

Washington University in St. Louis Washington University Open Scholarship

All Theses and Dissertations (ETDs)

Spring 4-22-2014

Power System State Estimation and Renewable Energy Optimization in Smart Grids

Peng Yang

Washington University in St. Louis

Follow this and additional works at: <https://openscholarship.wustl.edu/etd>



Part of the [Electrical and Computer Engineering Commons](#)

Recommended Citation

Yang, Peng, "Power System State Estimation and Renewable Energy Optimization in Smart Grids" (2014). *All Theses and Dissertations (ETDs)*. 1272.

<https://openscholarship.wustl.edu/etd/1272>

This Dissertation is brought to you for free and open access by Washington University Open Scholarship. It has been accepted for inclusion in All Theses and Dissertations (ETDs) by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

Washington University in St. Louis
School of Engineering and Applied Science
Department of Electrical & Systems Engineering

Dissertation Examination Committee:
Arye Nehorai, Chair
R. Martin Arthur
I. Norman Katz
Chenyang Lu
Heinz Schaettler
Ami Wiesel

Power System State Estimation and Renewable Energy Optimization in Smart Grids

by

Peng Yang

A dissertation presented to the Graduate School of Arts and Sciences
of Washington University in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

May 2014
Saint Louis, Missouri

© 2014, Peng Yang

Contents

List of Figures	v
List of Tables	vii
Acknowledgments	viii
Abstract	x
1 Introduction	1
1.1 Background	2
1.1.1 Power system state estimation	2
1.1.2 Renewable energy in power grids	4
1.2 Contributions of this work	5
1.3 Organization of the dissertation	7
1.4 Notations	7
2 State Estimation Using PMUs with Phase Mismatch	9
2.1 Introduction	9
2.2 Measurement and system model	11
2.2.1 Measurement model	11
2.2.2 Dynamic models for system state and phase mismatch	17
2.3 State estimation considering phase mismatch	18
2.3.1 State estimation using the static model	19
2.3.2 State estimation using the dynamic model	22
2.3.3 Other possible methods	25
2.4 Numerical examples	28
2.4.1 General setup	29
2.4.2 Example using static models	30
2.4.3 Example using dynamic models	32
2.4.4 Comparison between AM and PKF	34
2.4.5 Effect of PMU specifications	34
2.5 Summary	37
3 Performance Bound and Optimal PMU Placement	38
3.1 Introduction	38

3.2	Performance bound	40
3.3	PMU placement algorithms	46
3.3.1	Problem formulation	46
3.3.2	Greedy algorithm	48
3.3.3	Other placement strategies	52
3.4	Numerical examples	53
3.4.1	Comparison of different PMU placements	53
3.4.2	State estimation results under different placement	56
3.5	Summary	57
4	Parallel Load Scheduling with Renewable Distributed Generation	58
4.1	Introduction	58
4.2	Related work	60
4.3	System model	61
4.3.1	Model of users	61
4.3.2	Utility costs and electricity price	65
4.3.3	Optimization of load schedule	68
4.4	Parallel distributed optimization	70
4.4.1	Fixed penalty coefficient	72
4.4.2	Adaptive penalty coefficient	73
4.4.3	Convergence properties	74
4.5	Numerical examples	74
4.5.1	General setup	74
4.5.2	Simulation results	76
4.5.3	Convergence with a fixed penalty coefficient	78
4.5.4	Convergence with an adaptive penalty coefficient	80
4.5.5	Effect of distributed generators	81
4.6	Summary	82
5	Joint Storage and Renewable Generation Capacity Planning	83
5.1	Introduction	83
5.2	Related work	85
5.3	System model	86
5.3.1	Energy storage model	86
5.3.2	Generator model	88
5.3.3	Load balance constraint	89
5.4	Storage and renewable generation planning	90
5.4.1	Optimal planning problem	90
5.4.2	Formulation of consensus problem	92
5.4.3	Distributed optimization	94
5.5	Numerical examples	96
5.5.1	Data and parameters	97

5.5.2	Results of storage and generation planning	99
5.5.3	Results with different diesel generation costs	102
5.5.4	Results with different diesel generation capacities	103
5.5.5	Results with data from different geographic locations	104
5.6	Summary	105
6	Conclusions and Future Work	106
6.1	Summary and conclusions	106
6.2	Future directions	107
	References	109
Appendix A	Derivation of Condition (3.44)	119
Appendix B	Proof of Submodularity	121
Appendix C	Sketch of Convergence Proof	123
Vita	126

List of Figures

1.1	IEEE 14-bus system.	3
2.1	Bus branch model.	12
2.2	Block diagram for state estimation using parallel Kalman filter.	24
2.3	Comparison of state estimation results using AM, LS, and OLS on the IEEE 57-Bus system with 28 PMUs.	31
2.4	Comparison of state estimation results using KF, PKF, and OPK on the IEEE 57-Bus system with 28 PMUs.	33
2.5	RMSE as a function of the number of PMUs installed on the IEEE 57-bus system.	35
2.6	RMSE as a function of PMU synchronization accuracy $\sigma_{t,k}$ on the IEEE 57-bus system with 28 PMUs.	36
2.7	RMSE as a function of the time interval between consecutive synchronization on the IEEE 57-bus system.	36
3.1	Comparison of PMU placement for IEEE 14-bus system using different design criteria and different algorithms.	55
3.2	Comparison of PMU placement for IEEE 30-bus system using different design criteria and different algorithms.	55
3.3	Comparison of state estimation results using different PMU placement profiles on the IEEE 14-bus system.	56
4.1	Two-way power and information flow between the utility and users.	61
4.2	Example of a satisfaction cost function with different parameters.	63
4.3	Example of deterministic and expected cost as a function of the net load. $p_S(t)$ and $p_D(t)$ are set to be 20 cents/kWh and 10 cents/kWh, respectively.	69
4.4	Single user load schedule before and after optimization. Regular users and users with renewable generators are distinguished by R and D, respectively.	76
4.5	Aggregated hourly load to the utility company before and after optimization. Peak-hour load is significantly balanced, and load curve is balanced after optimization.	77
4.6	Reduction in daily bill for randomly selected users before and after optimization. Values calculated based on 200 realizations of randomly drawn DG generation data.	77
4.7	Convergence of parallel optimization with fixed penalty coefficients.	79

4.8	Convergence of parallel optimization with adaptive penalty coefficients. . . .	80
4.9	Effect of prediction error and DG penetration percentage on utility costs. . .	81
5.1	Plots for normalized load, solar generation, and wind generation data. On the left are box plots, and on the right are raw data heat maps for three years. .	100
5.2	Hourly energy shortage distribution. Note that a negative value denotes that generation is greater than demand.	101
5.3	Total cost and percentage of renewable generation as a function of diesel generation capacity ratio.	103
5.4	Geographical locations of the three cities for comparison.	104

List of Tables

2.1	Alternating algorithm for static state estimation	22
2.2	PMU locations for different test systems.	30
2.3	Root mean-squared error (RMSE) of estimation using static models on different test systems, using least-squares estimation (LS), alternating minimization (AM), and LS with perfect phase mismatch information (OLS).	32
2.4	Root mean-squared error (RMSE) of estimation using dynamic models on different test systems, using traditional Kalman filter (KF), parallel Kalman filter (PKF), and Kalman filter with perfect phase mismatch information (OKF).	33
3.1	Greedy algorithm for PMU placement.	48
4.1	Algorithm for utility company.	72
4.2	Algorithm for individual user.	72
4.3	Comparison of computation time.	79
5.1	Parameters for energy generators.	98
5.2	Parameters for energy storage.	98
5.3	Results for $r_{DC} = 0.50$, $r_{SD} = 0.05$ over a three-year design horizon.	101
5.4	Results for different diesel generation costs	102
5.5	Geographic locations and climates.	104
5.6	Comparison of optimization results for different geographic locations.	105

Acknowledgments

I am sincerely grateful to my advisor, Dr. Arye Nehorai, for his mentoring, guidance, and support during my research and life at Washington University. He has provided me a stimulating environment and lots of freedom to pursue the research topics I'm interested in. I also thank him for introducing me to the research area of the smart grid.

I wish to thank my dissertation defense committee members, Dr. R. Martin Arthur, Dr. I. Norman Katz, Dr. Chenyang Lu, Dr. Heinz Schaettler, and Dr. Ami Wiesel, and my preliminary research exam committee member Dr. Hiro Mukai, for their valuable suggestions and comments on improving my dissertation. I also wish to thank my instructors from both the University of Science and Technology of China, and Washington University in St. Louis, for helping me construct a solid background for my research.

I further thank Dr. Ami Wiesel, and my labmates Gongguo, Weifeng, Zhao, Elad and Phani, for collaborating on exciting topics. I'm also thankful to my labmates Patricio, Satya, Murat, Tao, Sandeep, Xiaoxiao, Keyong, Alex, Jichuan, and Mengxue, who have always been inspiring and supportive, and made my experience at Washington University a wonderful memory.

I owe my deepest thanks to my parents, for their endless love and encouragement. All my achievements would not be possible without their continuous support.

Peng Yang

Washington University in Saint Louis
May 2014

Dedicated to my parents.

ABSTRACT OF THE DISSERTATION

Power System State Estimation and Renewable Energy Optimization in Smart Grids

by

Peng Yang

Doctor of Philosophy in Electrical Engineering

Washington University in St. Louis, 2014

Professor Arye Nehorai, Chair

The future smart grid will benefit from real-time monitoring, automated outage management, increased renewable energy penetration, and enhanced consumer involvement. Among the many research areas related to smart grids, this dissertation will focus on two important topics: power system state estimation using phasor measurement units (PMUs), and optimization for renewable energy integration.

In the first topic, we consider power system state estimation using PMUs, when phase angle mismatch exists in the measurements. In particular, we build a measurement model that takes into account the measurement phase angle mismatch. We then propose algorithms to increase state estimation accuracy by taking into account the phase angle mismatch. Based on the proposed measurement model, we derive the posterior Cramér-Rao bound on the estimation error, and propose a method for PMU placement in the grid. Using numerical examples, we show that by considering the phase angle mismatch in the measurements, the estimation accuracy can be significantly improved compared with the traditional weighted least-squares estimator or Kalman filtering. We also show that using the proposed PMU

placement strategy can increase the estimation accuracy by placing a limited number of PMUs in proper locations.

In the second topic, we consider optimization for renewable energy integration in smart grids. We first consider a scenario where individual energy users own on-site renewable generators, and can both purchase and sell electricity to the main grid. Under this setup, we develop a method for parallel load scheduling of different energy users, with the goal of reducing the overall cost to energy users as well as to energy providers. The goal is achieved by finding the optimal load schedule of each individual energy user in a parallel distributed manner, to flatten the overall load of all the energy users. We then consider the case of a micro-grid, or an isolated grid, with a large penetration of renewable energy. In this case, we jointly optimize the energy storage and renewable generator capacity, in order to ensure an uninterrupted power supply with minimum costs. To handle the large dimensionality of the problem due to large historical datasets used, we reformulate the original optimization problem as a consensus problem, and use the alternating direction method of multipliers to solve for the optimal solution in a distributed manner.

Chapter 1

Introduction

Until recently, the power grid in the United States had not changed much (except for expansion in scale) since it was built about 100 years ago. The grid was constructed with enough extra capacity and redundancy to accommodate customer demand and recover from outages. Once constructed, an electric utility can manage the system only by matching generation to consumption. To monitor the grid status, system operators often rely on the traditional supervisory control and data acquisition (SCADA) system, which has a very low sample rate, and does not provide many pieces of important information about the grid state.

With the introduction of the “smart grid” concept [1, 2], there have been significant changes and developments in the electricity power grid. Although there is not an established accurate definition of smart grid, it is well accepted that a smart grid is a power grid modernized with advanced sensor, information, and communication technology. Such modernization enables automated data collection and system operation, in order to increase the grid efficiency and reliability, as well as making energy generation and distribution more economical and sustainable. The smart grid will be different from the power grid today, from at least the following three aspects.

Reliability and security: The smart grid will be equipped with advanced sensors and automatic outage management systems. Advances in both hardware and software will enable real-time system monitoring, which was not practical in the traditional power grid. Such an ability will help reduce system vulnerability to failures and physical or cyber attacks. The smart grid will also minimize consequences of disruptions, prevent cascading failures, and support self-healing.

Economic incentive: The advanced metering infrastructure, energy control systems, and home automation systems will enable more efficient energy management solutions, as well as real-time two-way communication between the utility company and end users. Such development has motivated research on demand-side management, peak load shaving, automatic energy scheduling and management, etc., which will increase energy efficiency and reduce costs to both the utility and end users.

Environmental sustainability: The penetration of renewable energy is expected to increase significantly during the coming years. By replacing traditional forms of generation with renewable energy sources, carbon emissions can be significantly reduced. The future grid also will tend to be decentralized. Decentralized micro-grids have their own generators, and supply most of the energy demand locally, often using renewable generators. This trend will further benefit the environment as it reduces losses from long-distance transmission and energy conversions.

To summarize, the future smart grid will benefit from real-time monitoring, automated outage management, increased renewable energy penetration, and enhanced consumer involvement. Due to its importance, the smart grid has attracted considerable attention since its introduction, and has become an active research area in recent years.

1.1 Background

Among the many research areas related to smart grids, this dissertation will focus on two important topics – power system state estimation and renewable energy integration. In this section, we provide a brief background on these two topics.

1.1.1 Power system state estimation

A electric power system is a networked system that enables generation, transmission, and distribution of electric power. A typical power system consists of multiple sub-stations interconnected by transmission lines, transformers, and other electric devices. Power is

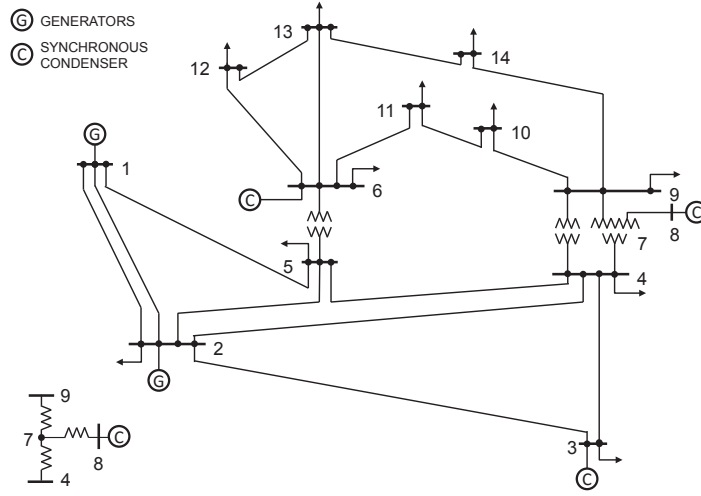


Figure 1.1: IEEE 14-bus system.

injected to the grid and absorbed by loads at these substations. As an example, Figure 1.1 illustrates the IEEE 14-Bus system.

The state of a power system is denoted as the complex voltage of each system bus. It is often denoted using the phasor representation $Ee^{j\phi}$, where E denotes the magnitude, and ϕ denotes the phase. State estimation provides overall monitoring of the system status, and is important for system operators to maintain the operating conditions in a normal and secure state [3]. Accurate state estimation is also critical for other applications, including electricity market operations [4].

A traditional approach to estimate power system state is to employ weighted-least-squares (WLS) estimation using SCADA measurements [3]. The SCADA system measures power injection and power flow on selected buses and branches, and the measurements are nonlinear functions of the system state. If we denote the measurements as a vector \mathbf{z} , and the system state as a vector \mathbf{s} , the relationship between the state and measurements can then be denoted as

$$\mathbf{z} = \mathbf{h}(\mathbf{s}) + \boldsymbol{\epsilon}, \quad (1.1)$$

where $\boldsymbol{\epsilon}$ denotes the measurement noise. Directly estimating the system state \mathbf{s} with the nonlinear vector function $\mathbf{h}(\mathbf{s})$ results in a non-convex optimization which is often very difficult to solve. Therefore, an iterative Gauss-Newton iteration is usually used, wherein each iteration the nonlinear vector function is linearized using Taylor series expansion. Let

\mathbf{H} denote the Jacobian matrix, and then the measurement model reduces to

$$\mathbf{z} = \mathbf{H}\mathbf{s} + \boldsymbol{\epsilon}. \quad (1.2)$$

The state estimate is then obtained using the WLS estimation

$$\hat{\mathbf{s}} = \arg \min_{\mathbf{x}} (\mathbf{z} - \mathbf{H}\mathbf{s})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \mathbf{H}\mathbf{s}), \quad (1.3)$$

where $\boldsymbol{\Sigma}$ is the (known) covariance matrix of the measurement noise.

The main limitations of traditional SCADA based state estimation include its low sample rate, asynchronous measurements, and nonlinear measurement model. Recently, phasor measurement units (PMUs) [5] have been used for power system state estimation. PMUs provide time-synchronized high sample rate measurements of complex bus voltages and currents. In Cartesian coordinates, the measurement model is a linear function of the real and imaginary parts of the system state. These advantages enable monitoring of transient states and other important applications. Despite PMUs' benefits, several issues have to be taken care of, including the phase angle mismatch that exists in PMU measurements, and PMU placement strategy to optimize the locations of a limited number of units.

1.1.2 Renewable energy in power grids

Renewable energy sources [6], including solar and wind energy, currently provide only about 3% of the electricity in the United States. However, a high penetration of renewable energy is projected for various reasons. The future shortage of fossil fuels, environmental concerns, and advances in smart grid technologies all stimulate an increasing penetration of renewable energy. Researchers have shown that supplying all the energy needs of the United States from renewable energy is realizable in the future [7]. According to the National Renewable Energy Laboratory (NREL), renewable energy potentially will support about 80% of the total electricity consumption in the U.S. in 2050 [8].

While there are centralized renewable generation facilities like solar parks or wind farms, a different and appealing approach is to use renewable energy locally. Local use avoids the problem of long-distance transmission and makes the grid more decentralized, and thus

more robust against large-scale blackouts. Based on the scale of the load and the normal operating mode, two scenarios can be considered. In the first scenario, small end users, e.g., households, install on-site low-profile renewable generators that partially satisfy their energy needs. These users are also connected to the main grid, with the ability to both draw electricity from the grid and inject excess generation into the grid. In this scenario, optimizing the energy consumption schedule is important as it can not only reduce costs for the end users, but also benefit the grid by shaving peak-hour loads and increasing energy efficiency. The second scenario of consideration is larger energy users who form a micro-grid and rely on a significant level of renewable penetration. Micro-grids can operate in isolation mode, when their single point of coupling with the main grid is detached. When a micro-grid is designed to be self-sustained most of the time, it is important to plan a combination of different energy generation and storage capacities, to ensure an uninterrupted energy supply to meet the grid needs. This is especially true when considering the intermittency of renewable generation and its varying geographical and temporal availability.

1.2 Contributions of this work

This dissertation first considers power system state estimation using PMUs with phase angle mismatch, and then optimization for renewable energy integration in different scenarios. We summarize the main contributions as follows.

State estimation using PMUs with phase mismatch: As mentioned in Section 1.1.1, PMUs are time-synchronized sensors primarily used for power system monitoring. However, inaccurate synchronization is an inevitable problem, and phase angle mismatch exists in PMU measurements. Here, we develop a model for power system state estimation using PMUs with phase mismatch. We then propose various methods to estimate system states based on this measurement model. Numerical examples on IEEE standard test systems are used to demonstrate the improved accuracy of our algorithms over traditional algorithms when imperfect synchronization is present. We conclude that when a sufficient number of PMUs with small delays are employed, the imperfect synchronization can be largely compensated for in the estimation stage.

Performance bound and optimal PMU placement: Heuristically it is difficult to tell how the placement of a PMU at a specific location in the power system will affect the state estimation error, not to say the effect of the PMU synchronization and measurement accuracy. Therefore we derive the posterior Cramér-Rao bound (PCRB) on the state estimation error, based on the measurement model we developed under the first topic. We then propose a PMU placement strategy using the derived PCRB. The greedy algorithm is used to solve the optimization problem. We compare the results with other heuristics, with the optimal solution achieved through exhaustive search (for small systems), and with a lower bound on the optimal placement obtained through convex relaxation. We show that for some design criteria, the objective functions are submodular, which guarantees a performance bound on the greedy solution. For other design criteria where the objective functions are not submodular, numerical examples demonstrate the effectiveness of the greedy algorithm. We also show that using the proposed PMU placement strategy can increase the estimation accuracy by placing the limited number of PMUs in proper locations.

Parallel load scheduling with distributed renewable generation: Under this topic, we propose a framework for demand response in smart grids, where individual energy users own on-site renewable generators, and can both purchase electricity from and sell it back to the main grid. The goal is to optimize the load schedule of users to minimize the utility company’s cost and the users’ bills, while considering user satisfaction. We employ a parallel autonomous optimization scheme, where each user requires only the knowledge of the aggregated load of other users, instead of the specific load profiles of individual users. All the users can execute distributed optimization simultaneously. The distributed optimization is coordinated through a soft constraint on changes of load schedules between iterations. Numerical examples show that our method can significantly reduce the peak-hour load, costs to the utility, and users’ bills. Since the autonomous user optimization is executed in parallel, our method also significantly decreases computation time and communication costs.

Joint storage and renewable generation capacity planning: In a micro-grid with a small carbon footprint, the penetration of renewable energy is usually high. In such power grids, energy storage is important to guarantee an uninterrupted and stable power supply for end users. Due to the different characteristics of energy storage systems and the different availability of renewable energy resources, joint capacity optimization for multiple types of energy storage and generation is important when designing this type of power systems.

Under this topic, we formulate a cost minimization problem for storage and generation planning, considering both the initial investment cost and operational/maintenance cost, and propose a distributed optimization framework to overcome the difficulty brought about by the large size of the optimization problem. The results are helpful in making decisions on energy storage and generation capacity planning in future decentralized power grids with high renewable penetrations.

1.3 Organization of the dissertation

The rest of the dissertation is organized as follows. Chapters 2 and 3 consider state estimation using PMUs. In Chapter 2 we develop a new measurement model, and propose algorithms to increase state estimation accuracy. In Chapter 3 we derive the posterior Cramér-Rao bound based on the proposed model, and propose a PMU placement strategy. Chapters 4 and 5 consider optimization for renewable energy integration in smart grids. We consider the case of parallel load scheduling with on-site renewable generators in Chapter 4, and then the optimization of renewable generation and storage capacity for micro-grids in Chapter 5. We finally summarize the dissertation in Chapter 6, and point out potential future directions.

