEE Trans. Power Syst.*, vol. 7, no. 3, pp. 1106–1122, Aug. 1992.

[74] L. Jiang and S. H. Low, "Real-time demand response with uncertain renewable energy in smart grid," in *Proc. of Allerton Conf. on Communication, Control, and Computing*, Monticello, IL, Sept. 2011.

[75] C. Jin and P. K. Ghosh, "Coordinated usage of distributed sources for energy cost saving in micro-grid," in *Proc. of North American Power Symp.*, Boston, MA, Aug. 2011.

[76] V. Kekatos, Y. Zhang, and G. B. Giannakis, "Electricity market forecasting via low-rank multi-kernel learning," *IEEE J. Sel. Topics Sig. Proc.*, vol. 8, no. 6, pp. 1182–1193, Dec. 2014.

[77] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 6, pp. 1211–1226, Aug. 2002.

[78] S.-J. Kim, E. Dall'Anese, and G. B. Giannakis, "Cooperative spectrum sensing for cognitive radios using kriged Kalman filtering," *IEEE J. Sel. Topics Sig. Proc.*, vol. 5, no. 1, pp. 24–36, Feb. 2011.

[79] S.-J. Kim and G. B. Giannakis, "Optimal resource allocation for MIMO ad hoc cognitive radio networks," *IEEE Trans. Info. Theory*, vol. 57, no. 5, pp. 3117–3131, May 2011.

[80] ——, "Scalable and robust demand response with mixed-integer constraints," *IEEE Trans. Smart Grid*, vol. 4, no. 4, pp. 2089–2099, Dec. 2013.

[81] S.-J. Kim, S. Jain, and G. B. Giannakis, "Backhaul-constrained multi-cell cooperation using compressive sensing and spectral clustering," in *Proc. of 13th Wrkshp. on Signal Process. Advances in Wireless Commun.*, Cesme, Turkey, June 2012.

[82] D. Kirschen and G. Strbac, *Power System Economics*. West Sussex, England: Wiley, 2010.

[83] K. C. Kiwiel, "Approximation in proximal bundle methods and decomposition of convex programs," *J. Optim. and Appl.*, vol. 84, no. 3, pp. 529–548, Mar. 1995.

[84] ——, "Convergence of approximate and incremental subgradient methods for convex optimization," *SIAM J. Optim.*, vol. 14, no. 3, pp. 807–840, 2004.

[85] A. J. Kleywegt, A. Shapiro, and T. Homem-De-Mello, "The sample average approximation method for stochastic discrete optimization," *SIAM J. Optim.*, vol. 12, no. 2, pp. 479–502, 2001.

[86] M. Kraning, E. Chu, J. Lavaei, and S. Boyd, "Dynamic network energy management via proximal message passing," *Found. Trends Optim.*, vol. 1, no. 2, pp. 70–122, Jan. 2014.

[87] E. H. Lieb, "Convex trace functions and the Wigner-Yanase-Dyson conjecture," *Advances in Math.*, vol. 11, no. 3, pp. 267–288, Dec. 1973.

[88] X. Liu, "Economic load dispatch constrained by wind power availability: A wait-and-see approach," *IEEE Trans. Smart Grid*, vol. 1, no. 3, pp. 347–355, Dec. 2010.

[89] X. Liu and W. Xu, "Economic load dispatch constrained by wind power availability: A here-and-now approach," *IEEE Trans. Sustain. Energy*, vol. 1, no. 1, pp. 2–9, Apr. 2010.

[90] J. Löfberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," in *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004. [Online]. Available: http://users.isy.liu.se/johanl/yalmip

[91] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Economics*, 2nd ed. New York: Wiley, 1999.

[92] Midcontinent Independent System Operator, Inc., "MISO market data." [Online]. Available: https://www.misoenergy.org/MarketsOperations/Pages/MarketsOperations.aspx

[93] ——, "MISO wind data." [Online]. Available: https://www.misoenergy.org/MarketsOperations/RealTimeMarketData/Pages/DayAheadWindForecast.aspx

[94] L. Mirsky, "On the trace of matrix products," *Mathematische Nachrichten*, vol. 20, no. 3–6, pp. 171–174, 1959.

[95] A.-H. Mohsenian-Rad, V. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia, "Autonomous demand side management based on game-theoretic energy consumption scheduling for the future smart grid," *IEEE Trans. Smart Grid*, vol. 1, no. 3, pp. 320–331, Dec. 2010.