1.4 Notations

We use lower-case italic symbols to denote scalars (a), bold lower-case italic symbols to denote vectors (\mathbf{a}), bold upper-case italic symbols to denote matrices (\mathbf{A}), and calligraphic symbols to denote sets (\mathcal{A}).¹

We use $\{\cdot\}^c$ to denote continuous signals, $\{\cdot\}^r$ to denote the real part, and $\{\cdot\}^i$ to denote the imaginary part. We use $\text{card}(\cdot)$ to denote the cardinality of a set, $\text{mod}(\cdot, \cdot)$ to denote the modulo operator, and superscript \top to denote matrix or vector transpose. We use $\text{tr}(\cdot)$ to denote the trace of a matrix, $\lambda_{\max}(\cdot)$ to denote the maximum eigenvalue of a matrix, and $\det(\cdot)$ to denote matrix determinant. We use $\mathbb{E}(\cdot)$ to denote expectation.

¹Exceptions are that we use bold calligraphic \mathcal{I} to denote the Fisher Information Matrix, and \mathcal{N} to denote the Gaussian distribution.

We use $\{\mathbf{a}^j\}$ to denote a collection of all \mathbf{a}^j 's for $j \in \mathcal{J}$, $(\mathbf{a})_j$ or a_j to denote the j th element of vector \mathbf{a} , $\|\mathbf{a}\|_p$ to denote the ℓ_p norm, and $\|\mathbf{a}\|_*$ to denote the nuclear norm. In some cases, we use $\|\mathbf{a}\|_{\Sigma^{-1}}^2$ to denote $\mathbf{a}^\top \Sigma^{-1} \mathbf{a}$ for simplicity. The concatenation of two vectors $[\mathbf{a}; \mathbf{b}]$ is equivalent to $[\mathbf{a}^\top, \mathbf{b}^\top]^\top$. We use $(\mathbf{A})_n$ to denote the n th column vector of \mathbf{A} , \mathbf{A}_{mn} to denote its (block) sub-matrix, and $(\mathbf{A})_{m,n}$ to denote its element on the m th row and n th column.

We use \in to denote “is a member of”, \subseteq to denote “is a subset of or equal to”, \forall to denote “for all”, \cup to denote set union, \cap to denote set intersection, and \setminus to denote set subtraction. We use $>$ and $<$ to for element-wise comparison, and \succ and \prec for comparison in the matrix sense, i.e., $\mathbf{A} \succ \mathbf{B}$ means $\mathbf{A} - \mathbf{B}$ is positive definite.

Chapter 2

State Estimation Using PMUs with Phase Mismatch

We will consider power system state estimation using PMUs in the following two chapters. In this chapter, we consider the problem of state estimation using PMU measurements, when phase angle mismatch exists in the measurements.²

2.1 Introduction

In recent years, phasor measurement units (PMUs) have become increasingly important in power system state estimation [3, 9, 10, 11, 12, 13]. The traditional supervisory control and data acquisition (SCADA) system has a low reporting rate, and requires complex nonlinear state estimation, since the SCADA measurements, e.g., the power flow and power injections, are nonlinear functions of the system states (complex bus voltages). PMUs provide synchronized phasor measurements of bus voltages and currents, which results in linear models in Cartesian coordinates for state estimation. Their sampling rate is much higher than that of SCADA systems, enabling real-time estimation of the power system's state and fast response to abnormalities.

²This chapter is based on P. Yang, Z. Tan, A. Wiesel, and A. Nehorai, "Power system state estimation using PMUs with imperfect synchronization," *IEEE Trans. on Power Systems*, Vol 28, No. 4, pp. 4162-4172, Nov. 2013. © IEEE 2013.

There has been ongoing research on state estimation using PMUs [14, 15, 16, 17, 18, 19]. Most of the recent work directly combines the SCADA data with data from PMUs, and uses weighted least squares (WLS) estimation or similar methods. One important issue with this approach is that the SCADA measurements are not synchronized, and the sample rates of SCADA and PMUs are different, causing time skewness problems [20].

Typically, PMUs use a global positioning system (GPS) radio clock, which sends one pulse per second (1 pps) synchronization signal [21]. In practice, PMU measurements are not perfectly synchronized, and phase angle mismatch exists in the PMU measurements for various reasons. Possible reasons include different standards, protocols, and designs employed by different vendors [22]; errors from the synchronization protocols [23, 24]; or malfunction of the PMU algorithms or timing circuits [18, 25]. According to the IEEE Standard for Synchrophasors for Power Systems (IEEE Std C37.118-2005) [26], a phase angle of 0.01 rad will cause 1% total vector error (TVE), which corresponds to a maximum time error of $\pm 26\mu\text{s}$ in 60Hz systems. The accuracy and consistency of all PMUs, regardless of their makes and models, is important for large scale PMU deployments [27]. However, much larger phase angle mismatch than the IEEE standard has been observed in real PMU measurements [18, 22, 28].

In large scale deployments, PMUs with alternative synchronization mechanisms may be used. These alternative synchronization mechanisms may include the Precision Time Protocol (PTP) as defined in IEEE-1588 standard [29, 23, 24], or time signal radio stations, e.g., WWVB located in Colorado, US [30]. According to the IEEE-1588 standard, instead of purely using a GPS radio clock for each of the devices, only the “masters” are equipped with global clocks. The “slaves” use local clocks, and a sync message is transmitted from a “master” to its “slaves” every few seconds. Alternatively, the WWVB radio station uses a pulse amplitude modulated signal with a bit rate of 1b/s to synchronize widely separated clocks, with lower accuracy than GPS radio clocks. As a general model, in this paper we consider different PMUs are synchronized every T_{sync} seconds, and use imperfect local clocks between consecutive synchronizations.

When the PMUs are not perfectly synchronized, the traditional measurement model which considers the phase mismatch resulting from dissynchronization as additive noise is no longer accurate. In fact, as our numerical example suggests, if the synchronization error and/or

the time between consecutive synchronizations increases, the traditional estimation methods will deteriorate significantly. To mitigate this problem, we introduce a new model for state estimation that takes PMU phase mismatch into consideration. We propose estimation algorithms based on alternating minimization (AM) and parallel Kalman filtering (PKF) [31], for estimation using the static and dynamic models respectively. The estimation algorithm based on the static model is simple, robust, and does not require any assumptions on the dynamics of the states and phase mismatch. Nevertheless, the filtering approach is preferable for tracking time varying phase mismatches under the standard dynamic state space model, as it is less computationally intensive than the AM method. We also briefly describe other potential solutions to the problem, including robust least squares estimation and low rank relaxations methods, which do not provide satisfying results in this particular application. Numerical examples demonstrate that our proposed algorithms provide more accurate state estimates when the PMUs are imperfectly synchronized. The estimation performance remains satisfactory when the synchronization error increases. We conclude that when a sufficient number of PMUs are employed and the mismatches are small, our methods can largely compensate for errors resulting from imperfect time synchronization.

The rest of this chapter is organized as follows. In Section 2.2 we describe the measurement and system model considering phase mismatch. In Section 2.3 we introduce the proposed algorithms for state estimation with phase mismatch. We show numerical examples in Section 2.4, and summarize in Section 2.5.

2.2 Measurement and system model

2.2.1 Measurement model

We consider a grid model with N buses connected via branches. The continuous voltage signal on bus p at time instance $t \in \mathbb{R}^+$ is defined as

$$\overline{E}_p^c(t) = E_p^c(t) \cos(2\pi f_c t + \varphi_p^c(t)), \quad (2.1)$$

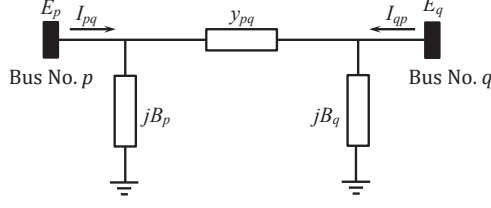


Figure 2.1: Bus branch model.

where f_c is the frequency. The phasor representation of $\overline{E}_p^c(t)$ is $E_p^c(t)e^{j\varphi_p^c(t)}$, with $E_p^c(t)$ denoting the magnitude and $\varphi_p^c(t)$ denoting the phase. For simplicity, we work with a time series of discrete phasors

$$\begin{aligned} E_{p,k} &= E_p^c(kT), \\ \varphi_{p,k} &= \varphi_p^c(kT), \end{aligned} \tag{2.2}$$

where T is the PMU reporting period, typically around tens of milliseconds, and $k \in \mathbb{N}^+$. In Cartesian coordinates this translates to

$$E_{p,k}^r = E_{p,k} \cos \varphi_{p,k}, \tag{2.3}$$

$$E_{p,k}^i = E_{p,k} \sin \varphi_{p,k}. \tag{2.4}$$

We define the state of the power grid at time instance k as a length $2N$ real valued vector

$$\mathbf{s}_k = [E_{1,k}^r, E_{1,k}^i, E_{2,k}^r, E_{2,k}^i, \dots, E_{N,k}^r, E_{N,k}^i]^\top. \tag{2.5}$$

The voltages on the buses are related to the currents through the branches, as illustrated in Fig. 2.1. We denote the susceptance at bus p as B_p , and the admittance at branch $\{p, q\}$ as y_{pq} , with

$$y_{pq} = g_{pq} + jb_{pq}, \tag{2.6}$$

where g_{pq} is the conductance, b_{pq} is the susceptance, and j equals $\sqrt{-1}$. These parameters are assumed to be known and constant. Consequently, the real and imaginary parts of the

current on branch $\{p, q\}$ are given by

$$I_{pq,k}^r = (E_{p,k}^r - E_{q,k}^r) g_{pq} - (E_{p,k}^i - E_{q,k}^i) b_{pq} - B_p E_{p,k}^i, \quad (2.7)$$

$$I_{pq,k}^i = (E_{p,k}^r - E_{q,k}^r) b_{pq} + (E_{p,k}^i - E_{q,k}^i) g_{pq} + B_p E_{p,k}^r. \quad (2.8)$$

To avoid the time skewness problems that arise when traditional SCADA measurements are combined with PMU measurements, here we consider state estimation using only PMUs. An alternative approach is to incorporate SCADA estimation as priors for the estimation based on PMU data [32], which could be easily incorporated into our model. In an ideal setting, the power system is monitored via a network of M perfectly synchronized PMUs measuring the voltages and currents located in a set of buses \mathcal{M} at time stamps kT for $k \in \mathbb{N}^+$, where $M = \text{card}(\mathcal{M})$. The PMU installed on the p th bus measures noisy versions of

$$\tilde{\mathbf{z}}_k^p = \mathbf{H}_p \mathbf{s}_k^p, \quad (2.9)$$

with

$$\tilde{\mathbf{z}}_k^p = [E_{p,k}^r, E_{p,k}^i, I_{pq_1,k}^r, I_{pq_1,k}^i, \dots, I_{pq_u,k}^r, I_{pq_u,k}^i]^\top, \quad (2.10)$$

$$\mathbf{s}_k^p = [E_{p,k}^r, E_{p,k}^i, E_{q_1,k}^r, E_{q_1,k}^i, \dots, E_{q_u,k}^r, E_{q_u,k}^i]^\top, \quad (2.11)$$

where $\{q_1, \dots, q_u\}$ are the indices of the neighboring buses to bus p . The matrix \mathbf{H}_p can be written as

$$\mathbf{H}_p = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \Upsilon_{pq_1} & \tilde{\Upsilon}_{pq_1} & \mathbf{0} & \dots & \mathbf{0} \\ \Upsilon_{pq_2} & \mathbf{0} & \tilde{\Upsilon}_{pq_1} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Upsilon_{pq_u} & \mathbf{0} & \mathbf{0} & \dots & \tilde{\Upsilon}_{pq_u} \end{bmatrix}, \quad (2.12)$$

where

$$\Upsilon_{pq_i} = \begin{bmatrix} g_{pq_i} & -b_{pq_i} - B_p \\ b_{pq_i} + B_p & -b_{pq_i} \end{bmatrix}, \quad (2.13)$$

$$\tilde{\Upsilon}_{pq_i} = \begin{bmatrix} -g_{pq_i} & b_{pq_i} \\ -b_{pq_i} & -g_{pq_i} \end{bmatrix}. \quad (2.14)$$

Stacking the noisy versions of (2.9) for all $p \in \mathcal{M}$ into one large model yields the traditional power grid observation model

$$\mathbf{z}_k = \tilde{\mathbf{z}}_k + \boldsymbol{\epsilon}_k = \mathbf{H} \mathbf{s}_k + \boldsymbol{\epsilon}_k, \quad (2.15)$$

where $\tilde{\mathbf{z}}_k^p$, \mathbf{H}_p , and \mathbf{s}_k^p are the appropriate sub-blocks of $\tilde{\mathbf{z}}_k$, \mathbf{H} , and \mathbf{s}_k corresponding to the PMU on the p th bus, and $\boldsymbol{\epsilon}_k$ denotes Gaussian measurement noise with $\boldsymbol{\epsilon}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{\epsilon,k})$. In the common case where the noise is independently and identically distributed (i.i.d.), the noise covariance satisfies $\mathbf{R}_{\epsilon,k} = \sigma^2 \mathbf{I}$.

We now propose a more realistic measurement model which takes into account the imperfect synchronization of the PMUs. The PMU installed on the p th bus measures its k th sample at time $kT + t_{p,k}$, where $t_{p,k}$ is the time delay of the p th PMU at the k th sample. Denote the delays of all PMUs as a vector $\mathbf{t}_k = [\dots, t_{p,k}, \dots]^\top$, where $p \in \mathcal{M}$.

The voltage at bus p at time instance $kT + t_{p,k}$ is

$$\begin{aligned} \overline{E}_p^c(kT + t_{p,k}) &= E_p^c(kT + t_{p,k}) \cos(2\pi f_c(kT + t_{p,k}) + \varphi_p^c(kT + t_{p,k})) \\ &\approx E_p^c(kT) \cos(2\pi f_c kT + 2\pi f_c t_{p,k} + \varphi_p^c(kT)) \\ &= E_{p,k} \cos(2\pi f_c kT + 2\pi f_c t_{p,k} + \varphi_{p,k}), \end{aligned} \quad (2.16)$$

where the approximation holds because $t_{p,k} \ll T$ (typically tens of microseconds in comparison to tens of milliseconds). Let f_c denote the power frequency, and thus the phase mismatch can be written as

$$\boldsymbol{\theta}_k = 2\pi f_c \mathbf{t}_k. \quad (2.17)$$

The phasor notation for the complex voltage is then $E_{p,k}e^{j(\varphi_{p,k}+\theta_{p,k})}$. The real and imaginary parts of this delayed voltage can be expressed as

$$\tilde{E}_{p,k}^r = E_{p,k} \cos(\varphi_{p,k} + \theta_{p,k}) \approx E_{p,k}^r - E_{p,k}^i \theta_{p,k}, \quad (2.18)$$

$$\tilde{E}_{p,k}^i = E_{p,k} \sin(\varphi_{p,k} + \theta_{p,k}) \approx E_{p,k}^i - E_{p,k}^r \theta_{p,k}, \quad (2.19)$$

where we have used the standard approximations

$$\sin \theta_{p,k} \approx \theta_{p,k}, \quad \cos \theta_{p,k} \approx 1, \quad (2.20)$$

which hold for small values of θ_p (typically less than 1 degree, corresponding to $46.3 \mu s$ delay at $f_c = 60 \text{Hz}$). The delayed currents are detailed in (2.21) and (2.22), where we temporarily omitted the subscript k for simplicity of notation.

$$\begin{aligned} \tilde{I}_{pq}^r &= (E_p \cos(\varphi_p + \theta_p) - E_q \cos(\varphi_q + \theta_p)) g_{pq} \\ &\quad - (E_p \sin(\varphi_p + \theta_p) - E_q \sin(\varphi_q + \theta_p)) b_{pq} - B_p E_p \sin(\varphi_p + \theta_p) \\ &\approx E_p^r (g_{pq} + b_{pq} \theta_p + B_p \theta_p) + E_q^r (-g_{pq} - b_{pq} \theta_p) \\ &\quad + E_p^i (-g_{pq} \theta_p - b_{pq} - B_p) + E_q^i (g_{pq} \theta_p + b_{pq}), \end{aligned} \quad (2.21)$$

$$\begin{aligned} \tilde{I}_{pq}^i &= (E_p \cos(\varphi_p + \theta_p) - E_q \cos(\varphi_q + \theta_p)) b_{pq} \\ &\quad + (E_p \sin(\varphi_p + \theta_p) - E_q \sin(\varphi_q + \theta_p)) g_{pq} + B_p E_p \cos(\varphi_p + \theta_p) \\ &\approx E_p^r (b_{pq} - g_{pq} \theta_p + B_p) + E_q^r (-b_{pq} + g_{pq} \theta_p) \\ &\quad + E_p^i (-b_{pq} \theta_p + g_{pq} - B_p \theta_p) + E_q^i (b_{pq} \theta_p - g_{pq}). \end{aligned} \quad (2.22)$$

Note that all the measurements from the PMU installed on a particular bus, namely the voltage and the currents on all the adjacent branches, are associated with the same phase mismatch θ_p , as they use the same time stamp. Thus, the imperfectly synchronized version of (2.9) is given by

$$\tilde{\mathbf{z}}_k^p = \mathbf{H}_p \mathbf{s}_k^p + \theta_{p,k} \mathbf{G}_p \mathbf{s}_k^p, \quad (2.23)$$

where \mathbf{H}_p is defined in (2.10), and \mathbf{G}_p is

$$\mathbf{G}_p = \begin{bmatrix} -\tilde{\mathbf{I}} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \Xi_{pq_1} & \tilde{\Xi}_{pq_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \Xi_{pq_2} & \mathbf{0} & \tilde{\Xi}_{pq_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Xi_{pq_u} & \mathbf{0} & \mathbf{0} & \cdots & \tilde{\Xi}_{pq_u} \end{bmatrix}, \quad (2.24)$$

with

$$\tilde{\mathbf{I}} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad (2.25)$$

$$\Xi_{pq_i} = \begin{bmatrix} b_{pq_i} + B_p & -g_{pq_i} \\ -g_{pq_i} & -b_{pq_i} - B_p \end{bmatrix}, \quad (2.26)$$

$$\tilde{\Xi}_{pq_i} = \begin{bmatrix} -b_{pq_i} & g_{pq_i} \\ g_{pq_i} & b_{pq_i} \end{bmatrix}. \quad (2.27)$$

Stacking these observations together and padding zeros in appropriate locations in corresponding matrices yields the following bilinear observation model³

$$\mathbf{z}_k = \left(\mathbf{H} + \sum_{m \in \mathcal{M}} \theta_{m,k} \mathbf{G}_m \right) \mathbf{s}_k + \boldsymbol{\epsilon}_k, \quad (2.28)$$

where $\boldsymbol{\epsilon}_k$ denotes Gaussian measurement noise following $\mathcal{N}(\mathbf{0}, \mathbf{R}_{\epsilon,k})$. In some settings, \mathbf{G}_m can be replaced by $\mathbf{G}_{m,k}$ to account for time varying system topology or parameters.

³It is worth mentioning that recently a bilinear state estimation approach [33, 34] has also been proposed as an alternative to conventional SCADA based state estimation using the well-known Gauss-Newton iterative schemes, where the original nonlinear measurement model is rephrased as two linear models, coupled through a nonlinear change of variables.

2.2.2 Dynamic models for system state and phase mismatch

System state dynamics

Following [35], we adopt a state space linear dynamic model for the system state:

$$\mathbf{s}_{k+1} = \mathbf{A}_{s,k}\mathbf{s}_k + \mathbf{B}_{s,k}\mathbf{u}_{s,k} + \mathbf{w}_{s,k}. \quad (2.29)$$

The matrix $\mathbf{A}_{s,k}$ relates the state at the previous time step to the state at the current time step. The matrix $\mathbf{B}_{s,k}$ relates the controls and other driving forces $\mathbf{u}_{s,k}$ to the state. The random vector $\mathbf{w}_{s,k}$ is assumed to be multivariate Gaussian with

$$\mathbf{w}_{s,k} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{s,k}). \quad (2.30)$$

The covariance $\mathbf{Q}_{s,k}$ can incorporate additional prior information from network topology or SCADA estimation, etc. In a dynamic state estimation scenario, the parameters $\mathbf{A}_{s,k}$, $\mathbf{B}_{s,k}$ and $\mathbf{Q}_{s,k}$ are calculated online through the parameter identification process [36].

Phase mismatch dynamics

We assume that the PMUs are jointly synchronized every T_{sync} seconds. Immediately after the synchronization, the delays \mathbf{t}_0 follow a Gaussian distribution, with

$$\mathbf{t}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{t,0}), \quad (2.31)$$

where the covariance matrix $\mathbf{R}_{t,0}$ depends on the synchronization accuracy of the specific synchronization mechanism employed. Also, depending on the synchronization mechanism, \mathbf{t}_0 can follow different probability distributions. Between two synchronizations, we assume \mathbf{t}_k follows the linear dynamic model

$$\mathbf{t}_{k+1} = \mathbf{A}_{t,k}\mathbf{t}_k + \mathbf{B}_{t,k}\mathbf{u}_{t,k} + \mathbf{w}_{t,k}. \quad (2.32)$$

The control variable $\mathbf{u}_{t,k}$ includes temperature and other control dynamics which affect the time synchronization. The covariance of $\mathbf{w}_{t,k}$, $\mathbf{Q}_{t,k}$, can either be white Gaussian, assuming

independent time drifts of different sensors, or can have topologically based structures associated with advanced distributed synchronization mechanisms. It has been shown in [23] that the clock state of an IEEE 1588 network satisfies a similar linear dynamic model.

Using (2.17), we can easily write the dynamics of the phase mismatch $\boldsymbol{\theta}_k$ as

$$\boldsymbol{\theta}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{\boldsymbol{\theta},0}), \quad (2.33)$$

$$\boldsymbol{\theta}_{k+1} = \mathbf{A}_{\boldsymbol{\theta},k} \boldsymbol{\theta}_k + \mathbf{B}_{\boldsymbol{\theta},k} \mathbf{u}_{\boldsymbol{\theta},k} + \mathbf{w}_{\boldsymbol{\theta},k}. \quad (2.34)$$

2.3 State estimation considering phase mismatch

In the previous section we formulated the power grid statistical models. State estimation is the problem of recovering \mathbf{s}_k given $\mathbf{z}_k, \mathbf{z}_{k-1}, \dots, \mathbf{z}_0$. The system state can be estimated using two types of techniques: static state estimation (SSE) and dynamic state estimation (DSE) [36]. SSE estimates the system state at time instant k using measurements for the same instant of time. The most commonly used method for SSE is the weighted least-squares (WLS) method [3]. DSE depends on physical modeling of the time varying nature of the power system, and employs dynamic state models, e.g., the model defined in (2.29), where the model parameters are estimated online. The estimation is traditionally obtained using Kalman filtering [13, 15, 37, 38]. In this chapter, we will not discuss system identification. Rather, we assume the system parameters (structures) are known, and propose methods for state estimation based on the static and the dynamic models considering synchronization errors.

We consider a more realistic model defined in (2.28), (2.29), and (2.34). This formulation involves additional nuisance parameters $\boldsymbol{\theta}_k$, which complicates the inference of \mathbf{s}_k . The optimal filter solution requires the computation of the posterior distribution of unknown state parameters marginalized over the nuisance parameters, and is clearly intractable. Instead, we first propose an alternating minimization based estimation approach under the static model, without considering the state or phase mismatch dynamics. Then we propose an approximate solution based on joint estimation of both \mathbf{s}_k and \mathbf{t}_k via two parallel yet coupled Kalman filters under the dynamic model. We also discuss some other possible state estimation

methods based on model (2.28), which do not provide satisfying results in this specific application.

2.3.1 State estimation using the static model

Static state estimation considers the estimation of \mathbf{s}_k given \mathbf{z}_k . It does not exploit previous measurements, nor does it assume the dynamic models in (2.29) and (2.34). Instead of a dynamic characterization, we assume the time dissynchronization follows a Gaussian prior distribution:

$$\mathbf{t}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{t,k}), \quad (2.35)$$

where $\mathbf{R}_{t,k}$ is calculated based on (2.33) and (2.34). In a simple case where $\mathbf{A}_{t,k} = \mathbf{I}$ and $\mathbf{B}_{t,k} = \mathbf{0}$, we obtain

$$\mathbf{R}_{t,k} = \mathbf{R}_{t,0} + \text{mod} \left(k, \frac{1}{T} \right) \mathbf{Q}_{t,k}. \quad (2.36)$$

The covariance $\mathbf{R}_{t,k}$ changes with k because the time delay evolves over time following (2.34), and is reset to $\mathbf{R}_{t,0}$ immediately after synchronization. Convert the units, and we obtain that the phase mismatch $\boldsymbol{\theta}_k$ follows a simple Gaussian prior $\mathcal{N}(\mathbf{0}, \mathbf{R}_{\theta,k})$. We model the states as deterministic unknown variables without any prior distribution. The algorithm can be easily modified to allow additional information, e.g., that provided by existing SCADA estimation. The main advantage of these static simplifications is the robustness to inaccurate state space modeling or inaccurate system identification.

Static state estimation with nuisance phase mismatches is a regularized structured total least squares (STLS) [39] problem, where errors exist not only in the observation vector, but also in the data matrix. The errors in the data matrix also exhibit certain structures, in this case an affine function of the matrices associated with each bus installed with PMUs. Estimating the system state is then a process of finding a solution to the STLS problem, with regularization for the phase mismatch prior. Total least squares (TLS) [40, 41] has a classical closed form solution based on singular value decomposition (SVD). However, STLS is an open problem which is still not fully understood. Two possible approaches to STLS are an alternating minimizations (AM) method which sequentially solves for the state or the

phases independently, or a low rank relaxation technique via nuclear norm minimization [42]. In our setting, the number of measurements is small compared to the number of unknown variables, and the phase mismatch is also relatively small. We found through exploratory experiments that in this case, the simple AM approach is more preferable than the low rank relaxation technique⁴. Due to the bilinear structure of (2.28), the optimal estimator for each of the unknown parameters, assuming the other is known, has a simple closed form solution. Thus, we propose the following AM approach:

Estimate \mathbf{s}_k , assuming $\boldsymbol{\theta}_k$ is known

Assuming $\boldsymbol{\theta}_k$ is known, and denoting

$$\boldsymbol{\Phi}(\boldsymbol{\theta}_k) = \left(\mathbf{H} + \sum_{m \in \mathcal{M}} \theta_{m,k} \mathbf{G}_m \right), \quad (2.37)$$

we can then write (2.28) as

$$\mathbf{z}_k = \boldsymbol{\Phi}(\boldsymbol{\theta}_k) \mathbf{s}_k + \boldsymbol{\epsilon}_k. \quad (2.38)$$

The maximum likelihood estimator of \mathbf{s}_k is then obtained by solving the weighted least-squares (WLS) problem:

$$\hat{\mathbf{s}}_k = \arg \min_{\mathbf{s}_k} (\mathbf{z}_k - \boldsymbol{\Phi}(\boldsymbol{\theta}_k) \mathbf{s}_k)^\top \mathbf{R}_{\epsilon,k}^{-1} (\mathbf{z}_k - \boldsymbol{\Phi}(\boldsymbol{\theta}_k) \mathbf{s}_k). \quad (2.39)$$

The closed form solution to (2.39) is

$$\hat{\mathbf{s}}_k = \left(\boldsymbol{\Phi}(\boldsymbol{\theta}_k)^\top \mathbf{R}_{\epsilon,k}^{-1} \boldsymbol{\Phi}(\boldsymbol{\theta}_k) \right)^{-1} \boldsymbol{\Phi}(\boldsymbol{\theta}_k)^\top \mathbf{R}_{\epsilon,k}^{-1} \mathbf{z}_k. \quad (2.40)$$

In this step, it is also possible to include the state estimation from the SCADA system, as prior information. Then we can assume a Gaussian prior distribution of the system state,

$$\mathbf{s}_k \sim \mathcal{N}(\check{\mathbf{s}}_k, \check{\mathbf{R}}_{s,k}), \quad (2.41)$$

⁴We will briefly describe the low rank relaxation technique in Section 2.3.3.

where $\check{\mathbf{s}}_k$ and $\check{\mathbf{R}}_{s,k}$ denote the system state and covariance, respectively, estimated using the SCADA system. With this prior information, we can then replace the maximum likelihood estimator (2.39) with a maximum-a-posteriori (MAP) estimator to estimate the system state \mathbf{s}_k .

Estimate $\boldsymbol{\theta}_k$, assuming \mathbf{s}_k is known

Assuming \mathbf{s}_k is known from (2.39), we estimate the phase mismatch $\boldsymbol{\theta}_k$. Let $\boldsymbol{\psi}_m(\mathbf{s}_k) = \mathbf{H}_m \mathbf{s}_k$ and $\boldsymbol{\Psi}(\mathbf{s}_k) = [\dots, \boldsymbol{\psi}_m(\mathbf{s}_k), \dots]$, where $m \in \mathcal{M}$. We can then write (2.28) as

$$\mathbf{z}_k = \mathbf{H} \mathbf{s}_k + \boldsymbol{\Psi}(\mathbf{s}_k) \boldsymbol{\theta} + \boldsymbol{\epsilon}_k. \quad (2.42)$$

The maximum-a-posteriori (MAP) estimator of $\boldsymbol{\theta}_k$ is then

$$\begin{aligned} \hat{\boldsymbol{\theta}}_k = \arg \min_{\boldsymbol{\theta}_k} & (\mathbf{z}_k - \mathbf{H} \mathbf{s}_k - \boldsymbol{\Psi}(\mathbf{s}_k) \boldsymbol{\theta}_k)^\top \mathbf{R}_{\epsilon,k}^{-1} (\mathbf{z}_k - \mathbf{H} \mathbf{s}_k - \boldsymbol{\Psi}(\mathbf{s}_k) \boldsymbol{\theta}_k) \\ & + \boldsymbol{\theta}_k^\top \mathbf{R}_{\theta,k}^{-1} \boldsymbol{\theta}_k, \end{aligned} \quad (2.43)$$

and the closed form solution is

$$\hat{\boldsymbol{\theta}}_k = (\boldsymbol{\Psi}(\mathbf{s}_k)^\top \mathbf{R}_{\epsilon,k}^{-1} \boldsymbol{\Psi}(\mathbf{s}_k) + \mathbf{R}_{\theta,k}^{-1})^{-1} \boldsymbol{\Psi}(\mathbf{s}_k)^\top \mathbf{R}_{\epsilon,k}^{-1} (\mathbf{z}_k - \mathbf{H} \mathbf{s}_k). \quad (2.44)$$

Alternating algorithm for joint estimation

The estimation algorithm, as described in Table 2.1, iterates between the two steps described above and solves for the state of the system.

A possible drawback of Algorithm 2.1 is that multiple iterations have to be executed for each estimation. However, under the assumption that the phase mismatch changes slowly, the phase mismatch from a previous time point can be used to “warm start” the current estimation, thus reducing the number of iterations.

Algorithm 2.1: Alternating minimization approach for static state estimation considering PMU phase mismatch.

Data: observations \mathbf{z}_k .

Result: state estimation \mathbf{s}_k .

begin

initialize $\boldsymbol{\theta}_k$;

repeat

 update \mathbf{s}_k using (2.40);

 update $\boldsymbol{\theta}_k$ using (2.44);

until convergence or max iterations achieved;

return \mathbf{s}_k ;

end

Table 2.1: Alternating algorithm for static state estimation

Scalability for large systems

The main computational complexity comes from the inverse operation in (2.40) and (2.44). For really large power systems with thousands of buses, the scalability issue can be resolved by dividing the entire grid into sub-grids, thus significantly reducing the computational complexity in each sub-grid. Various consensus algorithms can be employed for buses that lie on the boundaries of neighboring sub-grids.

2.3.2 State estimation using the dynamic model

In this subsection we consider the case of state estimation using the dynamic model, described by (2.29) and (2.34). Kalman filtering is widely used in on-line state estimation, and is employed for power system state estimation problems [13, 15, 37, 38]. For a perfectly synchronized linear dynamic model with white Gaussian noise, the Kalman filter is known to be optimal. Our mismatched model is bi-linear and more difficult. Two vectors of parameters, the phase mismatch $\boldsymbol{\theta}_k$ and the system state \mathbf{s}_k , have to be estimated jointly.

Based on the system dynamics and the measurement model, different approximation methods could potentially be used to dynamically estimate the state \mathbf{s}_k and phase mismatch $\boldsymbol{\theta}_k$,

including total Kalman filtering [43], and recursive total least squares [44]. Our experiments that follow suggest that parallel Kalman filtering (PKF) [31] is simple to use and produces accurate estimation results.

The idea of PKF comes from two-player dynamic game theory [45] and was proposed in [31]. The solution of the game, i.e., the equilibrium, is for each player to select the best strategy corresponding to the other player's strategy. In the context of our problem, we can consider the bilinear state estimation problem as two players, with their strategy sets being their estimates of the system state and the phase mismatch. The cost function of each player is the estimation error. Therefore the solution of the game should satisfy

$$\hat{\mathbf{s}}_{k|k}^* = \arg \min_{\hat{\mathbf{s}}_{k|k}} u_s \left(\hat{\mathbf{s}}_{k|k}, \hat{\boldsymbol{\theta}}_{k|k}^* \right), \quad (2.45)$$

$$\hat{\boldsymbol{\theta}}_{k|k}^* = \arg \min_{\hat{\boldsymbol{\theta}}_{k|k}} u_\theta \left(\hat{\mathbf{s}}_{k|k}^*, \hat{\boldsymbol{\theta}}_{k|k} \right). \quad (2.46)$$

In practice, we use the “optimal predicted estimates” $\hat{\boldsymbol{\theta}}_{k|k-1}^*$ and $\hat{\mathbf{s}}_{k|k-1}^*$ instead of $\hat{\boldsymbol{\theta}}_{k|k}^*$ and $\hat{\mathbf{s}}_{k|k}^*$ to parallel the two filters. In this way we reformulate the original problem into two interlaced estimation problems on two linear time-varying systems. To simplify the notations, we define

$$\mathbf{C}_\theta(\boldsymbol{\theta}_k) = \sum_{m \in \mathcal{M}} \theta_{m,k} \mathbf{G}_{m,k}, \quad (2.47)$$

$$\mathbf{C}_s(\mathbf{s}_k) = [\cdots, \mathbf{G}_{m,k} \mathbf{s}_k, \cdots], \text{ where } m \in \mathcal{M}, \quad (2.48)$$

and therefore the original model (2.28) can be rewritten as

$$\mathbf{z}_k = \mathbf{H} \mathbf{s}_k + \mathbf{C}_\theta(\boldsymbol{\theta}_k) \mathbf{s}_k + \boldsymbol{\epsilon}_k, \quad (2.49)$$

$$\mathbf{z}_k = \mathbf{H} \mathbf{s}_k + \mathbf{C}_s(\mathbf{s}_k) \boldsymbol{\theta}_k + \boldsymbol{\epsilon}_k. \quad (2.50)$$

We then decompose the original system into two subsystems. The subsystem for system state can then be characterized by (2.29) and (2.49), whereas the subsystem for phase mismatch can be characterized by (2.34) and (2.50).

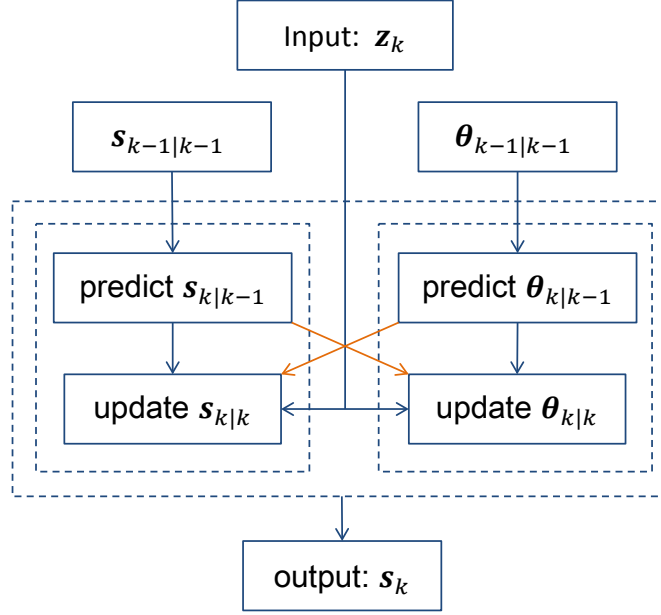


Figure 2.2: Block diagram for state estimation using parallel Kalman filter.

State and phase mismatch estimation

Based on the two subsystem models, we can couple two Kalman filters – one for state estimation and one for phase mismatch estimation. We show a block diagram illustration in Figure 2.2. First the predicted estimates $\hat{\theta}_{k|k-1}$ and $\hat{s}_{k|k-1}$ are calculated. When a measurement is received, the two filters update the estimates accordingly in parallel. The updated estimate will be used for prediction in the next time instant.

The formulas for the predictions and updates are listed below, where we use the superscripts s and θ to distinguish covariance matrices for the state and phase mismatch, respectively.

Prediction:

$$\begin{aligned}
 \hat{s}_{k|k-1} &= A_{s,k} \hat{s}_{k-1|k-1} + B_{s,k} u_{s,k} \\
 \hat{\theta}_{k|k-1} &= A_{\theta,k} \hat{\theta}_{k-1|k-1} + B_{\theta,k} u_{\theta,k} \\
 P_{k|k-1}^{(s)} &= A_{s,k} P_{k-1|k-1}^{(s)} A_{s,k}^\top + Q_{s,k} \\
 P_{k|k-1}^{(\theta)} &= A_{\theta,k} P_{k-1|k-1}^{(\theta)} A_{\theta,k}^\top + Q_{\theta,k}
 \end{aligned} \tag{2.51}$$

Update:

$$\begin{aligned}
\mathbf{K}_k^{(s)} &= \mathbf{P}_{k|k-1}^{(s)} \left(\mathbf{H} + \mathbf{C}_\theta(\hat{\boldsymbol{\theta}}_{k|k-1}) \right)^\top \\
&\quad \left(\left(\mathbf{H} + \mathbf{C}_\theta(\hat{\boldsymbol{\theta}}_{k|k-1}) \right) \mathbf{P}_{k|k-1}^{(s)} \left(\mathbf{H} + \mathbf{C}_\theta(\hat{\boldsymbol{\theta}}_{k|k-1}) \right)^\top + \mathbf{R}_{\epsilon,k} \right)^{-1} \\
\mathbf{K}_k^{(\theta)} &= \mathbf{P}_{k|k-1}^{(\theta)} \mathbf{C}_s(\hat{\mathbf{s}}_{k|k-1})^\top \\
&\quad \left(\mathbf{C}_s(\hat{\mathbf{s}}_{k|k-1}) \mathbf{P}_{k|k-1}^{(\theta)} \mathbf{C}_s(\hat{\mathbf{s}}_{k|k-1})^\top + \mathbf{R}_{\epsilon,k} \right)^{-1} \\
\mathbf{P}_{k|k}^{(s)} &= \mathbf{P}_{k|k-1}^{(s)} - \mathbf{K}_k^{(s)} \left(\mathbf{H} + \mathbf{C}_\theta(\hat{\boldsymbol{\theta}}_{k|k-1}) \right) \mathbf{P}_{k|k-1}^{(s)} \\
\mathbf{P}_{k|k}^{(\theta)} &= \mathbf{P}_{k|k-1}^{(\theta)} - \mathbf{K}_k^{(\theta)} \left(\mathbf{H} + \mathbf{C}_s(\hat{\mathbf{s}}_{k|k-1}) \right) \mathbf{P}_{k|k-1}^{(\theta)} \\
\hat{\mathbf{s}}_{k|k} &= \hat{\mathbf{s}}_{k|k-1} + \mathbf{K}_k^{(s)} \left(\mathbf{z}_k - \left(\mathbf{H} + \mathbf{C}_\theta(\hat{\boldsymbol{\theta}}_{k|k-1}) \right) \hat{\mathbf{s}}_{k|k-1} \right) \\
\hat{\boldsymbol{\theta}}_{k|k} &= \hat{\boldsymbol{\theta}}_{k|k-1} + \mathbf{K}_k^{(\theta)} \left(\mathbf{z}_k - \left(\mathbf{H} \hat{\mathbf{s}}_{k|k-1} + \mathbf{C}_s(\hat{\mathbf{s}}_{k|k-1}) \hat{\boldsymbol{\theta}}_{k|k-1} \right) \right)
\end{aligned} \tag{2.52}$$

Synchronization model

Our proposed phase mismatch model assumes that the PMUs are synchronized every few seconds, and this needs to be taken into consideration when coupling the two Kalman filters. For estimation of the phase mismatch, the matrix $\mathbf{P}_{k,k}^{(\theta)}$ has to be reset at the time of synchronization. In addition, since we are coupling this Kalman filter with the filter for state estimation, the state estimate also has to be changed at the time of synchronization. A reasonable strategy is to use the weighted least-squares estimate, without considering phase mismatch, to reset the filter for state estimation, since the phase mismatch is often very small immediately after synchronization. Due to the Gaussian linear model, all the prediction and update steps are closed-form and computationally efficient.

2.3.3 Other possible methods

In this section we consider two other possible non-iterative methods for state estimation using the bilinear observation model (2.28). The first is a robust min-max estimator [46], and the second is a low-rank relaxation method via nuclear norm minimization [42]. For

notationally simplicity, we omit subscript k in this section, since both methods use only the static model.

Min-max estimator

A robust estimator for the error-in-the-model scenario minimizes the worst case error [46]. In this problem, the model error has an affine structure. A robust estimator of the state \mathbf{s} , given that the phase mismatch is bounded, can be obtained by solving the following min-max problem:

$$\min_{\mathbf{s}} \max_{\|\boldsymbol{\theta}\| \leq \rho} \left\| \mathbf{z} - \left(\mathbf{H} + \sum_{m \in \mathcal{M}} \theta_m \mathbf{G}_m \right) \mathbf{s} \right\|_{\boldsymbol{\Sigma}^{-1}}^2. \quad (2.53)$$

Without loss of generality, we let $\rho = 1$. For a different ρ , it is easy to scale the problem to make it equal 1. Using the notation defined in (2.37), the residue can then be written as

$$r(\mathbf{H}, \mathbf{z}, \mathbf{s}) = \|\Phi(\boldsymbol{\theta})\mathbf{s} - \mathbf{z}\|. \quad (2.54)$$

Denote

$$\begin{aligned} \mathbf{M}(\mathbf{s}) &= [\mathbf{G}_1 \mathbf{s}, \dots, \mathbf{G}_M \mathbf{s}], \\ \mathbf{F} &= \mathbf{M}(\mathbf{s})^\top \mathbf{M}(\mathbf{s}), \\ \mathbf{g} &= \mathbf{M}(\mathbf{s})^\top (\mathbf{H} \mathbf{s} - \mathbf{z}), \\ h &= \|\mathbf{H} \mathbf{s} - \mathbf{z}\|^2, \end{aligned} \quad (2.55)$$

and we can then write the worst case squared residue as

$$r_S(\mathbf{H}, \mathbf{z}, \mathbf{s})^2 = \max_{\|\boldsymbol{\theta}\| \leq 1} \begin{bmatrix} 1 & \boldsymbol{\theta} \end{bmatrix} \begin{bmatrix} h & \mathbf{g}^\top \\ \mathbf{g} & \mathbf{F} \end{bmatrix} \begin{bmatrix} 1 \\ \boldsymbol{\theta} \end{bmatrix}. \quad (2.56)$$

Using the \mathcal{S} -procedure [47], and letting $\lambda > 0$, we then have that for every $\boldsymbol{\theta}$ satisfying $\boldsymbol{\theta}^\top \boldsymbol{\theta} \leq 1$,

$$\begin{bmatrix} 1 & \boldsymbol{\theta} \end{bmatrix} \begin{bmatrix} h & \mathbf{g}^\top \\ \mathbf{g} & \mathbf{F} \end{bmatrix} \begin{bmatrix} 1 \\ \boldsymbol{\theta} \end{bmatrix} \leq \lambda, \quad (2.57)$$

if and only if there exist $\gamma \geq 0$ such that

$$\begin{bmatrix} 1 & \boldsymbol{\theta} \end{bmatrix} \begin{bmatrix} \lambda - \gamma - h & -\mathbf{g}^\top \\ -\mathbf{g} & \gamma \mathbf{I} - \mathbf{F} \end{bmatrix} \begin{bmatrix} 1 \\ \boldsymbol{\theta} \end{bmatrix} \geq 0, \text{ for all } \boldsymbol{\theta}. \quad (2.58)$$

Therefore, for fixed \mathbf{s} , the worst-case residue can be computed by solving the following semi-definite programming:

$$\min_{\lambda, \gamma} \quad \lambda \quad (2.59)$$

$$\text{subject to} \quad \begin{bmatrix} \lambda - \gamma - h & -\mathbf{g}^\top \\ -\mathbf{g} & \gamma \mathbf{I} - \mathbf{F} \end{bmatrix} \geq 0. \quad (2.60)$$

Using the Schur-complement, the original problem can then be solved by the optimal solution $(\lambda, \gamma, \mathbf{s})$ of the following semi-definite programming:

$$\min_{\lambda, \gamma, \mathbf{s}} \quad \lambda \quad (2.61)$$

$$\text{subject to} \quad \begin{bmatrix} \lambda - \gamma & 0 & (\mathbf{H}\mathbf{s} - \mathbf{z})^\top \\ \mathbf{0} & \gamma \mathbf{I} & \mathbf{M}(\mathbf{s})^\top \\ \mathbf{H}\mathbf{s} - \mathbf{z} & \mathbf{M}(\mathbf{s}) & \mathbf{I} \end{bmatrix} \geq 0. \quad (2.62)$$

However, since the estimator minimizes the worst-case residue, it is robust but too conservative. The average performance is much worse than the alternating minimization method described in (2.3.1).