[96] J. M. Morales, L. Baringo, A. J. Conejo, and R. Mínguez, "Probabilistic power flow with correlated wind sources," *IET Generation, Transmission & Distribution*, vol. 4, no. 5, pp. 641–651, May 2010.

[97] J. M. Morales, A. J. Conejo, K. Liu, and J. Zhong, "Pricing electricity in pools with wind producers," *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1366–1376, Aug. 2012.

[98] J. M. Morales, M. Zugno, S. Pineda, and P. Pinson, "Electricity market clearing with improved scheduling of stochastic production," *Eur. J. Oper. Res.*, vol. 235, no. 3, pp. 765–774, June 2014.

[99] ——, "Redefining the merit order of stochastic generation in forward markets," *IEEE Trans. Power Syst.*, vol. 29, no. 2, pp. 992–993, Mar. 2014.

[100] J. M. Morales, A. J. Conejo, and J. Pérez-Ruiz, "Short-term trading for a wind power producer," *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 554–564, Feb. 2010.

[101] B. Narayanaswamy, V. K. Garg, and T. S. Jayram, "Online optimization for the smart (micro) grid," in *Proc. of ACM e-Energy*, Madrid, Spain, May 2012.

[102] A. Nedić and A. Ozdaglar, "Approximate primal solutions and rate analysis for dual subgradient methods," *SIAM J. Optim.*, vol. 19, no. 4, pp. 1757–1780, 2009.

[103] A. Nemirovski and A. Shapiro, "Convex approximations of chance constrained programs," *SIAM J. Optim.*, vol. 17, no. 4, pp. 969–996, Nov. 2006.

[104] C. Ng and H. Huang, "Linear precoding in cooperative MIMO cellular networks with limited coordination clusters," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1446–1454, Dec. 2010.

[105] K.-L. Noh, E. Serpedin, and K. Qaraqe, "A new approach for time synchronization in wireless sensor networks: Pairwise broadcast synchronization," *IEEE Trans. Wireless Commun.*, vol. 7, no. 9, pp. 3318–3322, Sept. 2008.

[106] E. Oh, B. Krishnamachari, X. Liu, and Z. Niu, "Toward dynamic energy-efficient operation of cellular network infrastructure," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 56–61, June 2011.

[107] F. Paganini, P. Belzarena, and P. Monzón, "Decision making in forward power markets with supply and demand uncertainty," in *Proc. of Conf. Info. Sci. and Syst.*, Princeton, NJ, Mar. 2014.

[108] D. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 46, no. 8, pp. 1439–1451, Aug. 2006.

[109] D. P. Palomar, "Convex primal decomposition for multicarrier linear MIMO transceivers," *IEEE Trans. Sig. Proc.*, vol. 53, no. 12, pp. 4661–4674, Dec. 2005.

[110] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Sig. Proc.*, vol. 51, no. 9, pp. 2381–2401, Sept. 2003.

[111] A. Papavasiliou and S. Oren, "Supplying renewable energy to deferrable loads: Algorithms and economic analysis," in *Proc. of IEEE PES General Meeting*, Minneapolis, MN, July 2010.

[112] P. Pinson and G. Kariniotakis, "Conditional prediction intervals of wind power generation," *IEEE Trans. Power Syst.*, vol. 25, no. 4, pp. 1845–1856, Nov. 2010.

[113] G. Pritchard, G. Zakeri, and A. Philpott, "A single-settlement, energy-only electric power market for unpredictable and intermittent participants," *Oper. Res.*, vol. 58, no. 4, pp. 1210–1219, July 2010.

[114] P. Ranci and G. Cervigni, *The Economics of Electricity Markets: Theory and Policy*. Northampton, MA: Edward Elgar Publishing, 2013.

[115] J. Rao and A. Fapojuwo, "On the tradeoff between spectral efficiency and energy efficiency of homogeneous cellular networks with outage constraint," *IEEE Tans. Veh. Technol.*, vol. 62, no. 4, pp. 1801–1814, May 2013.

[116] M. Razaviyayn, M. Sanjabi, and Z.-Q. Luo, "Linear transceiver design for interference alignment: Complexity and computation," *IEEE Trans. Info. Theory*, vol. 58, no. 5, pp. 2896–2910, May 2012.