Low-rank relaxation method

Another non-iterative approach to estimate system state \mathbf{s} based on (2.28) is through low-rank relaxation. Let $\mathbf{B} = \mathbf{s} \begin{bmatrix} 1 & \boldsymbol{\theta}^\top \end{bmatrix}$. Clearly \mathbf{B} is of rank 1. To estimate \mathbf{B} , we will solve

the constrained optimization problem:

$$\min_{\mathbf{B}} \left\| \mathbf{z} - \sum_{n=0}^N \mathbf{G}_n \mathbf{b}_n \right\|_2 \quad (2.63)$$

$$\text{subject to} \quad \text{rank}(\mathbf{B}) = 1, \quad (2.64)$$

where $\mathbf{G}_0 = \mathbf{H}$ and $\mathbf{B} = [\mathbf{b}_0 \ \mathbf{b}_1 \ \cdots \ \mathbf{b}_N]$. However, the rank-1 constraint makes the problem non-convex and difficult to solve. Therefore, we propose to use a convex relaxation using the nuclear norm. The nuclear norm of matrix \mathbf{A} is defined by $\|\mathbf{A}\|_* = \text{trace}(\sqrt{\mathbf{A}^* \mathbf{A}})$, and is often used to relax rank-minimization problems [42]. In the noiseless case, we can solve for \mathbf{B} by minimizing its nuclear norm under the equality constraint, i.e.,

$$\min_{\mathbf{B}} \quad \|\mathbf{B}\|_* \quad (2.65)$$

$$\text{subject to} \quad \mathbf{z} = \sum_{n=0}^N \mathbf{G}_n \mathbf{b}_n, \quad (2.66)$$

where $\|\cdot\|_*$ denotes the nuclear norm. In a noisy case, the estimation can be found by solving a regularized optimization problem:

$$\min_{\mathbf{B}} \left\| \mathbf{z} - \sum_{n=0}^N \mathbf{G}_n \mathbf{b}_n \right\|_2 + \lambda \|\mathbf{B}\|_*. \quad (2.67)$$

After obtaining \mathbf{B} , the vectors \mathbf{s} and $\boldsymbol{\theta}$ will be found by executing singular value decomposition (SVD) on \mathbf{B} . Numerical experiments show that this method works well for randomly generated matrices $\{\mathbf{G}_n\}$. However, in our application, the matrices $\{\mathbf{G}_n\}$ are structured and highly sparse, and thus, the low rank relaxation approach does not provide satisfying estimation results.

2.4 Numerical examples

In this section we use numerical examples to illustrate the improvement achieved by the proposed state estimation algorithms compared with commonly used methods. We also illustrate the effect of the PMU specifications on the estimation performance, including

the number of PMUs installed, the time synchronization parameters, and the length of the synchronization interval.

2.4.1 General setup

We assume the PMUs report 30 times per second and are synchronized every second, unless otherwise specified. At time of synchronization the time delay follows (2.31), and between consecutive synchronizations, the time delay follows the linear dynamic model (2.32). Without loss of generality, we assume the synchronizations of different PMUs are independent, and therefore $\mathbf{R}_{t,0} = \sigma_{t,0}^2 \mathbf{I}$ and $\mathbf{Q}_{t,k} = \sigma_{t,k}^2 \mathbf{I}$. This assumption is not necessary here, and is only for ease and clarity of performance evaluation. In the numerical examples, we set $\sigma_{t,0} = \sigma_{t,k} = 5\mu\text{s}$, except in the example illustrating the effect of synchronization accuracy on the estimation performance.

The power system is quasi static, and the change of states is relatively slow compared to the high PMU reporting rate. Therefore in the simulation we select $\mathbf{Q}_{s,k} = \sigma_{s,k}^2 \mathbf{I}$, with $\sigma_{s,k} = 10^{-3} \text{p.u.}$ [36]. We set $\mathbf{A}_{s,k}$ to be an identity matrix [35], and $\mathbf{B}_{s,k}$ to be a zero matrix in the simulations. This specific choice of parameter does not affect the generality of our algorithm. As we mentioned before, these system parameters are estimated online through system identification in the dynamic state estimation scenario. The initial state \mathbf{s}_0 is defined via its magnitude \mathbf{E}_0 and angle $\boldsymbol{\varphi}_0$ and generated from the following distributions:

$$\begin{aligned} \mathbf{E}_0 &\sim \mathcal{N}(\mathbf{1}, 0.05^2 \mathbf{I}) \quad \text{p.u.}, \\ \boldsymbol{\varphi}_0 &\sim \mathcal{U}(\mathbf{0}, 2\boldsymbol{\pi}) \quad \text{rad.} \end{aligned} \tag{2.68}$$

These initial distributions are used for data generation only, and are not exploited by the estimation methods. The measurement noise is assumed to be i.i.d. Gaussian, with $\mathbf{R}_{\epsilon,k} = \sigma_{\epsilon}^2 \mathbf{I}$, and $\sigma_{\epsilon} = 5 \times 10^{-3} \text{ p.u.}$

We test the performance of our algorithms on the IEEE 14, 30, 57, and 118 Bus systems. For each test system, we consider two scenarios - one with the minimum number of PMUs installed for full (topological) observation, and one with redundant observations on selected buses. In this numerical example, the placement of the minimum number of PMUs for the first scenario is obtained using the method proposed in [48]. For the second scenario, buses

with redundant observations are randomly selected. In reality, we can assign more redundant observations on more important buses. We include the PMU placement profile for different test cases and different scenarios in Table 2.2.

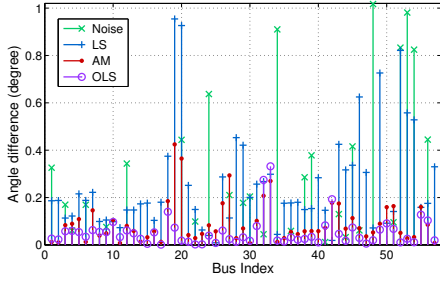
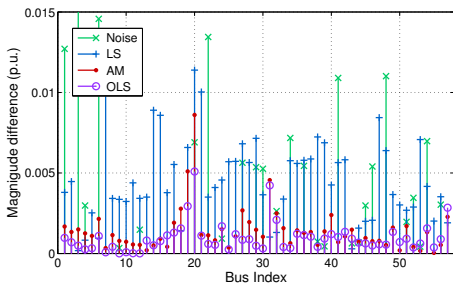
Table 2.2: PMU locations for different test systems.

# of Buses	# of PMUs	Indices of Buses with PMUs
14	4	2, 6, 7, 9
	6	2, 4, 6, 7, 9, 13
30	10	1, 7, 9, 10, 12, 18, 24, 25, 27, 28
	16	3, 4, 5, 7, 10, 11, 12, 17, 19, 22, 24, 25, 26, 28, 29, 30
57	17	1, 4, 6, 13, 20, 22, 25, 27, 29, 32, 36, 39, 41, 45, 47, 51, 54
	28	1, 3, 4, 6, 9, 12, 20, 22, 24, 27, 29, 30, 32, 34, 36, 38, 39, 41, 43, 44, 45, 46, 48, 51, 52, 53, 54, 56
118	32	3, 7, 9, 11, 12, 17, 21, 25, 28, 34, 37, 41, 45, 49, 53, 56, 62, 63, 68, 70, 71, 76, 79, 85, 86, 89, 92, 96, 100, 105, 110, 114
	54	2, 5, 8, 10, 12, 15, 17, 21, 22, 25, 26, 29, 31, 36, 37, 41, 42, 43, 44, 46, 53, 54, 56, 57, 58, 62, 64, 67, 68, 69, 70, 71, 73, 75, 77, 79, 80, 83, 85, 87, 90, 91, 94, 100, 101, 102, 105, 107, 110, 112, 114, 115, 116, 118

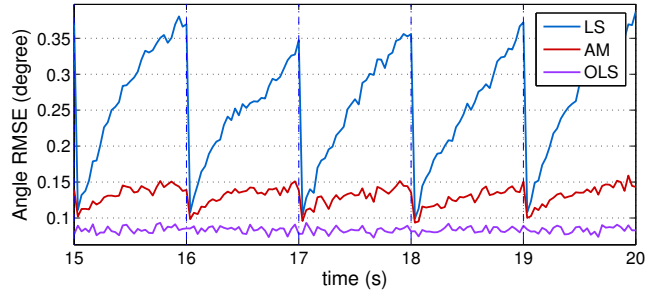
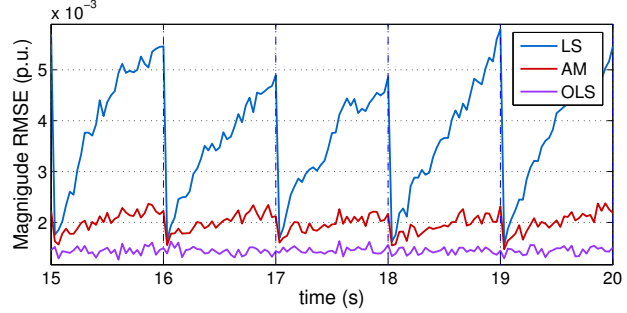
We use the root mean squared error (RMSE) as the performance measure. For each time point, we calculate the RMSE of the magnitude and angle of the bus voltages, and then average the RMSE over all the time points. In each test we execute 20 Monte-Carlo simulations with 600 data points (equivalent to 20 seconds of data).

2.4.2 Example using static models

We first consider state estimation using static models, where the state at time point k is estimated based on measurements at the same time point only, and does not consider system dynamics. We compare our AM method with the traditional least-squares (LS) estimation.



(a) single snapshot



(b) 5 seconds

Figure 2.3: Comparison of state estimation results using AM, LS, and OLS on the IEEE 57-Bus system with 28 PMUs.

Under the general setup, we run Algorithm 2.1 for static state estimation. The estimation error at one time point is shown in Figure 2.3a⁵, where we compare the absolute magnitude and angle difference for the estimation of each bus state with the actual observed noise. We also include an “oracle least-squares” (OLS) case for comparison. The oracle estimation is obtained by assuming perfect knowledge of the phase mismatch, and using simple LS to estimate the states. Note that in the model we used real and imaginary parts of the complex-valued states and measurements. Here we convert them into magnitudes and angles for easier comparison. We also compare the magnitude and angle RMSE for five consecutive seconds (150 samples) in Figure 2.3b. When considering phase mismatch in the estimation model, and employing the AM algorithm for state estimation, we observe that the estimation errors in both magnitude and angle are reduced, and the errors from AM are comparable to the oracle case.

⁵For illustrative purposes, we use the IEEE 57-bus system as an example, to ensure adequate number of buses and avoid overwhelming graph complexity. Results for other systems are available in Table 2.3.

Table 2.3: Root mean-squared error (RMSE) of estimation using static models on different test systems, using least-squares estimation (LS), alternating minimization (AM), and LS with perfect phase mismatch information (OLS).

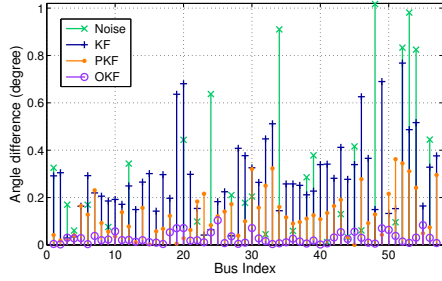
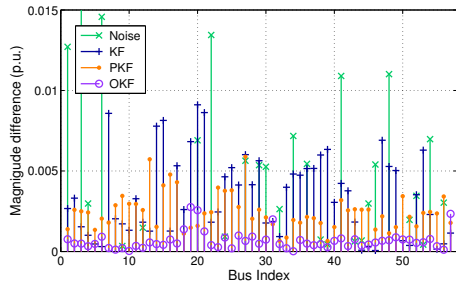
# of Buses N	# of PMUs M	Least Squares		Alternating Minimization		Oracle		Improvement	
		Magnitude (p.u.)	Angle (degrees)	Magnitude (p.u.)	Angle (degrees)	Magnitude (p.u.)	Angle (degrees)	Magnitude	Angle
14	4	5.020e-03	0.2827	2.976e-03	0.1781	1.454e-03	0.0832	40.72%	36.99%
	6	4.946e-03	0.2488	2.251e-03	0.1323	1.016e-03	0.0586	54.49%	46.81%
30	10	4.467e-03	0.2788	4.187e-03	0.2432	2.998e-03	0.1719	6.26%	12.75%
	16	2.918e-03	0.2202	2.180e-03	0.1402	1.640e-03	0.0941	25.29%	36.36%
57	17	5.865e-03	0.3519	5.718e-03	0.3320	4.219e-03	0.2416	2.50%	5.67%
	28	3.711e-03	0.2648	2.021e-03	0.1331	1.447e-03	0.0832	45.55%	49.75%
118	32	6.943e-03	0.4709	3.691e-03	0.2198	2.223e-03	0.1278	46.84%	53.31%
	54	5.875e-03	0.3864	1.933e-03	0.1180	7.257e-04	0.0419	67.09%	69.45%

We show the RMSE of the LS, AM, and OLS in all the test scenarios in Table 2.3. The improvement is defined by $(1 - \text{RMSE}_{\text{AM}}/\text{RMSE}_{\text{LS}}) \times 100\%$, which indicates the reduction in RMSE when using AM instead of LS. We observe that in all the test scenarios, AM provides amore accurate estimate than LS. The improvement is more significant when there are redundant observations from additional PMUs, as these provide a more accurate estimate of PMU phase mismatch. We also observe that when the number of PMUs increases, the AM estimation becomes closer to the oracle case.

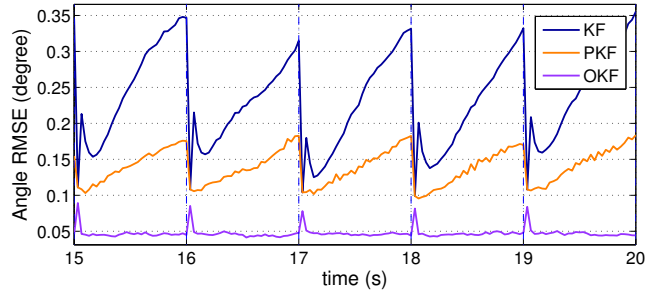
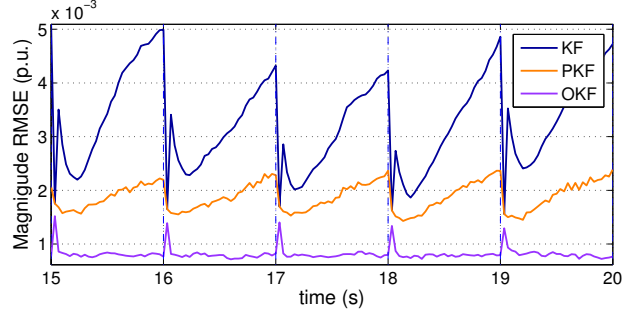
2.4.3 Example using dynamic models

We then consider the state estimation using dynamic models, where the state at time point k is estimated using all the measurements until the k th time point. We compare our parallel Kalman filtering (PKF) method with the traditional Kalman filter (KF) and the oracle Kalman filter case (OKF). The OKF assumes perfect knowledge of phase mismatch dynamics.

The state estimation error at one time point is shown in Figure 2.4a, and for five consecutive seconds is shown in Figure 2.4b. The figures indicate that by using PKF we can accurately estimate the phase mismatch, which significantly increases the estimation accuracy of the



(a) single snapshot



(b) 5 seconds

Figure 2.4: Comparison of state estimation results using KF, PKF, and OPK on the IEEE 57-Bus system with 28 PMUs.

Table 2.4: Root mean-squared error (RMSE) of estimation using dynamic models on different test systems, using traditional Kalman filter (KF), parallel Kalman filter (PKF), and Kalman filter with perfect phase mismatch information (OKF).

# of Buses N	# of PMUs M	Kalman Filter		Parallel Kalman Filter		Oracle		Improvement	
		Magnitude (p.u.)	Angle (degrees)	Magnitude (p.u.)	Angle (degrees)	Magnitude (p.u.)	Angle (degrees)	Magnitude	Angle
14	4	4.774e-03	0.2725	2.840e-03	0.1590	8.723e-04	0.0505	40.51%	41.63%
	6	4.655e-03	0.2336	2.720e-03	0.1395	7.000e-04	0.0402	41.57%	40.25%
30	10	3.061e-03	0.2048	2.563e-03	0.1630	1.287e-03	0.0732	16.29%	20.43%
	16	2.410e-03	0.1928	1.758e-03	0.1269	9.057e-04	0.0518	27.03%	34.17%
57	17	3.751e-03	0.2389	3.613e-03	0.2170	1.607e-03	0.0930	3.67%	9.19%
	28	3.200e-03	0.2375	1.868e-03	0.1392	8.163e-04	0.0471	41.64%	41.39%
118	32	6.010e-03	0.4127	3.277e-03	0.2098	9.473e-04	0.0547	45.47%	49.16%
	54	5.546e-03	0.3640	3.117e-03	0.2037	5.111e-04	0.0296	43.80%	44.02%

system state. Table 2.4 compares the RMSE using KF, PKF, and OKF in different test scenarios. Similar to the static case, the improvement is defined by $(1 - \text{RMSE}_{\text{PKF}}/\text{RMSE}_{\text{KF}}) \times 100\%$. We observe improvements similar to those in the static case.

2.4.4 Comparison between AM and PKF

In this subsection we compare the performance of AM and PKF. The performance of these two algorithms can be compared using Tables 2.3 and 2.4, as the two methods were tested using the same setup and data. PKF is more computationally efficient than AM, as it does not require iterative methods for the estimation at each time point. In many cases, PKF also provides slightly more accurate estimation than AM. However, the AM does not make assumptions on the dynamics of the system or the phase mismatch, and does not rely on an accurate system identification process. Therefore, the AM is more robust to inaccurate system identification than the PKF.

2.4.5 Effect of PMU specifications

Number of PMUs

The number of installed PMUs affects the estimation performance. In this experiment, we use a numerical example to show how the estimation error depends on the number of PMUs. We compare AM and LS in this subsection.

We assume the installed PMUs have the same synchronization accuracy, with $\sigma_{t,k} = 5 \mu\text{s}$. The optimal PMU placement involves minimizing the estimation error and other cost functions, and will be investigated in the next chapter. Here, we employ an *ad hoc* approach to determine PMU placement for illustration purposes. Let \mathcal{M} denote the set of buses with PMUs. We first select the minimum set of PMUs for full observation using the method from [48], and assign this minimum set to \mathcal{M} . Then we rank the remaining buses in decreasing order according to the number of their adjacent buses, and add new buses to \mathcal{M} following this order. In this way, buses with more adjacent buses have higher priority for being selected for PMU installation.

In Figure 2.5 we plot the estimation error as a function of the number of PMUs installed. We observe that the performance of AM improves as more PMUs are installed. The improvement is most significant when many PMUs are installed, i.e., our method approaches the oracle. This observation provides an intuition for optimal placement of PMUs when deciding the

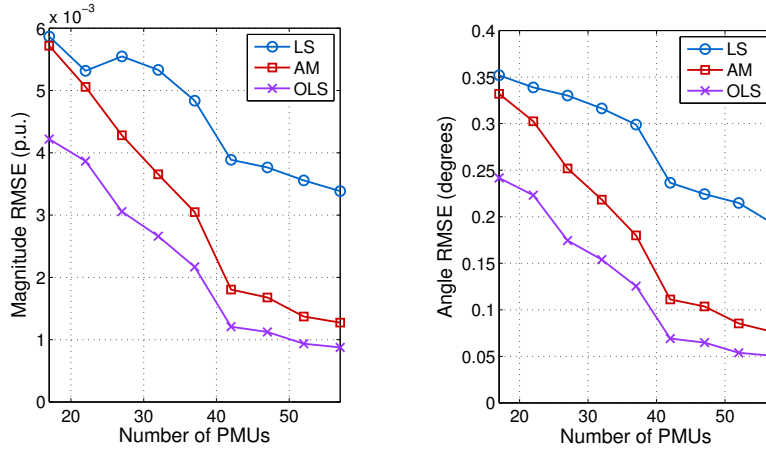


Figure 2.5: RMSE as a function of the number of PMUs installed on the IEEE 57-bus system.

number of PMUs to use. Note that in this example we set the PMU time synchronization specifications $\sigma_{t,k}$ to be the same for all PMUs. In practice, these specifications can be different, adding another design variable to the optimal PMU placement problem.

PMU synchronization specifications

We finally illustrate the effect of PMU synchronization specifications on the estimation results. The synchronization specifications include two aspects: the synchronization accuracy $\sigma_{t,k}$, and the time interval between two consecutive synchronizations.

In Figure 2.6 we plot the estimation error as a function of $\sigma_{t,k}$ on the IEEE 57-bus system with 28 PMUs installed. We observe that as $\sigma_{t,k}$ increases, the error also increases. However, since in our proposed methods we estimate the phase mismatch, the error of our proposed method is always lower than that of the least-squares approach. The improvement is more significant when $\sigma_{t,k}$ is large. The error of the oracle case does not change because the perfect phase mismatch is always assumed known.

Finally we illustrate the impact of synchronization intervals on the estimation performance. Different synchronization mechanisms have different synchronization intervals. For example, PMUs using a GPS clock are synchronized every 1 second, whereas according to the IEEE 1588 standard, the clocks are synchronized every few seconds. In Figure 2.7 we show the

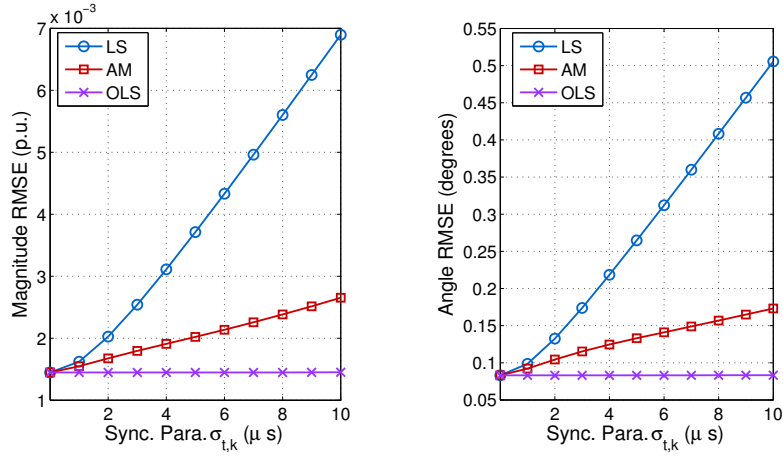


Figure 2.6: RMSE as a function of PMU synchronization accuracy $\sigma_{t,k}$ on the IEEE 57-bus system with 28 PMUs.

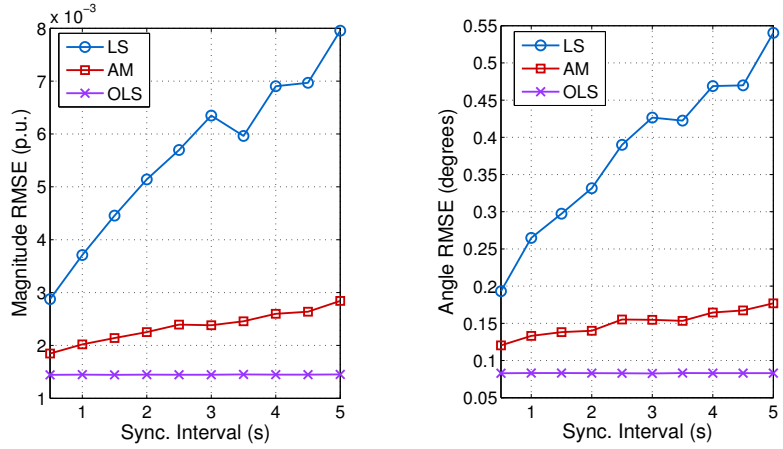


Figure 2.7: RMSE as a function of the time interval between consecutive synchronization on the IEEE 57-bus system.

estimation performance when the synchronization interval changes. We observe that as the time interval between consecutive synchronization increases, the estimation error also increases. However, using our methods, the increase in RMSE is much slower. We also performed experiments using KF and PKF, and the results are similar.

2.5 Summary

In this chapter we proposed a model for power system state estimation using PMUs with imperfect synchronization. We then proposed estimation algorithms using the static and dynamic models, with different assumptions. For estimation using the static models, we proposed to use alternating minimization to jointly estimate the phase mismatch of PMUs and the state of the power system. This approach does not rely on an accurate dynamic model for the state and phase mismatch, which is its main advantage over the dynamic counterpart. For estimation using the dynamic model, we proposed to couple two Kalman filters to estimate the phase mismatch and system state in parallel. Given a proper dynamic model for the system, this approach is more preferable as no iterations are required. Numerical examples showed that our methods significantly improve the estimation performance compared with traditional least-squares and Kalman filtering methods when the PMUs are not perfectly synchronized. We also illustrated the effect of PMU numbers and PMU synchronization on state estimation, and showed that when a sufficient number of PMUs are installed, the phase mismatch can be largely compensated for by using signal processing techniques, which encourages large scale deployment of imperfect PMUs. In addition, we showed that when using our methods, the estimation accuracy degrades slowly when the time interval between two synchronizations increases, which potentially encourages the use of alternative synchronization mechanisms with longer synchronization intervals.

Chapter 3

Performance Bound and Optimal PMU Placement

In the previous chapter, we proposed a model and developed multiple algorithms for power system state estimation using PMUs with phase mismatch. In this chapter, we consider the problem of PMU placement in power grids.⁶

3.1 Introduction

Since the number of PMUs currently deployed is limited, the optimal locations for PMU installation should be carefully selected. There has been active research on the topic of PMU placement [32, 48, 49, 50, 51, 52], using different methods under various models. Generally, there are two types of design criteria for PMU placement, topological observability and state estimation performance. In [48], the author proposed a generalized integer linear programming formulation for optimal PMU placement, and analyzed the cases for redundant PMU placement, full observability, and incomplete observability. This strategy, however, is based on binary observability, and does not take into account transmission line and other parameters. In [32], the PMU deployment was optimized based on estimation-theoretic criteria. Since the objective function is combinatorial, the authors proposed a convex relaxation approach to approximate the binary constraints, and formulated the optimization problem as

⁶Part of this chapter is based on P. Yang, Z. Tan, A. Wiesel and A. Nehorai, “Performance bounds and sensor placement for state estimation using PMUs with phase mismatch,” in *Proc. IEEE Power and Energy Society (PES) General Meeting*, Vancouver BC, Canada, July, 2013. © IEEE 2013.

a semi-definite program. The approximation, however, may not always be accurate, and the problem is difficult to solve for large systems. Li *et al.* [50] proposed a greedy approach for PMU placement, where they considered only voltage measurements from PMUs. They assumed PMU-instrumented buses were accurately measured, and the measurements were directly used as estimates. The objective was to minimize the estimation error for non-PMU buses. Later, they proposed an information-theoretic based design criterion in [51], where they used the mutual information between the PMU measurements and power system states. In both works, the PMU measurements are considered to be extremely accurate, which may not be realistic in practice.

In this chapter, we consider the problem of optimal PMU placement in the presence of measurement angle phase mismatch, using the bilinear observation model developed in Chapter 2. Heuristically it is difficult to tell how the placement of a PMU at a specific location in the power system would affect the state estimation error, not to say the effect of the PMU synchronization and measurement accuracy. Therefore, we derive the posterior Cramér-Rao bound (PCRB), which is a lower bound on the estimation error, and use it for optimal placement design. The bound does not depend on the instantaneous dynamical system state or time synchronization error, but rather depends on the network topology and parameters, and the statistics (probability distribution) of the states and time synchronization error.

We consider the classical experiment design criteria, and propose a greedy algorithm to solve the optimization problem. The use of greedy algorithms makes the optimization problem scalable for large systems. For some design criteria, it is possible to show that the objective functions are non-decreasing submodular [53], and therefore performance guarantees on the greedy algorithm exist. For the other design criteria, the objective functions of which are usually not submodular, numerical examples also indicate that the greedy solutions are close to the optimal solutions or lower bounds. We also perform state estimation on different placement strategies and demonstrate the improvement in estimation accuracy using the proposed PMU placement strategy.

The rest of this chapter is organized as follows. In Section 3.2 we derive the PCRB on the estimation error under the measurement model we derived in Section 2.2. In Section 3.3 we describe the PMU placement algorithms based on the derived PCRB. We show numerical

examples for PMU placement and the actual estimation error using different placement strategies in Section 3.4, and summarize the chapter in Section 3.5.

3.2 Performance bound

In this section, we analyze the estimation performance with a given PMU allocation under the model (2.28). This metric will be used in the next section for finding an optimal allocation. A direct analysis of the estimators in [54] is difficult. Instead, we derive a closed form bound on the estimation error based on the posterior Cramér-Rao bound (PCRB) [55, 56], which was presented in [57] as a global (or Bayesian) Cramér-Rao bound.

Let \mathbf{z} denote a vector of measured data, $\boldsymbol{\theta}$ be a vector of random parameters, $p_{\mathbf{z},\boldsymbol{\theta}}(\mathbf{Z}, \boldsymbol{\Theta})$ be the joint probability density of the pair $(\mathbf{z}, \boldsymbol{\theta})$, and $\hat{\boldsymbol{\theta}} = h(\mathbf{z})$ be an estimate of $\boldsymbol{\theta}$. Then we have the following inequality:

$$\mathbb{E} \{ [h(\mathbf{z}) - \boldsymbol{\theta}] [h(\mathbf{z}) - \boldsymbol{\theta}]^\top \} \geq \boldsymbol{\mathcal{I}}^{-1}, \quad (3.1)$$

where

$$\boldsymbol{\mathcal{I}} = \mathbb{E} \left[-\frac{\partial^2 \log p_{\mathbf{z},\boldsymbol{\theta}}(\mathbf{Z}, \boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}^2} \right] \quad (3.2)$$

is the Fisher information matrix (FIM) [58]. Therefore the PCRB, defined as the inverse of $\boldsymbol{\mathcal{I}}$, provides a lower bound on the covariance of any unbiased estimators for the unknown parameters under the measurement model.

The PMU installed on each bus takes measurements on its bus voltage and line currents. Let L denote the total number of measurements from all the PMUs. Since each measurement and each bus state consists of real and imaginary parts, the dimension of the matrix \mathbf{H} , which relates system states to PMU measurements, is then $2L \times 2N$. We decompose the matrix \mathbf{H} defined in (2.28) into the summation of N matrices, denoted as $\{\mathbf{H}_n\}_{n=1}^N$, where each matrix relates the measurements from the n th bus to the system states, with zeros padded in appropriate locations to maintain the same matrix dimensions as \mathbf{H} . In other words, matrix \mathbf{H}_n denotes a matrix with non-zero entries only in rows related to the measurements

from the n th bus, and

$$\mathbf{H} = \sum_{n=1}^N \mathbf{H}_n. \quad (3.3)$$

Let \mathcal{X} denote the set of buses with PMUs installed, and $\text{card}(\mathcal{X})$ denote the cardinality of \mathcal{X} , which is the total number of installed PMUs. We use a vector $\mathbf{x} = [x_1, \dots, x_N]^\top$ as an indicator vector denoting whether a PMU is installed on each bus. Each element x_n is binary with

$$x_n = \begin{cases} 1 & \text{if bus } n \in \mathcal{X}, \\ 0 & \text{otherwise.} \end{cases} \quad (3.4)$$

Under these definitions, we can rewrite the measurement model (2.28) with a given PMU placement profile \mathcal{X} as

$$\begin{aligned} \mathbf{z} &= \mathbf{f}(\mathbf{s}, \boldsymbol{\theta}) + \boldsymbol{\epsilon} \\ &= \sum_{n=1}^N x_n \mathbf{H}_n \mathbf{s} + \sum_{n=1}^N x_n \theta_n \mathbf{G}_n \mathbf{s} + \boldsymbol{\epsilon}, \end{aligned} \quad (3.5)$$

where we omit the subscript k for notational simplicity. The measurement noise $\boldsymbol{\epsilon}$ follow zero-mean Gaussian distribution with

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_\epsilon). \quad (3.6)$$

We assume the prior distributions for the system state \mathbf{s} and phase mismatch $\boldsymbol{\theta}$ are also Gaussian, with

$$\mathbf{s} \sim \mathcal{N}(\boldsymbol{\mu}_s, \mathbf{R}_s), \quad (3.7)$$

$$\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_\theta), \quad (3.8)$$

where the prior for \mathbf{s} can be obtained from SCADA estimates [32]. The phase mismatches of different sensors are assumed to be uncorrelated, with

$$\mathbf{R}_\theta = \begin{bmatrix} \sigma_{\theta,1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{\theta,2}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{\theta,N}^2 \end{bmatrix}. \quad (3.9)$$

This is a reasonable assumption since in practice, the PMUs are widely separated geographically, and the phase mismatch in measurements come from various sources as described in Section 2.1.

The probability density function of measurement \mathbf{z} , given system state \mathbf{s} and phase mismatch $\boldsymbol{\theta}$, is then

$$p(\mathbf{z}|\mathbf{s}, \boldsymbol{\theta}) = \frac{1}{(2\pi)^L |\mathbf{R}_\epsilon|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\mathbf{z} - \mathbf{f}(\mathbf{s}, \boldsymbol{\theta}))^\top \mathbf{R}_\epsilon^{-1} (\mathbf{z} - \mathbf{f}(\mathbf{s}, \boldsymbol{\theta})) \right\}. \quad (3.10)$$

To calculate the joint probability of \mathbf{s} and $\boldsymbol{\theta}$ given \mathbf{z} , we need the joint prior distribution of \mathbf{s} and $\boldsymbol{\theta}$. Since the system state is independent of the phase mismatch, we have

$$p(\mathbf{s}, \boldsymbol{\theta}) = p(\mathbf{s})p(\boldsymbol{\theta}). \quad (3.11)$$

The joint probability of \mathbf{s} , $\boldsymbol{\theta}$, and \mathbf{z} can then be calculated from

$$p(\mathbf{s}, \boldsymbol{\theta}, \mathbf{z}) = p(\mathbf{z}|\mathbf{s}, \boldsymbol{\theta})p(\mathbf{s})p(\boldsymbol{\theta}). \quad (3.12)$$

We can further obtain the log likelihood function as

$$\begin{aligned} l(\mathbf{s}, \boldsymbol{\theta}, \mathbf{z}) = & -\frac{1}{2} (\mathbf{z} - \mathbf{f}(\mathbf{s}, \boldsymbol{\theta}))^\top \mathbf{R}_\epsilon^{-1} (\mathbf{z} - \mathbf{f}(\mathbf{s}, \boldsymbol{\theta})) \\ & -\frac{1}{2} \boldsymbol{\theta}^\top \mathbf{R}_\theta^{-1} \boldsymbol{\theta} - \frac{1}{2} (\mathbf{s} - \boldsymbol{\mu}_s)^\top \mathbf{R}_s^{-1} (\mathbf{s} - \boldsymbol{\mu}_s) \\ & + \text{constant}. \end{aligned} \quad (3.13)$$

For simplicity of notation, we denote $\boldsymbol{\xi} = [\mathbf{s}^\top, \boldsymbol{\theta}^\top]^\top$. The Fisher information matrix (FIM) can then be written as

$$\begin{aligned}\mathcal{I}(\boldsymbol{\xi}) &= \mathbb{E} \left\{ -\frac{\partial^2 \mathbf{l}(\boldsymbol{\xi}, \mathbf{z})}{\partial \boldsymbol{\xi}^2} \right\} \\ &= \mathbb{E} \left\{ - \begin{bmatrix} \frac{\partial^2 \mathbf{l}(\boldsymbol{\xi}, \mathbf{z})}{\partial \mathbf{s}^2} & \frac{\partial^2 \mathbf{l}(\boldsymbol{\xi}, \mathbf{z})}{\partial \mathbf{s} \partial \boldsymbol{\theta}} \\ \frac{\partial^2 \mathbf{l}(\boldsymbol{\xi}, \mathbf{z})}{\partial \boldsymbol{\theta} \partial \mathbf{s}} & \frac{\partial^2 \mathbf{l}(\boldsymbol{\xi}, \mathbf{z})}{\partial \boldsymbol{\theta}^2} \end{bmatrix} \right\}.\end{aligned}\quad (3.14)$$

From (3.5) and (3.13), we derive that

$$\frac{\partial^2 \mathbf{l}(\boldsymbol{\xi}, \mathbf{z})}{\partial \mathbf{s}^2} = -(\mathbf{H}_x + \mathbf{F})^\top \mathbf{R}_\epsilon^{-1} (\mathbf{H}_x + \mathbf{F}) - \mathbf{R}_s^{-1}, \quad (3.15)$$

$$\frac{\partial^2 \mathbf{l}(\boldsymbol{\xi}, \mathbf{z})}{\partial \mathbf{s} \partial \boldsymbol{\theta}} = -(\mathbf{H}_x + \mathbf{F})^\top \mathbf{R}_\epsilon^{-1} \mathbf{L} + \mathbf{D}, \quad (3.16)$$

$$\frac{\partial^2 \mathbf{l}(\boldsymbol{\xi}, \mathbf{z})}{\partial \boldsymbol{\theta} \partial \mathbf{s}} = -\mathbf{L}^\top \mathbf{R}_\epsilon^{-1} (\mathbf{H}_x + \mathbf{F}) + \mathbf{D}^\top, \quad (3.17)$$

$$\frac{\partial^2 \mathbf{l}(\boldsymbol{\xi}, \mathbf{z})}{\partial \boldsymbol{\theta}^2} = -\mathbf{L}^\top \mathbf{R}_\epsilon^{-1} \mathbf{L} - \mathbf{R}_\theta^{-1}, \quad (3.18)$$

where

$$\mathbf{H}_x = \sum_{n=1}^N x_n \mathbf{H}_n, \quad (3.19)$$

$$\mathbf{D} = [\cdots, x_n \mathbf{G}_n^\top \mathbf{R}_\epsilon^{-1} (\mathbf{z} - \mathbf{f}(\boldsymbol{\xi})), \cdots], \quad (3.20)$$

$$\mathbf{F} = \sum_{n=1}^N x_n \theta_n \mathbf{G}_n, \quad (3.21)$$

$$\mathbf{L} = [\cdots, x_n \mathbf{G}_n \mathbf{s}, \cdots]. \quad (3.22)$$

In the derivations we used the properties that

$$\mathbf{H}_n^\top \mathbf{H}_m = \mathbf{0}, \quad \mathbf{G}_n^\top \mathbf{G}_m = \mathbf{0}, \quad \text{for } n \neq m, \quad (3.23)$$

and that

$$x_n^2 = x_n, \quad \text{for } x_n \in \{0, 1\}. \quad (3.24)$$

We then take expectations of (3.15)-(3.18) to obtain the FIM in the following form:

$$\mathcal{I}(\boldsymbol{\xi}) = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{12}^\top & \mathbf{C}_{22} \end{bmatrix}, \quad (3.25)$$

where the submatrices \mathbf{C}_{mn} are detailed below.

The matrix \mathbf{C}_{11} has the form

$$\mathbf{C}_{11} = \sum_{n=1}^N x_n \mathbf{H}_n^\top \mathbf{R}_\epsilon^{-1} \mathbf{H}_n + \sum_{n=1}^N x_n \sigma_{\theta,n}^2 \mathbf{G}_n^\top \mathbf{R}_\epsilon^{-1} \mathbf{G}_n + \mathbf{R}_s^{-1}. \quad (3.26)$$

Let $(\mathbf{C}_{12})_n$ denote the n th column of \mathbf{C}_{12} , and then

$$(\mathbf{C}_{12})_n = x_n \mathbf{H}_n^\top \mathbf{R}_\epsilon^{-1} \mathbf{G}_n \boldsymbol{\mu}_s. \quad (3.27)$$

The matrix \mathbf{C}_{22} has the form

$$\mathbf{C}_{22} = \begin{bmatrix} x_1 d_1 & 0 & \cdots & 0 \\ 0 & x_2 d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_N d_N \end{bmatrix} + \mathbf{R}_\theta^{-1} \quad (3.28)$$

where

$$d_n = \text{tr}(\mathbf{G}_n^\top \mathbf{R}_\epsilon^{-1} \mathbf{G}_n \mathbf{R}_s) + \boldsymbol{\mu}_s^\top \mathbf{G}_n^\top \mathbf{R}_\epsilon^{-1} \mathbf{G}_n \boldsymbol{\mu}_s. \quad (3.29)$$

Under the assumption of uncorrelated phase mismatch, matrix \mathbf{C}_{22} is then a diagonal matrix, with the n th diagonal element given by

$$(\mathbf{C}_{22})_{n,n} = x_n d_n + \frac{1}{\sigma_{\theta,n}^2}. \quad (3.30)$$

The PCRB for the state \mathbf{s} is the upper left block of the inverse of $\mathcal{I}(\boldsymbol{\xi})$, which can be obtained from block matrix inversion:

$$\text{PCRB}_s = (\mathbf{C}_{11} - \mathbf{C}_{12} \mathbf{C}_{22}^{-1} \mathbf{C}_{12}^\top)^{-1} \triangleq \mathcal{I}_{ss}^{-1}. \quad (3.31)$$

We now simplify the formula for $\mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{12}^\top$. Denote

$$\mathbf{v}_n = \mathbf{H}_n^\top \mathbf{R}_\epsilon^{-1} \mathbf{G}_n \boldsymbol{\mu}_s, \quad (3.32)$$

$$\varphi_n = x_n d_n + \frac{1}{\sigma_{\theta,n}^2}, \quad (3.33)$$

$$\gamma_n = 1 / \left(d_n + \frac{1}{\sigma_{\theta,n}^2} \right), \quad (3.34)$$

then $\mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{12}^\top$ can be written as

$$\begin{aligned} \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{12}^\top &= [x_1 \mathbf{v}_1, x_2 \mathbf{v}_2, \dots, x_N \mathbf{v}_N] \begin{bmatrix} \varphi_1^{-1} & 0 & \dots & 0 \\ 0 & \varphi_2^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \varphi_N^{-1} \end{bmatrix} \begin{bmatrix} x_1 \mathbf{v}_1^\top \\ x_2 \mathbf{v}_2^\top \\ \vdots \\ x_1 \mathbf{v}_1^\top \end{bmatrix} \\ &= \sum_{n=1}^N x_n^2 \varphi_n^{-1} \mathbf{v}_n \mathbf{v}_n^\top \\ &= \sum_{n=1}^N x_n \gamma_n \mathbf{v}_n \mathbf{v}_n^\top, \end{aligned} \quad (3.35)$$

where the last equation holds since $x_n \in \{0, 1\}$. When $x_n = 0$, the corresponding term in the summation equals zero. When $x_n = 1$, the corresponding term in the summation equals $\gamma_n \mathbf{v}_n \mathbf{v}_n^\top$. Therefore, the \mathcal{I}_{ss} can be rewritten as

$$\mathcal{I}_{ss} = \sum_{n=1}^N x_n \mathbf{A}_n + \mathbf{R}_s^{-1}, \quad (3.36)$$

where

$$\mathbf{A}_n = \mathbf{H}_n^\top \mathbf{R}_\epsilon^{-1} \mathbf{H}_n + \sigma_{\theta,n}^2 \mathbf{G}_n^\top \mathbf{R}_\epsilon^{-1} \mathbf{G}_n - \gamma_n \mathbf{v}_n \mathbf{v}_n^\top. \quad (3.37)$$

We now have derived and simplified the PCRB for state estimation. In the next section we will discuss PMU placement based on the derived PCRB.