[117] C. H. Rentel and T. Kunz, "A mutual network synchronization method for wireless ad hoc and sensor networks," *IEEE Trans. Mobile Comput.*, vol. 7, no. 5, pp. 633–646, May 2008.

[118] R. T. Rockafellar, "Augmented Lagrangians and applications of the proximal point algorithms in convex programming," *Math. Oper. Res.*, vol. 1, no. 2, pp. 97–116, May 1976.

[119] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *J. Risk*, vol. 2, no. 3, pp. 21–41, 2000.

[120] ——, "Conditional value-at-risk for general loss distributions," *J. of Banking Finance*, vol. 26, pp. 1443–1471, 2002.

[121] J. Rogers and K. Porter, "Wind power and electricity markets," Oct. 2011, [Online]. Available: http://variablegen.org/wp-content/uploads/2012/11/windinmarketstableOct2011.pdf.

[122] P. W. Sauer and M. A. Pai, *Power System Dynamics and Stability*. Upper Saddle River, NJ: Prentice Hall, 1998.

[123] G. Scutari and D. P. Palomar, "MIMO cognitive radio: A game-theoretical approach," *IEEE Trans. Sig. Proc.*, vol. 58, no. 2, pp. 761–780, Feb. 2010.

[124] M. Shahidehpour, H. Yamin, and Z. Li, *Market Operations in Electric Power Systems*. New York, NY: John Wiley, 2002.

[125] M. Shenouda and T. N. Davidson, "On the design of linear transceivers for multiuser systems with channel uncertainty," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 6, pp. 1015–1024, Aug. 2008.

[126] Y. Shi, J. Zhang, and K. Letaief, "Group sparse beamforming for green cloud-RAN," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2809–2823, May 2014.

[127] E. Sjödin, D. F. Gayme, and U. Topcu, "Risk-mitigated optimal power flow for wind powered grids," in *Proc. of American Control Conf.*, Montréal, Canada, June 2012.

[128] E. Song, Q. Shi, M. Sanjabi, R. Sun, and Z.-Q. Luo, "Robust SINR-constrained MISO downlink beamforming: When is semidefinite programming relaxation tight?" *EURASIP J. Wireless Commun. Netw.*, vol. 2012, pp. 1–11, Feb. 2012.

[129] P. Stluka, D. Godbole, and T. Samad, "Energy management for buildings and microgrids," in *Proc. of IEEE Conf. on Decision and Control and Euro. Control Conf.*, Orlando, FL, Dec. 2011.

[130] S. Stoft, *Power System Economics: Designing Markets for Electricity*. New York, NY: Wiley-IEEE Press, 2002.

[131] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optim. Methods Softw.*, vol. 11–12, pp. 625–653, Aug. 1999.

[132] C. W. Tan, M. Chiang, and R. Srikant, "Maximizing sum rate and minimizing MSE on multiuser downlink: Optimality, fast algorithms and equivalence via max-min SINR," *IEEE Trans. Sig. Proc.*, vol. 59, no. 12, pp. 6127–6143, Dec. 2011.

[133] K. C. Toh, M. J. Todd, and R. H. Tutuncu, "SDPT3 – a matlab software package for semidefinite programming," *Optim. Method Softw.*, vol. 11, pp. 545–581, 1999. [Online]. Available: http://www.math.nus.edu.sg/~mattohkc/sdpt3.html

[134] A. Tolli, H. Pennanen, and P. Komulainen, "Decentralized minimum power multi-cell beamforming with limited backhaul signaling," *IEEE Trans. Wireless Commun.*, vol. 10, no. 2, pp. 570–580, Feb. 2011.

[135] M. Trifkovic, W. Marvin, P. Daoutidis, and M. Sheikhzadeh, "Dynamic real-time optimization and control of a hybrid energy system," *AIChE J.*, vol. 60, no. 7, pp. 2546–2556, July 2014.

[136] M. Trifkovic, M. Sheikhzadeh, K. Nigim, and P. Daoutidis, "Modeling and control of a renewable hybrid energy system with hydrogen storage," *IEEE Trans. Contr. Syst. Technol.*, vol. 22, no. 1, pp. 169–179, Jan. 2014.