3.3 PMU placement algorithms

3.3.1 Problem formulation

The PCRB calculated in the previous section is a function of \mathbf{x} , which is determined by the set \mathcal{X} of buses installed with PMUs. In the rest of the chapter, we will use the notations $f(\mathbf{x})$ and $f(\mathcal{X})$ interchangeably. A placement \mathbf{x}^* is considered to be more preferable than $\mathbf{x} \neq \mathbf{x}^*$ when $\mathcal{I}_{ss}^{-1}(\mathbf{x}^*) \prec \mathcal{I}_{ss}^{-1}(\mathbf{x})$. Since the ordering over non-negative definite matrices is partial, the experiment design involves the optimization of a scalar-valued function of $\mathcal{I}_{ss}^{-1}(\mathbf{x})$ [59, 60]. This approach was also used in other optimal placement papers [32, 50]. Commonly used criteria include:

A-optimal: $f_A(\mathbf{x}) = \text{tr}(\mathcal{I}_{ss}^{-1}(\mathbf{x}))$,

D-optimal: $f_D(\mathbf{x}) = \log \det(\mathcal{I}_{ss}^{-1}(\mathbf{x}))$,

E-optimal: $f_E(\mathbf{x}) = \lambda_{\max}(\mathcal{I}_{ss}^{-1}(\mathbf{x}))$,

M-optimal: $f_M(\mathbf{x}) = \max_n \left\{ (\mathcal{I}_{ss}^{-1}(\mathbf{x}))_{n,n} \right\}$,

where $\text{tr}(\cdot)$ denotes the trace of a matrix, $\lambda_{\max}(\cdot)$ denotes the maximum eigenvalue of a matrix, and $\det(\cdot)$ denotes matrix determinant. The objective of the optimal placement problem then becomes minimizing the scalar functions of the PCRB, and the choice of the specific scalar function depends on specific applications and design criteria. One motivation for using the **A-optimal** criterion is that, as proved in [61], if an optimal solution (in the matrix sense) exists, then the optimal solution can be found using the **A-optimal** criterion.

In practice, full PMU observability is often preferred. The reasons are threefold. First, SCADA and PMUs have different sample rates, and SCADA measurements are not time synchronized. Directly combining these two types of measurements for state estimation results in time-skewness problems [20]. Second, although the SCADA estimation can be used as prior information for PMU based state estimation [32, 54], joint estimation without PMU observability does not fully exploit the benefit of real-time monitoring enabled by PMUs. Third, more PMUs will eventually be installed in the grid, and it will be possible in

the future to estimate system state purely based on PMU measurements. As an example, the 13-bus Central New York State Power System has already been made observable by the installation of 6 multi-channel PMUs [25].

When full PMU observability is a requirement, the first step is to select PMU locations to guarantee observability of the system. Both the buses with PMUs and the adjacent buses can be observed. Let \mathbf{M}_A denote the adjacency matrix of the bus-branch network. Then the indicator vector \mathbf{x}_{\min} , corresponding to a minimum set of PMUs \mathcal{X}_{\min} for topological observability [48], is obtained by solving

$$\begin{aligned} \mathbf{x}_{\min} &= \arg \min_{\mathbf{x}} \mathbf{x}^\top \mathbf{1}_N \\ &\text{subject to } \mathbf{M}_A \mathbf{x} \geq \mathbf{1}_N, \\ &\mathbf{x} \in \{0, 1\}^N, \end{aligned} \tag{3.38}$$

where $\mathbf{1}_N$ denotes an all-one vector with dimension N . This binary integer optimization can be efficiently solved for a moderate size system. For large systems, recent works, e.g., [62], have proposed approaches to solve this binary integer programming problem. When multiple solutions exist for full observability, the best placement profile is selected by comparing the corresponding design criteria.

The next step is to find optimal locations for additional PMUs, based on the minimum set, to minimize the objective function of the selected design criterion. Therefore the optimization problem for PMU placement can be formulated as

$$\begin{aligned} &\min_{\mathbf{x}} f_Q(\mathbf{x}) \\ &\text{subject to } \mathcal{X}_{\min} \subseteq \mathcal{X}, \\ &\mathbf{x}^\top \mathbf{1} = M, \\ &\mathbf{x} \in \{0, 1\}^N, \end{aligned} \tag{3.39}$$

where $Q \in \{A, D, E, M\}$ denotes the specified design criterion, and M denotes the total number of PMUs.

Remark 3.1. *Full observability is not required under the Bayesian framework. If, in practice, observability is not a compulsory requirement, or the number of PMUs is too limited to*

achieve full observability, \mathcal{X}_{\min} can be set as \emptyset . In this case, there is no need to solve the binary integer programming defined in (3.38).

Remark 3.2. Critical buses that must be installed with or made observable by PMUs can be guaranteed by configuring corresponding parameters in (3.38) and/or (3.39).

3.3.2 Greedy algorithm

The problem (3.39) is combinatorial and intractable when N and M are large. We propose to use a greedy approach (Algorithm 3.1) to solve the problem.

Table 3.1: Greedy algorithm for PMU placement.

Algorithm 3.1: Greedy algorithm for PMU placement.
Data: design objective function f_Q , minimum set of buses \mathcal{X}_{\min} , number of PMUs.
Result: PMU placement \mathcal{X} .
begin
initialize $\mathcal{X} \leftarrow \mathcal{X}_{\min}$;
repeat
$u^* \leftarrow \arg \min_{u \notin \mathcal{X}} f_Q(\mathcal{X} \cup \{u\})$;
$\mathcal{X} \leftarrow \mathcal{X} \cup \{u^*\}$;
until $\text{card}(\mathcal{X}) = M$;
return \mathcal{X} ;
end

The greedy algorithm has $O(N)$ complexity to find the location of one PMU, and thus is scalable for large scale systems. In practice, PMUs are often not installed all at once in the grid in practice. The greedy algorithm fits well in this scenario, as in each iteration it find the best location for the next PMU.

When the objective function is monotonic and submodular, a bound exists for the performance of the greedy algorithm. A set function $f(\mathcal{X}) : \mathcal{X} \rightarrow \mathbb{R}$ is submodular if it satisfies two criteria:

(i) $f(\emptyset) = 0$;

$$(ii) f(\mathcal{X} \cup \mathcal{Y}) + f(\mathcal{X} \cap \mathcal{Y}) \leq f(\mathcal{X}) + f(\mathcal{Y}) \text{ for all sets } \mathcal{X} \text{ and } \mathcal{Y}.$$

The second criterion is equivalent to $f(\mathcal{X} \cup \{i\}) - f(\mathcal{X}) \geq f(\mathcal{Y} \cup \{i\}) - f(\mathcal{Y})$ for all $\mathcal{X} \subseteq \mathcal{Y}$ and $i \notin \mathcal{Y}$. Nemhauser *et. al.* [53] proved the following bound on the performance of a greedy algorithm applied to maximization of submodular and nondecreasing set functions.

Proposition 3.1. (Nemhauser [53]) *Let $f(\mathcal{X})$ be a submodular and nondecreasing set function. For the optimization problem $\max_{\text{card}(\mathcal{X}) \leq M} f(\mathcal{X})$ where $M \geq 1$, the solution from the greedy algorithm is always at least $1 - [(M-1)/M]^M$ times the optimal solution. This bound can be achieved for each M , and the limiting case is $(e-1)/e$.*

In our problem, for the **A-optimal** and **D-optimal** criteria, we can rewrite the optimization objectives as

$$\tilde{g}_A(\mathcal{X}) = -\text{tr}(\mathcal{I}_{ss}^{-1}(\mathbf{x})) \quad (3.40)$$

$$\tilde{g}_D(\mathcal{X}) = \log \det(\mathcal{I}_{ss}(\mathbf{x})). \quad (3.41)$$

Since the minimum set \mathcal{X}_{\min} has to be a subset of the selected buses, we redefine functions $g_A(\mathcal{X})$ and $g_D(\mathcal{X})$ based on $\tilde{g}_A(\mathcal{X})$ and $\tilde{g}_D(\mathcal{X})$ to make this constraint implicit. To be more specific, the input \mathcal{X} of these two functions are revised to be the set of *additional buses* instead of the entire set of buses installed with PMUs. In addition, since a submodular function has to be zero when the input is an empty set, we need to add a constant to each of the two objective functions. The objective functions are then rewritten as

$$g_A(\mathcal{X}) = -\text{tr}(\mathcal{I}_{ss}^{-1}(\mathcal{X} \cup \mathcal{X}_{\min})) + \text{tr}(\mathcal{I}_{ss}^{-1}(\mathcal{X}_{\min})) \quad (3.42)$$

$$g_D(\mathcal{X}) = \log \det(\mathcal{I}_{ss}(\mathcal{X} \cup \mathcal{X}_{\min})) - \log \det(\mathcal{I}_{ss}(\mathcal{X}_{\min})). \quad (3.43)$$

The added constants do not change the solution of the optimization problems. Before we proceed, we introduce the following assumption.

Assumption 3.1. *The matrices \mathbf{A}_n , $n = 1 \dots N$, as defined in (3.37) are symmetric non-negative definite.*

This assumption can be easily justified. The symmetry is directly obtained from the definition. Since $\sigma_{\theta,n}^2$ is small in practice, the first term in (3.37) is much larger than the second and third term, and therefore \mathbf{A}_n is positive semidefinite. An intuitive interpretation for this

assumption is that adding a PMU will increase the “information” for state estimation. The assumption can also be verified numerically.

Let \mathcal{X} and \mathcal{Y} denote two subset of buses with PMUs, and $\mathcal{X} \subseteq \mathcal{Y}$. Then denote $\mathbf{A}_{[\mathcal{X}]} = \mathcal{I}_{ss}(\mathcal{X})$ and $\mathbf{A}_{[\mathcal{Y}]} = \mathcal{I}_{ss}(\mathcal{Y})$. For the submodularity condition of g_A to hold, the following condition is sufficient:

$$\text{tr} \left(\mathbf{A}_{[\mathcal{X}]}^{-1} (\mathbf{A}_{[\mathcal{X}]} + \mathbf{A}_i)^{-1} \mathbf{A}_i \right) \geq \text{tr} \left(\mathbf{A}_{[\mathcal{Y}]}^{-1} (\mathbf{A}_{[\mathcal{Y}]} + \mathbf{A}_i)^{-1} \mathbf{A}_i \right). \quad (3.44)$$

The derivation for (3.44) is included in Appendix A. In general, Assumption 3.1 does not guarantee $g_A(\mathcal{X})$ to be submodular. However, numerical tests indicate that the submodularity conditions are satisfied in the test cases. An intuitive argument can be found in [50], where the authors used an additional assumption that the columns of the matrix \mathbf{H} are *nearly orthogonal*. Under this assumption, the matrices in (3.44) are nearly diagonal, and it is easy to justify that (3.44) holds.

The submodularity of $g_D(\mathcal{X})$ can be rigorously proved (see Appendix B), and we have the following proposition.

Proposition 3.2. *Under Assumption 3.1, the function $g_D(\mathcal{X})$ as defined in (3.43), with $\mathcal{I}_{ss}(\mathcal{X})$ defined in (3.36), is a submodular and non-decreasing function.*

Therefore, using the **D-optimal** design criteria, the performance bound for greedy algorithm solutions described in Proposition 3.1 applies to the revised objective function $g_D(\mathcal{X})$. We then have the following performance bounds on the greedy solutions for (3.39) in the **D-optimal** design case.

Theorem 3.1. *For the **D-optimal** design, let f_D^g and f_D^* denote the objective function evaluated at the greedy solution and the optimal solution to problem (3.39), respectively, and f_D^{\min} denote the value of the objective function evaluated at the minimum set \mathcal{X}_{\min} . The objective function of the greedy solution is bounded by*

$$f_D^g \leq \left(1 - \frac{1}{e}\right) f_D^* + \frac{1}{e} f_D^{\min}. \quad (3.45)$$

Proof. From Propositions 3.1, we have that

$$g_D(\mathcal{X}_g) \geq \left(1 - \frac{1}{e}\right) g_D(\mathcal{X}_*), \quad (3.46)$$

where \mathcal{X}_g and \mathcal{X}_* denote the *additional* set of buses obtained from the greedy algorithm and the optimal solution, respectively. Plugging (3.43) into (3.46), we obtain that

$$\begin{aligned} & \log \det(\mathcal{I}_{ss}(\mathcal{X}_g \cup \mathcal{X}_{\min})) - \log \det(\mathcal{I}_{ss}(\mathcal{X}_{\min})) \\ & \geq \left(1 - \frac{1}{e}\right) \left(\log \det(\mathcal{I}_{ss}(\mathcal{X}_* \cup \mathcal{X}_{\min})) - \log \det(\mathcal{I}_{ss}(\mathcal{X}_{\min})) \right). \end{aligned} \quad (3.47)$$

Following the definitions in Theorem 3.1, we then obtain that

$$-f_D^g + f_D^{\min} \geq \left(1 - \frac{1}{e}\right) (-f_D^* + f_D^{\min}), \quad (3.48)$$

which implies that

$$f_D^g \leq \left(1 - \frac{1}{e}\right) f_D^* + \frac{1}{e} f_D^{\min} \quad (3.49)$$

after simple calculations. □

As we argued above, although the **A-optimal** objective function is not rigorously proved to be submodular, the submodularity condition is often satisfied. Therefore the performance bound in Theorem 3.1 also holds for the **A-optimal** design criterion. For the **E-optimal** and **M-optimal** criteria, the objective functions are usually not submodular, and therefore we cannot apply the performance bound on the greedy algorithm in these two cases. However, the greedy algorithm also works well in both cases, as we will illustrate in the numerical examples.

3.3.3 Other placement strategies

We will compare the greedy algorithm with other heuristics and optimization methods, including random placement, rank by branch, exhaustive search, and convex relaxation [32]. A brief description of each method is given below.

Random placement

This is a non-optimized placement strategy. In this case, the buses for PMU installation are randomly selected sequentially. In other words, let \mathcal{M} denote the buses already installed with PMUs, and Ω denote the set of all the buses. The next bus will be selected randomly from the set $\Omega \setminus \mathcal{M}$ following a uniform distribution.

Rank by branch

This is a heuristic approach for placing PMUs, based on how many branches a bus is connected to. The candidate buses are ranked by the number of branches they connect with. Buses with more connected branches have higher priority for PMU installation. To be specific, the next selected bus will be the bus with the greatest number of connected branches.

Exhaustive search

This is the optimal case, where all the combinations of PMUs are searched and the optimal combination is selected. However, this approach is NP-hard [63], and can be solved only for small systems.

Convex relaxation

In this approach, the binary constraint in (3.39) is relaxed to a continuous constraint. To be specific, the problem is relaxed as follows:

$$\begin{aligned}
& \min_{\mathbf{x}} f_Q(\mathbf{x}) \\
& \text{subject to } \mathcal{X}_{\min} \subseteq \mathcal{X}, \\
& \quad \mathbf{x}^\top \mathbf{1} = M, \\
& \quad \mathbf{0} \leq \mathbf{x} \leq \mathbf{1}.
\end{aligned} \tag{3.50}$$

By replacing the constraint $\mathbf{x} \in \{0, 1\}^N$ with $\mathbf{x} \in [0, 1]^N$, the relaxed version of problem (3.39) can be formulated as a semidefinite programming (SDP) problem [32], which can be solved efficiently for problems of a moderate size. The resulting vector \mathbf{x} will be a vector with each element within the interval $[0, 1]$. The M buses corresponding to the largest M entries in \mathbf{x} will be selected. This convex relaxation method provides a solution that is optimal for (3.50), but not necessarily for the original problem (3.39), as will be illustrated in the numerical examples. However, the relaxed problem can provide a lower bound on the original problem (3.39), which can be used to evaluate different placements.

3.4 Numerical examples

3.4.1 Comparison of different PMU placements

We first show the placement results using different design criteria and different placement algorithms. The PMUs are assumed to have the same time-synchronization accuracy, where the time delay has a standard deviation of $25\mu s$ (equivalent to 0.54 degrees of phase mismatch at $f_c = 60\text{Hz}$). The additive measurement noise of each PMU is i.i.d. Gaussian with a standard deviation of $\sigma_\epsilon = 1 \times 10^{-3}\text{p.u.}$. The standard deviation for bus voltage is selected to be $\sigma_s = 0.01\text{p.u.}$.

In Fig.3.1 we show the results of the five placement strategies using the **A-**, **D-**, **E-** and **M-optimal** criteria on the IEEE 14-bus system. We enforce the full observability requirement in this numerical example. Note that for the 14-bus system, at least 4 PMUs are needed for observability, and therefore the minimum number of PMUs in Fig. 3.1 is 4.

For the random placement case, we ran the algorithm 10 times and took the average of the objective functions. A single trial of random placement does not result in the smooth curve shown in Fig.3.1. We observe that the greedy algorithm produces similar results to the optimal solution through exhaustive search. The random selection is in general worse than the greedy algorithm. The rank-by-branch algorithm is the worst of the tested algorithms. One possible explanation is that it employs only the information of the network connectivity, but fails to consider the contribution of each branch to the estimation accuracy. Namely, within a densely connected cluster of buses, each bus will have a high priority to be selected, which will reduce the estimation error of the buses within this cluster. However, for buses outside the cluster, the improvement in estimation accuracy is limited. In this test case, the convex relaxation algorithm produces results similar to those of the greedy algorithm most of the time.

In Fig.3.2 we show results on a larger system - the IEEE 30-bus system. We observe that the greedy algorithm performs well most of the time. However, as noted in the **A-**, **E-**, and **M-optimal** design cases, the first few points of results from the convex relaxation approach are not quite satisfactory. There are multiple “small” entries in the resulting vector \mathbf{x} , and only part of them are selected due to the limitation on the number of PMUs. Therefore the solution is not optimal for the original binary programming problem. The results from the greedy algorithm are very close to the lower bound obtained from the convex relaxation, which demonstrates that the greedy solution is close to the optimal solution.

For larger systems with hundreds or even thousands of buses, the greedy algorithm is still efficient, because for each additional PMU, the time complexity to find its best location is $O(|\Omega \setminus \mathcal{M}|)$. However, the complexity of the convex relaxation will increase significantly since a much larger semi-definite program has to be solved.

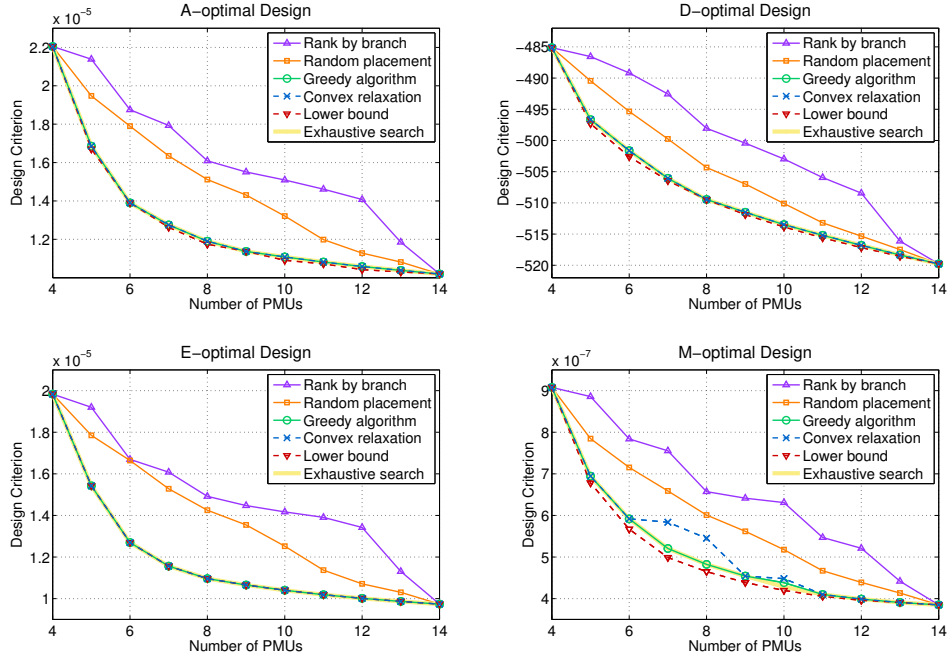


Figure 3.1: Comparison of PMU placement for IEEE 14-bus system using different design criteria and different algorithms.

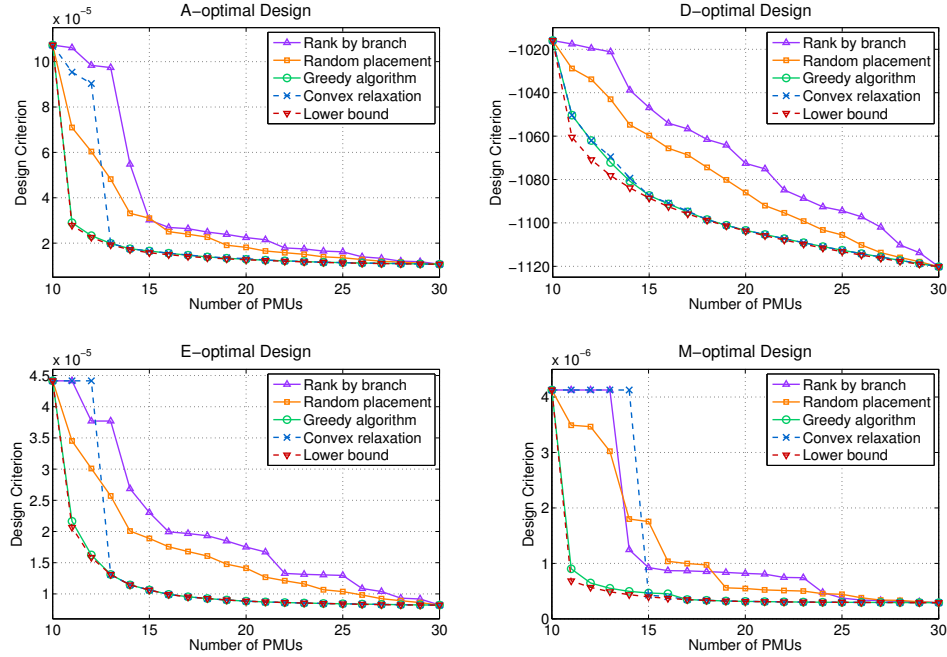


Figure 3.2: Comparison of PMU placement for IEEE 30-bus system using different design criteria and different algorithms.

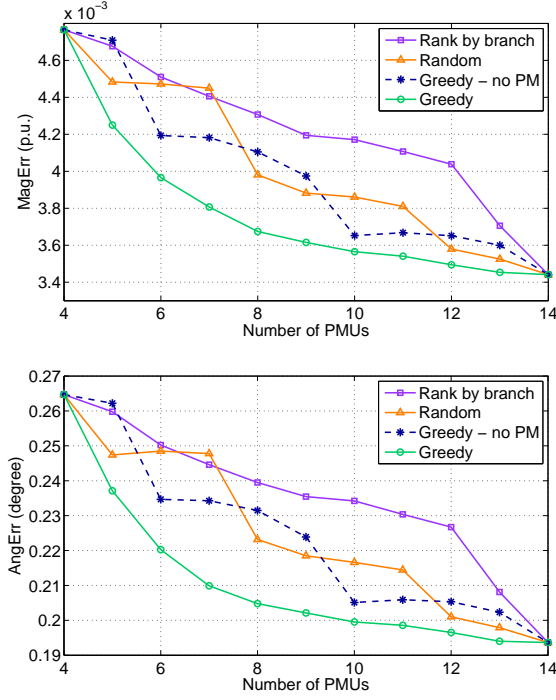


Figure 3.3: Comparison of state estimation results using different PMU placement profiles on the IEEE 14-bus system.

3.4.2 State estimation results under different placement

In Fig. 3.3, we show the estimation performance using the estimator we proposed in [54], under placement profiles from different strategies based on the **A-optimal** criterion. The estimator employs the alternating minimization approach described in Section 2.3.1, where the state and phase mismatch are estimated iteratively. Since the greedy, convex relaxation, and exhaustive search algorithms all provide the same placements in this case, we show only the estimation results using the greedy placement, which represents the other two placements also. The results indicate trends similar to the comparison of theoretical performance bounds shown in Fig. 3.1. Overall, the placement from the greedy algorithm provides the best estimation results. We also include a placement profile from the greedy algorithm, without considering the contribution of phase mismatch in the design (the no PM case in Fig. 3.3), and observe that the estimation performance is improved when taking PMU phase mismatch into account for optimal PMU placement.

3.5 Summary

In this chapter we considered the problem of placing imperfectly synchronized PMUs in the power grid. We derived the posterior Cramér-Rao bound on the state estimation error using a measurement model that we proposed in Section 2.2. Based on the PCRB, we proposed to employ a greedy algorithm for PMU placement using different design criteria. We showed that for some commonly used design criteria, the optimization problem can be written as maximizing a submodular and monotonic set function. Therefore a performance guarantee exists for these cases. For other cases, although the objective functions are usually not submodular, numerical examples illustrated that the greedy algorithm also provides good results.

We compared the greedy algorithm with other strategies, including the method that employs the idea of convex relaxation, and concluded that the greedy algorithm works well in all the tested cases, and the greedy solution is close to the optimal solution (through exhaustive search) and the lower bound (through relaxed optimization). We also compared state estimation performance for different placement strategies, using the estimator we proposed in Section 2.3.1. The results demonstrated that the greedy algorithm is efficient, and that taking PMU phase mismatch into account improves the PMU placement.

Chapter 4

Parallel Load Scheduling with Renewable Distributed Generation

In the following two chapters, we will consider the second topic of this dissertation, i.e., renewable energy integration in smart grids. In this chapter, we will consider parallel load schedule optimization with distributed on-site renewable generation.⁷

4.1 Introduction

In an electricity market, demand response [64, 65] is defined as a mechanism used to manage customer consumption of electricity in response to the supply conditions. It is usually achieved through optimal time-dependent pricing [66, 67, 68, 69, 70]. The principal goal of demand response is to reduce the generation cost of electricity by reducing the peak load and by shifting peak-hour demand to off-peak hours. In a traditional electricity market, demand response is often accomplished by time-of-use (TOU) pricing. A TOU pricing strategy divides a day into several time intervals and assigns different prices for electricity in each interval. To reduce their electricity bills, users will adjust their consumption of electricity based on the different prices. In [67, 68], and our prior work [69], multiple methods are developed to determine optimal TOU prices for a centralized generator system.

⁷This chapter is based on P. Yang, P. Chavali, E. Gilboa, and A. Nehorai, “Parallel load schedule optimization with renewable distributed generators in smart grids,” *IEEE Trans. on Smart Grid*, Vol 4, No. 3, pp. 1431-1441, Sep. 2013. © IEEE 2013.

In a smart grid, the advanced metering infrastructure (AMI) [71] and the energy-management controller (EMC) enable communication between the users and the utility company, and make it possible to implement more effective demand response strategies. The AMI device collects data on the electricity usage and communicates with other AMI devices and the system controller. The EMC device helps the users manage and schedule their consumption of electricity in order to minimize their cost and inconvenience/dissatisfaction.

Renewable distributed generators (DGs) [72] are becoming an increasingly important component of the smart grid. DGs are attractive options to mitigate problems caused by traditional centralized generators. For example, an outage due to the breakdown of a centralized generator can result in various critical services being shut down. When distributed generators are employed, in the event of failure of a few generators, such critical services can be powered using other generators. DGs are also environmentally friendly, since they use renewable resources such as solar or wind energy. Customers with DGs can sell extra generation to the grid, thus forming a new electricity market paradigm.

In this chapter we propose a unified framework for demand response incorporating all three elements: the AMI, EMC, and DGs. We consider a smart grid in which part of the users are equipped with on-site renewable generators, for example, roof-top wind turbines or solar panels that are installed on user properties and that partially satisfy these users electricity demand. When the output of DGs is less than the users' demand, these users will purchase electricity from the utility company (we will use "utility" for short); when the generation of the DGs exceeds the users' demand, these users sell the excess electricity to the utility. The EMC will help users schedule their use of electricity autonomously. The user schedules in each iteration are reported only to the utility through the AMI. The utility aggregates the information from all users, and provides the users with feedback on the prices, which are determined by the aggregated scheduled load of all users. Instead of the sequential *Gauss-Siedel* algorithm [73] commonly used in the literature, we consider a parallel distributed optimization approach, which will significantly reduce the time complexity and communication costs. In addition, individual users do not need to disclose their load schedule to other users in the grid, which avoids potential privacy issues. Using the AMI, the users will be billed for their hourly consumption of electricity at rates that depend on the generation costs. This will incentivize users to shift their peak-hour loads. As we illustrate using numerical examples,

the aggregated load curve of the users is balanced, and the costs to both the utility and the users decrease after the optimization.

The rest of this chapter is organized as follows. In Section 4.2 we briefly review related work in the literature. In Section 4.3 we introduce the model for the users, the pricing strategy, and the formulation of the optimization problem. In Section 4.4 we propose a parallel distributed optimization framework. We show numerical simulations in Section 4.5, and summarize this chapter in Section 4.6.

4.2 Related work

There are many recent works on load scheduling and demand response in smart grids [74, 75, 76, 77, 78, 79]. In [74], the authors propose a game-theoretic approach for scheduling home appliances. With this method, the users always wish to schedule their use of electricity for hours when the price is low. This results in a shift of peak load, but the peak load itself is not significantly reduced. To mitigate this problem, the authors propose to use real-time pricing (RTP). In the case of RTP, a good price forecasting algorithm has to be developed [75, 78], which is often difficult. A more attractive approach is to enable communications among the users and the utility. In [79], the authors propose a cooperative game among users and use distributed optimization to coordinate the user demands. This method is effective in leveling user demand and reducing the cost of generation. However, the distributed optimization has to be executed sequentially, meaning that two users cannot optimize their schedules at the same time, which results in significant delay and control cost when the number of users is large. In addition, in this framework, users are billed based on their total daily use of electricity, and there is no price difference between on-peak load and off-peak load. Each user has to broadcast his or her schedule in each iteration to all the users in the network, incurring high communication costs and raising privacy issues. In our proposed approach, the optimization of individual users can be executed in parallel, and users are billed based on their hourly consumption of electricity. The load schedules are transmitted only to the utility, therefore providing user privacy and reducing communication and control costs.

Distributed generation is another important component of our work. In [80] the authors consider a game-theoretic approach for controlling renewable distributed generation. The

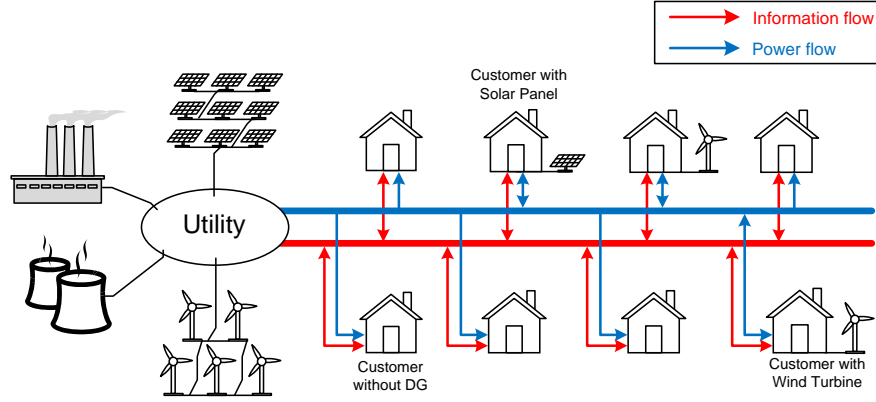


Figure 4.1: Two-way power and information flow between the utility and users.

equilibrium distributed power generation at each bus is studied, with consideration of physical regulations of the DG power injection. In this work, we consider a different scenario, where the DGs are low-capacity on-site generators primarily for the needs of the users who install them. Power injection to the grid occurs only when there is extra generation, and the amount is small.

4.3 System model

We consider a power grid with one utility, N regular users without DGs, and M users with DGs (as illustrated in Fig. 4.1). For each user, a smart meter measures the hourly consumption of electricity and communicates with the utility. An EMC helps each user optimize their load schedule and controls the operation of appliances. For users with DGs, the smart meters also measure the amount of electricity they sell to the utility. Two-way communication between the users and the utility is enabled by the AMI.

4.3.1 Model of users

We divide a day into T time periods. The loads of regular (non-DG) users and users with DGs are denoted as $\{l_{R,n}(t)\}_{t=1}^T$ and $\{l_{D,m}(t)\}_{t=1}^T$, respectively, where $n = 1, 2, \dots, N$ and $m = 1, 2, \dots, M$. In the rest of this paper we also use the notations $\mathbf{l}_{R,n} = [l_{R,n}(1), \dots, l_{R,n}(T)]^\top$

and $\mathbf{l}_{D,m} = [l_{D,m}(1), \dots, l_{D,m}(T)]^\top$ for notational simplicity. The user load is divided into three parts: base load, flexible load, and schedulable load, denoted by the superscripts B, F, and S, with

$$l_{Q,k}(t) = l_{Q,k}^B(t) + l_{Q,k}^F(t) + l_{Q,k}^S(t), \forall Q \in \{R, D\}, t, k. \quad (4.1)$$

A brief description about each type of load follows.

Base load

Base load is the electricity needed to satisfy basic user needs, e.g., refrigerators and lighting. This load is relatively fixed, and cannot be adjusted or rescheduled on a daily basis⁸. Therefore, we express it as constants in the optimization process.

Flexible load

Flexible load can be adjusted by users, but the adjustment will result in a satisfaction cost or gain. For example, when the electricity price is high during the daytime in the summer, users may choose to raise the temperature setting of their air conditioners to reduce energy cost. We use a function $s_{Q,k}^t(l_{Q,k}^F(t), d_{Q,k}^F(t))$ to quantify the satisfaction cost of users due to the difference between the nominal flexible demand $d_{Q,k}^F(t)$ and the actual flexible load $l_{Q,k}^F(t)$.

If the actual flexible load is smaller than the nominal demand, the function value is positive, meaning the users are not satisfied, which results in a loss of satisfaction that comes at a cost to the end-user. The value of the function increases faster as the actual load decreases. If the actual load is greater than the user demand, the function value is negative, meaning the users are satisfied. The decrease of the function value, however, slows down as the actual load continues to increase, because the satisfaction will saturate as the users use more electricity. When the actual load equals the user demand, the function value is zero. The satisfaction function $s_{Q,k}^t$ should meet the following conditions [69]:

⁸Indeed, lighting costs can be significantly reduced if, for example, incandescent light bulbs are replaced by fluorescent or LED light bulbs. Such changes from replacing appliances or devices are mostly one-time changes, which cannot be adjusted or scheduled on a daily bases.

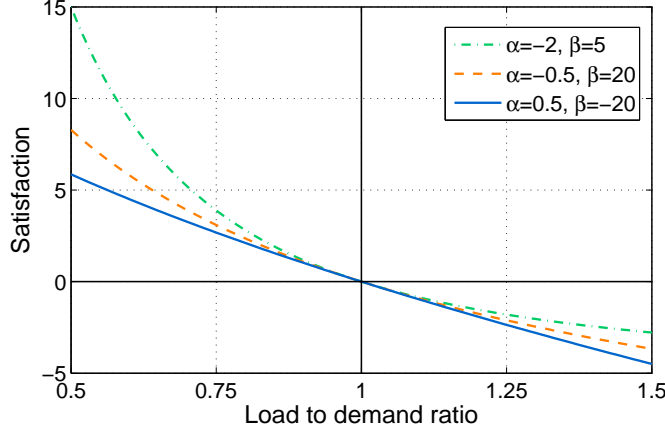


Figure 4.2: Example of a satisfaction cost function with different parameters.

If $l_{Q,k}^F(t) = d_{Q,k}^F(t)$, **then** $s_{Q,k}^t = 0$,

If $l_{Q,k}^F(t) > d_{Q,k}^F(t)$, **then** $s_{Q,k}^t < 0$, $\frac{\partial s_{Q,k}^t}{\partial l_{Q,k}^F(t)} < 0$, $\frac{\partial^2 s_{Q,k}^t}{\partial l_{Q,k}^F(t)^2} > 0$,

If $l_{Q,k}^F(t) < d_{Q,k}^F(t)$, **then** $s_{Q,k}^t > 0$, $\frac{\partial s_{Q,k}^t}{\partial l_{Q,k}^F(t)} < 0$, $\frac{\partial^2 s_{Q,k}^t}{\partial l_{Q,k}^F(t)^2} > 0$.

Therefore, the function $s_{Q,k}^t$ is a convex decreasing function of $l_{Q,k}^F(t)$. The parameters of the function can be different to reflect different users' preferences, e.g., their sensitivity in the difference between nominal demand and actual load.

In [69], we proposed a satisfaction function that satisfies all the aforementioned conditions:

$$s_{Q,k}^t(l_{Q,k}^F(t), d_{Q,k}^F(t)) = d_{Q,k}^F(t) \beta_{Q,k}(t) \left(\left(\frac{l_{Q,k}^F(t)}{d_{Q,k}^F(t)} \right)^{\alpha_{Q,k}(t)} - 1 \right), \quad (4.2)$$

where $\alpha_{Q,k}(t)$ and $\beta_{Q,k}(t)$ are parameters with $\alpha_{Q,k}(t) < 1$ and $\alpha_{Q,k}(t)\beta_{Q,k}(t) < 0$. Different parameters of $\alpha_{Q,k}(t)$ and $\beta_{Q,k}(t)$ reflect the different preferences of the users. An example of 4.2 with different parameters is shown in Fig. 4.2.

Schedulable load

Schedulable load is the load that users can schedule under certain constraints. The schedulable load for an individual user consists of the loads from multiple schedulable appliances. Let $\mathcal{A}_{Q,k}$ denote the set of schedulable appliances of the k th type Q user, we have

$$l_{Q,k}^S(t) = \sum_{a \in \mathcal{A}_{Q,k}} l_{Q,k}^a(t), \forall Q \in \{R, D\}, k, t. \quad (4.3)$$

For each appliance, there is a starting time and completion deadline for the task, denoted by t_a^s and t_a^e , respectively, where the subscript a denotes the index of the appliance. To simplify the notation, we also use $\mathcal{T}_{Q,k}^a = \{t | t_s^a \leq t \leq t_e^a\}$ to denote the feasible operation time periods for each appliance. For each appliance, there are also lower and upper bounds on the power consumption during its operation period. In addition, to complete a specific task, there is a constraint on the total energy consumed by the appliance over the operation time periods. The corresponding constraints are expressed as follows:

$$\sum_{t \in \mathcal{T}_{Q,k,a}} l_{Q,k}^a(t) = E_{Q,k}^a, \forall a \in \mathcal{A}_{Q,k}, Q \in \{R, D\}, k, \quad (4.4)$$

$$\begin{cases} l_{Q,k}^{a,\min} \leq l_{Q,k}^a(t) \leq l_{Q,k}^{a,\max}, & \text{when } t \in \mathcal{T}_{Q,k}^a, \\ l_{Q,k}^a(t) = 0, & \text{otherwise,} \end{cases} \quad (4.5)$$

where $E_{Q,k}^a$ denotes the total energy consumption requirement of appliance a , and $l_{Q,k}^{a,\min}$ and $l_{Q,k}^{a,\max}$ denote the lower and upper bounds on the hourly power consumption of the appliance during its operational hours. Similar models for schedulable appliances have been used in the literature [81, 82]. Note that although many appliances can only be turned on and off, resulting in discrete changes in their power consumption at a specific time point, the hourly consumption of a specific appliance can be continuously changed, depending on the length of time it operates during that hour. We also model electric vehicle charging using the same model as for appliances.

The total load of a single user is bounded by the base load and an upper bound $l_{Q,k}^{\max}$, which is determined either by physical line capacity or by the utility:

$$l_{Q,k}^B(t) \leq l_{Q,k}(t) \leq l_{Q,k}^{\max}, \quad \forall t, k. \quad (4.6)$$

Distributed generation

Let $g_{D,m}(t)$ denote the actual generation of the m th user with DG during time period t . When $g_{D,m}(t) < l_{D,m}(t)$, the user purchases electricity from the utility company at the same retail price $p_s(t)$ paid by regular users. When $g_{D,m}(t) > l_{D,m}(t)$, the extra electricity is sold back to the utility at a different buyback price $p_D(t)$. Since the scheduling is performed day-ahead, the prediction of renewable generation is not perfect. Therefore we assume there is error in the prediction of renewable generation,

$$\mathbf{g}_{D,m}(t) = \tilde{\mathbf{g}}_{D,m}(t) + \mathbf{e}_{D,m}^g, \quad (4.7)$$

where $\tilde{\mathbf{g}}_{D,m}(t)$ denotes the predicted generation, and $\mathbf{e}_{D,m}^g$ denotes the prediction error. In this chapter, we assume the prediction error follows a Gaussian distribution $\mathcal{N}(\mathbf{0}, \sigma_e^2 \mathbf{I})$.

Energy storage

We do not specifically model energy storage in this work, because the cost of energy storage devices is usually high. Considering the relatively low on-site energy generation capacity, and the fact that the users are connected to the grid with two-way energy flow, it is reasonable to avoid energy storage devices for each individual energy user⁹. However, we would like to point out that the use of energy storage will potentially be helpful in smoothing out each individual user's load consumption curve and mitigating the prediction errors of on-site renewable generation¹⁰.

4.3.2 Utility costs and electricity price

Cost to the utility is a non-decreasing convex function of the total generation. Usually a piece-wise linear function or a quadratic function is used as the generation cost function.

⁹Nevertheless, we will specifically model energy storage systems in the next chapter, when dealing with renewable generation in a different scenario.

¹⁰In recent work, electric vehicles are sometimes considered as home energy storage devices, and the vehicle-to-grid (V2G) effects are studied [83, 84].

Let $g_u(t)$ denote the generation of the utility at time period t . We model the generation cost using a non-decreasing convex function $C_t(g_u(t))$ at time t .

We assume that the load balance constraint is satisfied, i.e.,

$$g_u(t) = \sum_{n=1}^N l_{R,n}(t) + \sum_{m=1}^M (l_{D,m}(t) - g_{D,m}(t)), \quad \forall t, \quad (4.8)$$

and the generation is bounded by the system limit,

$$g_u(t) \leq g_u^{\max}, \quad \forall t. \quad (4.9)$$

Rate-of-return regulations [85] allow the utility to earn a fair rate-of-return on its investment. Thus in this chapter, we set the time-dependent unit retail price of electricity to be proportional to the time-dependent generation cost,

$$p_S(t) = \mu(t) \frac{C_t(g_u(t))}{g_u(t)}, \quad (4.10)$$

where $\mu(t) > 1$ is a preset profit coefficient. Other pricing models can also be chosen according to different pricing policies. For example, when a fixed profit is allowed for each unit of electricity, the retail price (4.10) can be modified as

$$p_S(t) = \frac{C_t(g_u(t))}{g_u(t)} + p_F, \quad (4.11)$$

where p_F denotes the fixed unit profit.

In addition to the retail price, the utility needs to decide the buyback price $p_D(t)$ of electricity from users. In this chapter, we set the buyback price to be equal to the marginal generation cost, i.e.,

$$p_D(t) = \frac{C_t(g_u(t))}{g_u(t)}. \quad (4.12)$$

When the retail price is high, the buyback price is also high, which guarantees user benefits. Also, the buyback price would incentivize the utility to buy from users; otherwise, if the utility needed to generate on its own, due to the convexity of the generation cost function, the marginal cost would increase. Therefore this buyback price is fair for both users and

the utility company. Assuming the utility buys back all the excess generation of the users, the total cost $C_{u,g}$ for electricity to the utility consists of the generation cost and the cost of purchasing electricity from users with DGs,

$$C_{u,g} = \sum_{t=1}^T \left(C_t(g_u(t)) + p_D(t) \sum_{m=1}^M (g_{D,m}(t) - l_{D,m}(t))^+ \right),$$

where

$$(g - l)^+ = \begin{cases} g - l & \text{if } g > l, \\ 0 & \text{otherwise.} \end{cases}$$

Due to the nature of electricity markets, the utility has to be regulated, and they are required to satisfy the electricity users. Therefore the total cost to the utility, considering social welfare, also includes the user satisfaction cost, i.e.,

$$C_u = C_{u,g} + \sum_{t=1}^T \left(\sum_{n=1}^N s_{R,n}^t + \sum_{m=1}^M s_{D,m}^t \right). \quad (4.13)$$

The goal of the utility would be to minimize the total cost (4.13) by optimizing user load schedules. In practice, the user preference and actual user load are hard for the utility to quantify or predict, since they are decided by individual users when they make the tradeoff between the satisfaction associated with the usage of electricity and the cost they pay. As a result, the minimization of C_u with respect to each individual user's load is intractable in a centralized manner. A more reasonable approach is to allow individual users to optimize their own load schedules, to achieve a minimized overall cost. We need to point out that the objective function of the utility (4.13) is not exactly the same as the summation of the individual objective functions of the users, unless μ is chosen to be μ_{th} where

$$\mu_{th}(t) = 1 + \frac{\sum_{m=1}^M (g_{D,m}(t) - l_{D,m}(t))^+}{g_u(t) + \sum_{m=1}^M (g_{D,m}(t) - l_{D,m}(t))^+}. \quad (4.14)$$

However, by minimizing individual users' costs, the overall load curve is flattened, which in turn reduces the cost to the utility.