[137] P. Tseng, "Further results on approximating nonconvex quadratic optimization by semidefinite programming relaxation," *SIAM J. Optim.*, vol. 14, no. 1, pp. 268–283, July 2003.

[138] Power systems test case archive. University of Washington. [Online]. Available: http://www.ee.washington.edu/research/pstca/

[139] D. Villanueva, A. Feijóo, and J. L. Pazos, "Simulation of correlated wind speed data for economic dispatch evaluation," *IEEE Trans. Sustain. Energy*, vol. 3, no. 1, pp. 142–149, Jan. 2012.

[140] N. Vucic and H. Boche, "Robust QoS-constrained optimization of downlink multiuser MISO systems," *IEEE Trans. Sig. Proc.*, vol. 57, no. 2, pp. 714–725, June 2009.

[141] N. Vucic, H. Boche, and S. Shi, "Robust transceiver optimization in downlink multiuser MIMO systems," *IEEE Trans. Sig. Proc.*, vol. 57, no. 9, pp. 3576–3587, Sept. 2009.

[142] P. Vytelingum, T. D. Voice, S. D. Ramchurn, A. Rogers, and N. R. Jennings, "Agent-based micro-storage management for the smart grid," in *Proc. 9th Intl. Conf. Autonomous and Multiagent Systems*, Toronto, Canada, May 2010, pp. 39–46.

[143] G. Wang, U. V. Shanbhag, T. Zheng, E. Litvinov, and S. Meyn, "An extreme-point subdifferential method for convex hull pricing in energy and reserve markets – Part I: Algorithm structure," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2111–2120, Aug. 2013.

[144] J. Wang, G. Scutari, and D. P. Palomar, "Robust MIMO cognitive radio via game theory," *IEEE Trans. Sig. Proc.*, vol. 59, no. 3, pp. 1183–1201, Mar. 2011.

[145] X. Wang, G. B. Giannakis, and A. G. Marques, "A unified approach to QoS-guaranteed scheduling for channel-adaptive wireless networks," *Proc. IEEE*, vol. 95, no. 12, pp. 2410–2431, Dec. 2007.

[146] X. Wang, Y. Zhang, G. B. Giannakis, and S. Hu, "Robust smart-grid powered cooperative multipoint systems," *IEEE Trans. Wireless Commun.*, June 2015 (accepted).

[147] C. Wu, H. Mohsenian-Rad, J. Huang, and Y. Wang, "Demand side management for wind power integration in microgrid using dynamic potential game theory," in *Proc. of 2011 IEEE Globecom Wksps.*, Houston, TX, Dec. 2011.

[148] L. Xiao and S. Boyd, "Simultaneous routing and resource allocation via dual decomposition," *IEEE Trans. Commun.*, vol. 52, no. 7, pp. 1136–1144, July 2004.

[149] L. Xie, Y. Gu, X. Zhu, and M. G. Genton, "Power system economic dispatch with spatio-temporal wind forecasts," in *Proc. of IEEE EnergyTech*, Cleveland, OH, May 2011.

[150] J. Xu, L. Duan, and R. Zhang, "Cost-aware green cellular networks with energy and communication cooperation," *IEEE Commun. Mag.*, vol. 53, no. 5, pp. 257–263, May 2015.

[151] J. Xu, Y. Guo, and R. Zhang, "CoMP meets energy harvesting: A new communication and energy cooperation paradigm," in *Proc. of Global Commun. Conf.*, Atlanta, GA, Dec. 2013.

[152] J. Xu and R. Zhang, "Cooperative energy trading in CoMP systems powered by smart grids," Mar. 2014. [Online]. Available: http://arxiv.org/pdf/1403.5735v1.pdf

[153] A. L. Yuille and A. Rangarajan, "The concave-convex procedure," *Neural Computation*, vol. 15, no. 4, pp. 915–936, Apr. 2003.

[154] V. Zavala, M. Anitescu, and J. Birge, "A stochastic electricity market clearing formulation with consistent pricing properties," Oper. Res. (submitted) 2014, [Online]. Available: http://www.mcs.anl.gov/~anitescu/PUBLICATIONS/2014/zavala-stochpricing-2014.pdf.

[155] D. Zennaro, E. Dall'Anese, T. Erseghe, and L. Vangelista, "Fast clock synchronization in wireless sensor networks via ADMM-based consensus," in *Proc. of Intl. Sym. on Mod. and Opt. in Mobile, Ad Hoc and Wireless Net.*, Princeton, NJ, May 2011, pp. 148–153.

[156] J. Zhang, R. Chen, J. Andrews, A. Ghosh, and R. W. Heath, "Networked MIMO with clustered linear precoding," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1910–1921, Apr. 2009.

[157] L. Zhang, Y.-C. Liang, and Y. Xin, "Joint beamforming and power allocation for multiple access channels in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 38–51, Jan. 2008.

[158] L. Zhang, Y.-C. Liang, Y. Xin, and H. V. Poor, "Robust cognitive beamforming with partial channel state information," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4143–4153, Aug. 2009.

[159] R. Zhang and Y.-C. Liang, "Exploiting multi-antennas for opportunistic spectrum sharing in cognitive radio networks," *IEEE J. Sel. Topics Sig. Proc.*, vol. 2, no. 1, pp. 88–102, Feb. 2008.

[160] Y. Zhang, E. Dall'Anese, and G. B. Giannakis, "Distributed optimal beamformers for cognitive radios robust to channel uncertainties," *IEEE Trans. Sig. Proc.*, vol. 60, no. 12, pp. 6495–6508, Dec. 2012.

[161] ——, "Distributed robust beamforming for MIMO cognitive networks," in *Proc. of Intl. Conf. on Acoustics, Speech, and Signal Process.*, Kyoto, Japan, Mar. 2012.

[162] Y. Zhang, N. Gatsis, and G. B. Giannakis, "Robust distributed energy management for microgrids with renewables," in *Proc. of Intl. Conf. Smart Grid Commun.*, Tainan, Taiwan, Nov. 2012.

[163] ——, "Disaggregated bundle methods for distributed market clearing in power networks," in *Proc. of Global Conf. on Signal and Info. Process.*, Austin, TX, Dec. 2013.

[164] ——, "Risk-constrained energy management with multiple wind farms," in *Proc. of 4th IEEE Conf. Innovative Smart Grid Tech.*, Washington, D.C., Feb. 2013.

[165] ——, "Robust energy management for microgrids with high-penetration renewables," *IEEE Trans. Sustain. Energy*, vol. 4, no. 4, pp. 944–953, Oct. 2013.

[166] ——, "Robust energy management for microgrids with high-penetration renewables," University of Minnesota, Tech. Rep., May 2013. [Online]. Available: http://arxiv.org/pdf/1207.4831.pdf

[167] Y. Zhang, N. Gatsis, V. Kekatos, and G. B. Giannakis, "Risk-aware management of distributed energy resources," in *Proc. of 18th Intl. Conf. on Digital Signal Process.*, Santorini Island, Greece, July 2013.

[168] Y. Zhang and G. B. Giannakis, "Distributed stochastic market clearing with high-penetration wind power," *IEEE Trans. Power Syst.*, Apr. 2015 (accepted).

[169] Y. Zhang, S.-J. Kim, and G. B. Giannakis, "Short-term wind power forecasting using nonnegative sparse coding," in *Proc. of 49th Conf. on Info. Sci. and Syst.*, Baltimore, MD, Mar. 2015.

[170] Y. Zhang, X. Wang, G. B. Giannakis, and S. Hu, "Distributed robust resource allocation for renewable powered wireless cellular networks," in *Proc. of Black-SeaCom Conf.*, Constanta, Romania, May 2015.

[171] L. Zhao and B. Zeng, "Robust unit commitment problem with demand response and wind energy," University of South Florida, Tech. Rep., Oct. 2010, [Online]. Available: http://www.optimization-online.org/DB_FILE/2010/11/2784.pdf.

[172] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Sig. Proc. Mag.*, vol. 24, no. 3, pp. 79–89, May 2007.

[173] G. Zheng, K.-K. Wong, and B. Ottersten, "Robust cognitive beamforming with bounded channel uncertainties," *IEEE Trans. Sig. Proc.*, vol. 57, no. 12, pp. 4871–4881, Dec. 2009.

[174] X. Zhu and M. G. Genton, "Short-term wind speed forecasting for power system operations," *Intl. Stat. Rev.*, vol. 80, no. 1, pp. 2–23, Apr. 2012.

[175] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "MATPOWER: steady-state operations, planning and analysis tools for power systems research and education," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, Feb. 2011.

Appendix A

Notations

The following notational conventions are adopted in this thesis. Other notations are explicitly specified when necessary.

Table A.1: Notations

Notation	Definition
$\mathbf{A}$	Upper boldface letters represent matrices
$\mathbf{x}$	Lower boldface letters represent column vectors
$\mathcal{S}$	Calligraphic letters stand for sets
$\mathbb{R}$ ($\mathbb{C}$)	Space of real (complex) numbers
$\mathbb{R}^n$ ($\mathbb{C}^n$)	Space of $n \times 1$ real (complex) vectors
$\mathbb{R}^n_+$	n-dimensional non-negative orthant
$\mathbb{R}^{n_1 \times n_2}$ ($\mathbb{C}^{n_1 \times n_2}$)	Space of $n_1 \times n_2$ real (complex) matrices
$\mathbb{H}^{n \times n}$	Space of $n \times n$ Hermitian matrices
$\mathbf{x}'$ and $\mathbf{x}^{\mathcal{H}}$	Transpose and Hermitian transpose of $\mathbf{x}$
$\|\mathbf{x}\|_2$	Euclidean norm of $\mathbf{x}$
$\|\mathbf{A}\|_F$	Frobenius norm of $\mathbf{A}$
$\mathbf{x} \cdot \mathbf{y}$	Inner product of $\mathbf{x}$ and $\mathbf{y}$
$\mathbf{A} \otimes \mathbf{B}$	Kronecker product of $\mathbf{A}$ and $\mathbf{B}$
$\mathrm{Tr}(\mathbf{A})$	Trace of $\mathbf{A}$
$\lfloor \mathcal{I} \rfloor$	Lower endpoint of the interval set $\mathcal{I}$

(continued on next page)

(continued from previous page)

Notation	Definition
$[a]^+ := \max\{a, 0\}$	Projection to the nonnegative reals
$\|a\|$	Absolute (magnitude) value of $a \in \mathbb{R}$ ($a \in \mathbb{C}$)
$\lceil a \rceil$	Smallest integer greater than or equal to $a \in \mathbb{R}$
$\Re(a)$	Real part of $a \in \mathbb{C}$
$\mathbf{x} \preceq \mathbf{y}$	Element-wise inequality between vectors $\mathbf{x}$ and $\mathbf{y}$
$\mathbf{A} \preceq \mathbf{B}$	Matrix $\mathbf{B} - \mathbf{A}$ is positive semidefinite
$\mathbb{E}[\cdot]$	Expectation

Appendix B

Acronyms

A number of acronyms are frequently used throughout this thesis for the sake of conciseness. These acronyms and their meanings are alphabetically listed in the following table.

Table B.1: Acronyms

Acronym	Meaning
ADMM	Alternating Direction Method of Multipliers
BCA	Block Coordinate Ascent
CDF	Cumulative Distribution Function
CLT	Central Limit Theorem
CPM	Cutting Plane Method
CoMP	Coordinated Multipoint
CR	Cognitive Radio
CVaR	Conditional Value-at-Risk
DA	Day-Ahead
DR	Demand Response
DSM	Demand Side Management
ED	Economic Dispatch
i.i.d.	Independent and Identically Distributed
ISO	Independent System Operator

(continued on next page)

(continued from previous page)

Acronym	Meaning
LMI	Linear Matrix Inequality
LMPs	Locational Marginal Prices
LOLP	Loss-of-Load Probability
LP	Linear Program
MC	Market Clearing
MIMO	Multiple-Input Multiple-Output
MO	Market Operator
MSE	Mean Squared Error
OPF	Optimal Power Flow
PDF	Probability Density Function
PHEV	Plug-in Hybrid Electric Vehicle
QoS	Quality-of-Service
QP	Quadratic Program
RES	Renewable Energy Sources
RT	Real-Time
RTO	Regional Transmission Organization
SAA	Sample Average Approximation
SCED	Security-Constrained Economic Dispatch
SCUC	Security-Constrained Unit Commitment
SDP	Semi-Definite Program
SINR	Signal-to-Interference-plus-Noise Ratio
SOLP	Satisfaction-of-Load Probability
UC	Unit Commitment
VaR	Value-at-Risk
w.l.o.g.	Without Loss of Generality
WPPs	Wind Power Producers
w.r.t.	With Respect To