4.3.3 Optimization of load schedule

The idea of distributed optimization is to allow users to autonomously minimize their own costs by properly scheduling their loads, which in turn flattens the aggregated load curve and reduces the cost to the utility. Since the retail price p_S is a function of g_u , which depends on the total user load, the users wish to cooperate in order to minimize their own costs.

The total cost to a user includes the user satisfaction cost in addition to the electricity bill. Realistically, people tend to balance the money they pay with the benefit they obtain. For a user without DG, the daily cost function includes the utility bill $P_{R,n}$ and the satisfaction cost $S_{R,n}$,

$$\begin{aligned} C_{R,n}(\mathbf{l}_{R,n}) &= P_{R,n}(\mathbf{l}_{R,n}) + S_{R,n}(\mathbf{l}_{R,n}) \\ &= \sum_{t=1}^T p_S(t) l_{R,n}(t) + \sum_{t=1}^T s_{R,n}^t(l_{R,n}^F(t), d_{R,n}^F(t)). \end{aligned} \quad (4.15)$$

For a user with DG, the daily cost will also include the profit they make by selling electricity to the utility. Since the prediction of renewable generation is not accurate, with errors defined in (4.7), we will use the expected cost for this type of users. Therefore the cost function is

$$\begin{aligned} C_{D,m}(\mathbf{l}_{D,m}) &= \mathbb{E}(P_{D,m}(\mathbf{l}_{D,m})) + S_{D,m}(\mathbf{l}_{D,m}) \\ &= \mathbb{E} \left(\sum_{t=1}^T p_S(t) (l_{D,m}(t) - g_{D,m}(t))^+ - \sum_{t=1}^T p_D(t) (g_{D,m}(t) - l_{D,m}(t))^+ \right) \\ &\quad + \sum_{t=1}^T s_{D,m}^t(l_{D,m}^F(t), d_{D,m}^F(t)). \end{aligned} \quad (4.16)$$

Note that since the users make money by selling the electricity they generate, the corresponding term in the cost function carries a negative sign. Under the Gaussian assumption on the prediction error, we can then calculate the expectation of the bill $P_{D,m}$. Denote the net load as

$$\mathbf{h}_{D,m} = \mathbf{l}_{D,m} - \tilde{\mathbf{g}}_{D,m}, \quad (4.17)$$

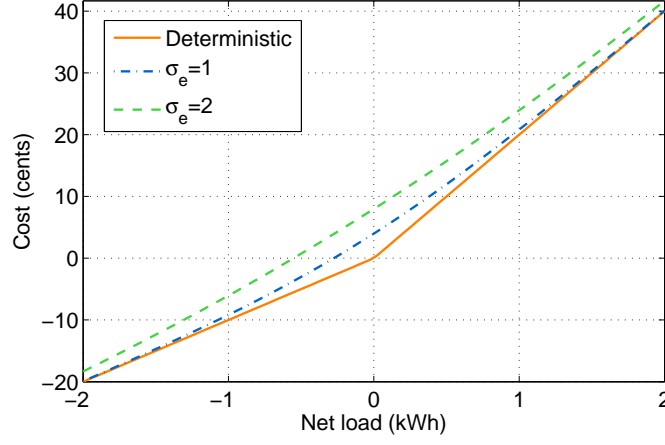


Figure 4.3: Example of deterministic and expected cost as a function of the net load. $p_S(t)$ and $p_D(t)$ are set to be 20 cents/kWh and 10 cents/kWh, respectively.

and the expected bill for the m th user with DG is then calculated as follows:

$$\begin{aligned}
\mathbb{E}(P_{D,m}(\mathbf{l}_{D,m})) &= \sum_{t=1}^T \left(\int_{-\infty}^{h_{D,m}(t)} p_S(t) (h_{D,m}(t) - e_{D,m}^g(t)) f_e(e_{D,m}^g(t)) de_{D,m}^g(t) \right. \\
&\quad \left. + \int_{h_{D,m}(t)}^{+\infty} p_D(t) (h_{D,m}(t) - e_{D,m}^g(t)) f_e(e_{D,m}^g(t)) de_{D,m}^g(t) \right) \\
&= \sum_{t=1}^T \left[h_{D,m}(t) \left(p_S(t) \Phi \left(\frac{h_{D,m}(t)}{\sigma_e} \right) + p_D(t) \left(1 - \Phi \left(\frac{h_{D,m}(t)}{\sigma_e} \right) \right) \right) \right. \\
&\quad \left. + (p_S(t) - p_D(t)) \frac{\sigma_e}{\sqrt{2\pi}} \exp \left(-\frac{h_{D,m}(t)^2}{2\sigma_e^2} \right) \right], \tag{4.18}
\end{aligned}$$

where $f_e(\cdot)$ denotes the probability density function of the prediction error, and $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution. Fig. 4.3 shows an illustrative example of the deterministic cost function, and two expected cost functions with different prediction errors σ_e . The deterministic cost can be considered as the expected cost when $\sigma_e = 0$ kWh.

The objective of the k th user of type $Q \in \{R, D\}$ is to optimize $\mathbf{l}_{Q,k}$ so that the cost $C_{Q,k}$ is minimized. In (4.1) we decomposed the load into three components. The optimization is over the flexible load and the schedulable load, where the schedulable load can be further decomposed into the loads of different appliances. In this parallel optimization framework,

each user optimizes his own load, and the optimization for each user can be written as

$$\begin{aligned}
& \min_{l_{Q,k}^F, l_{Q,k}^S} C_{Q,k}(l_{Q,k}) \\
& \text{subject to} \quad l_{Q,k}(t) = l_{Q,k}^B(t) + l_{Q,k}^F(t) + l_{Q,k}^S(t), \forall Q \in \{R, D\}, t, k, \\
& \quad l_{Q,k}^S(t) = \sum_{a \in \mathcal{A}_{Q,k}} l_{Q,k}^a(t), \forall Q \in \{R, D\}, k, t, \\
& \quad \sum_{t \in \mathcal{T}_{Q,k,a}} l_{Q,k}^a(t) = E_{Q,k}^a, \forall a \in \mathcal{A}_{Q,k}, Q \in \{R, D\}, k, \\
& \quad \begin{cases} l_{Q,k}^{a,\min} \leq l_{Q,k}^a(t) \leq l_{Q,k}^{a,\max}, & \text{when } t \in \mathcal{T}_{Q,k}^a, \\ l_{Q,k}^a(t) = 0, & \text{otherwise,} \end{cases} \\
& \quad l_{Q,k}^B(t) \leq l_{Q,k}(t) \leq l_{Q,k}^{\max}, \quad \forall t, k.
\end{aligned} \tag{4.19}$$

The local optimization can be efficiently solved using standard convex optimization methods, e.g., the interior point method [60]. However, convergence of this algorithm requires coordination among the users. In the next section, we will discuss how to coordinate the parallel distributed optimizations.

4.4 Parallel distributed optimization

In [79], the authors propose to solve a distributed optimization for energy consumption scheduling following the *Gauss-Seidel* mapping framework. Each local optimization of a single user is carried out successively, and after each optimization, the user broadcasts the new schedule to all other users in the network. All other users update this information, and then the next user starts optimization. This algorithm is guaranteed to converge [73], but limitations exist. First, although the optimization is distributed, it has to be carried out sequentially. When the number of users is large, the computation time is not trivial, and the control cost is high. Second, each user has to broadcast their schedule to all the other users, which results in privacy issues and significant communication costs. The broadcast also requires additional communication infrastructure among the users.

We propose to solve (4.19) in a parallel distributed manner. After receiving updates from the utility, the users are able to execute their local optimizations simultaneously. However,

one concern with this approach is the convergence of the parallel distributed algorithm. If the behavior of the users is not well coordinated, the distributed optimization will very likely result merely in shifted peaks. All the users would wish to schedule their load for time periods when the price is low, and thus after the distributed optimization is finished, the aggregated load schedule would exhibit shifted peak loads, resulting in high costs for both users and the utility. In [86], the authors propose to use a penalty on the distance between two iteration steps to control the parallel optimization in the tensor field estimation. Motivated by this idea, we introduce a penalty on changes of load schedule between two consecutive iterations for each user.¹¹ The objective function for the optimization of user k , type Q, at the i th iteration is then revised as

$$\min_{\mathbf{l}_{Q,k}^{F,(i)}, \mathbf{l}_{Q,k}^{S,(i)}} C_{Q,k}(\mathbf{l}_{Q,k}^{(i)}) + \lambda_{Q,k} \left\| \mathbf{l}_{Q,k}^{(i)} - \mathbf{l}_{Q,k}^{(i-1)} \right\|_2^2, \quad (4.20)$$

where the superscript i denotes the number of the iteration. Tables 4.1 and 4.2 show the algorithms for the parallel distributed optimization. The only information users require from the utility is the price as a function of their load schedules. In the case of residential load scheduling scenario, the load of each individual user is small compared with the aggregated load. Then users need only the prices \mathbf{p}_S and \mathbf{p}_D , since the impact of an individual user's load schedule on the price is trivial. Individual users need to report their load schedules only to the utility, not to all the other users, which protects user privacy. The utility will aggregate the information and update the users for the next iteration. The distributed optimization stops when the changes in user load schedule and utility cost in consecutive iterations are within (low) preset thresholds.

The additional term $\lambda_{Q,k} \left\| \mathbf{l}_{Q,k}^{(i)} - \mathbf{l}_{Q,k}^{(i-1)} \right\|_2^2$ penalizes a change of load schedule between two iterations (stepsize), so that each user is discouraged from making big changes in consecutive iterations. The selection of the parameter $\lambda_{Q,k}$, which we refer to as the penalty coefficient, is important as it determines the convergence and the performance of the distributed optimization. As an intuitive interpretation, a large $\lambda_{Q,k}$ strictly limits the step size, and therefore significantly slows down the convergence. A very small $\lambda_{Q,k}$, however, does not guarantee the convergence of the distributed optimization, and may cause oscillations due to shifting

¹¹A similar penalty term was recently used in [87] to coordinate electric vehicle charging.

Table 4.1: Algorithm for utility company.

Algorithm 4.1: algorithm for utility company.
begin
$ite = 0$; % Iteration counter
Obtain initial $\mathbf{l}_{R,n}^{ite}, \mathbf{l}_{D,m}^{ite}$ from all users;
Calculate $\mathbf{g}_u^{ite}, C_u^{ite}$, prices;
do
Broadcast updates to all users;
Obtain $\mathbf{l}_{R,n}^{ite}, \mathbf{l}_{D,m}^{ite}$ from all users;
Calculate $\mathbf{g}_u^{ite+1}, C_u^{ite+1}$, prices;
$ite := ite + 1$;
until $\ C_u^{ite} - C_u^{ite-1}\ < \delta_1, \ \mathbf{l}_{Q,k}^{ite} - \mathbf{l}_{Q,k}^{ite-1}\ < \delta_2, \forall Q, k$
end

Table 4.2: Algorithm for individual user.

Algorithm 4.2: algorithm for the k th user of type Q.
begin
$ite = 0$; % Iteration counter
Initialize $\mathbf{l}_{Q,k}^{ite} = \mathbf{d}_{Q,k}$;
Report $\mathbf{l}_{Q,k}^{ite}$ to the utility;
While Updates received from the utility
Solve local optimization and obtain $\mathbf{l}_{Q,k}^{ite+1}$;
Report $\mathbf{l}_{Q,k}^{ite+1}$ to the utility;
$ite := ite + 1$;
end
end

peaks in consecutive iterations. In the following subsections, we will discuss two strategies to select the penalty coefficient.

4.4.1 Fixed penalty coefficient

In the general case, since different users consume different amounts of electricity, the effects of their individual load schedules on the total load vary. We select the penalty coefficient,

$\lambda_{Q,k}$ using

$$\lambda_{Q,k} \propto \left(\sum_{t=1}^T g_u(t) / \sum_{t=1}^T l_{Q,k}(t) \right)^\varepsilon, \quad (4.21)$$

where $\varepsilon \geq 0$. For users who consume little electricity, the penalty coefficient is selected to be a large value to constrain the change of their load schedules in consecutive iteration steps. Otherwise these users tend to schedule their load without considering their aggregate impact on the retail price. Large consumers, however, have a more significant impact on the retail price, and they need to consider this in optimizing their own schedule. Therefore a small penalty coefficient is more suitable for these users.

Other aspects should also be considered when deciding constraints on the step size penalty. For example, the penalty coefficient should reflect that not all consumers are of equal priority. Let $\rho_{Q,k}$ denote the priority measurement, and then $\lambda_{Q,k}$ should be chosen as

$$\lambda_{Q,k} \propto \frac{1}{\rho_{Q,k}}, \quad (4.22)$$

so that users with higher priority will be constrained less on their load schedules.

When using a fixed penalty coefficient, it is important to select it properly because the convergence of the distributed optimization is heavily dependent on this choice. In general, a large penalty coefficient is preferred as it ensures convergence, but at the cost of slowing down the convergence.

4.4.2 Adaptive penalty coefficient

As an alternative to the fixed penalty coefficient, we can choose $\lambda_{Q,k}$ based on the previous stepsize and the iteration number:

$$\lambda_{Q,k}^{(i)} = \max \left\{ \tau_{Q,k} \left\| \mathbf{l}_{Q,k}^{(i)} - \mathbf{l}_{Q,k}^{(i-1)} \right\|, \kappa_{Q,k} i \right\}, \quad (4.23)$$

where the superscript i denotes the i th iteration, and $\tau_{Q,k}$ and $\kappa_{Q,k}$ are constants which take into account the factors considered in (4.21) and (4.22), i.e.,

$$\tau_{Q,k}, \kappa_{Q,k} \propto \frac{1}{\rho_{Q,k}} \left(\sum_{t=1}^T g_u(t) / \sum_{t=1}^T l_{Q,k}(t) \right)^\varepsilon. \quad (4.24)$$

The adaptive penalty coefficient consists of two terms. When i is small, the first term in (4.23) dominates the second term. In this case, if a user makes a big change in the i th iteration, the stepsize penalty will be higher in the next iteration, and thus the user will make a smaller change in the next iteration. On the contrary, if a user makes a small change in the i th iteration, more flexibility will be given in the next iteration. When the number of iterations goes up, the second term in (4.23) will dominate the first term. Since the second term increases with the number of iterations, the stepsize will be increasingly constrained to ensure convergence.

4.4.3 Convergence properties

As we discussed, a larger stepsize penalty coefficient is in general beneficial for the convergence, but potentially decrease the speed of convergence. In Appendix C we sketch a preliminary proof for convergence when the stepsize penalty coefficient is large. We also use numerical examples in Section 4.5.3 and Section 4.5.4 to show the convergence properties using different fixed and adaptive stepsize penalty coefficients.

4.5 Numerical examples

4.5.1 General setup

We consider a power system with one utility company, 80 regular users without DGs, and 20 users with DGs. A day is divided into $T = 24$ time periods. We randomly generate the daily user demand data and DG generation data based on the MISO daily report by the U.S. Federal Energy Regulatory Commission (FERC) [88], and the simulated wind generation data

provided by the Ontario Power Authority [89]. We set the standard deviation of prediction error of DG generation to be $\sigma_e = 0.5$ kWh for each time period, unless otherwise specified.

For the schedulable load, we assume each user has a random number of schedulable appliances from a predefined set of appliances, e.g., dishwashers, washing machines, and dryers. We also assume part of the users own electric vehicles with different battery capacities. The starting time and completion deadline of the tasks are randomly generated.

We employ the satisfaction function as described in (4.2). In this distributed optimization scenario, the utility company does not need information about these parameters.

For the generation cost of the utility company, we use the quadratic function

$$C_t(g_u(t)) = c_1^t g_u(t) + c_2^t (g_u(t))^2, \quad (4.25)$$

with $c_1^t = 0.1$ cents/kWh and $c_2^t = 0.025$ cents/kWh². The profit coefficient μ is selected to be 1.2. We assume all users are of equal priority. For the fixed λ case, we select the base $\lambda_{Q,k}$ to be

$$\lambda_{Q,k} = \left(\sum_{t=1}^T g_u(t) / \sum_{t=1}^T l_{Q,k}(t) \right)^{\frac{1}{2}}. \quad (4.26)$$

For the adaptive λ case, we select the parameter $\tau_{Q,k}$ of term I to be the same as $\lambda_{Q,k}$ in the fixed λ case, and select the parameter $\kappa_{Q,k}$ of term II to be

$$\kappa_{Q,k} = \frac{1}{160} \lambda_{Q,k}. \quad (4.27)$$

The initial stepsize is set to be 1, and the stopping criteria for the iterative algorithm are set to be $\delta_1 = 10^{-2}$ and $\delta_2 = 10^{-3}\sqrt{T}$.

In Section 4.5.2 we will compare the load and costs before and after optimization. In Sections 4.5.3 and 4.5.4 we will show examples of convergence using fixed λ and adaptive λ , respectively. Examples illustrating the effects of DG generation will be included in Section 4.5.5.

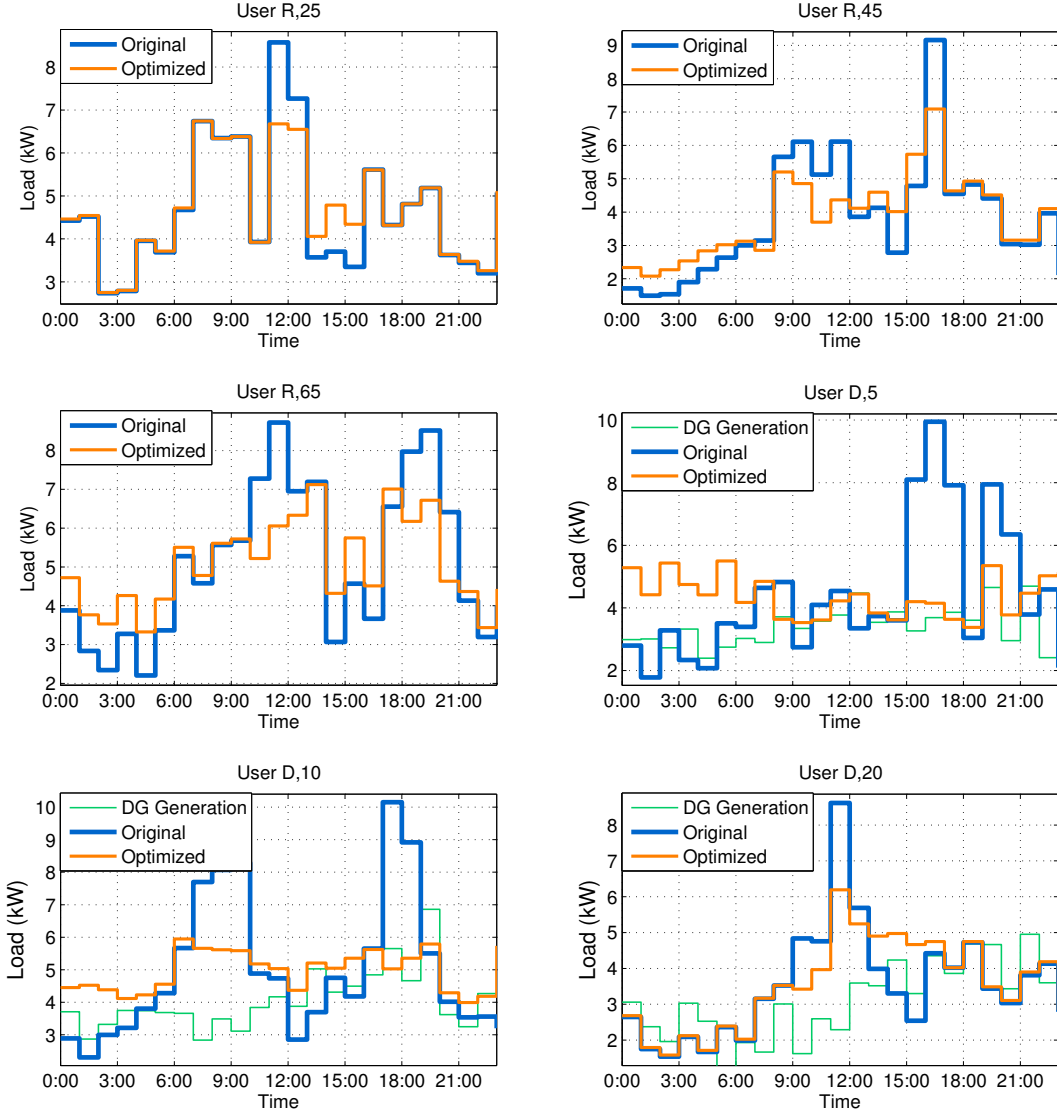


Figure 4.4: Single user load schedule before and after optimization. Regular users and users with renewable generators are distinguished by R and D, respectively.

4.5.2 Simulation results

In this section we show results when the fixed λ is selected according to (4.26). Using the described setup, we ran the algorithms in Tables 4.1 and 4.2. In Fig. 4.4 we show the load profile of three regular users and three users with renewable generators before and after optimization. All the users reduce their peak loads by shifting part of their on-peak

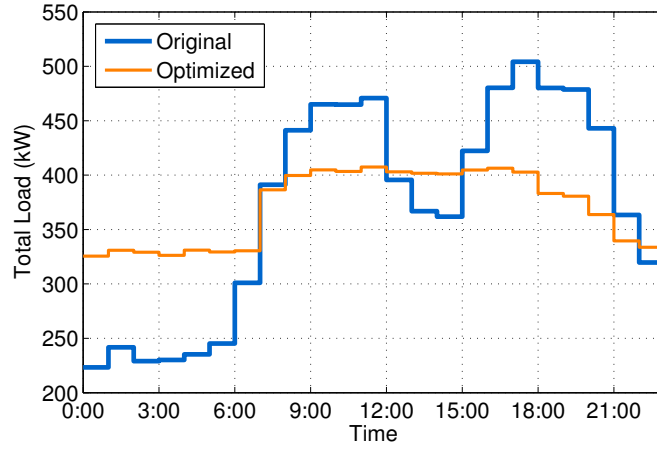


Figure 4.5: Aggregated hourly load to the utility company before and after optimization. Peak-hour load is significantly balanced, and load curve is balanced after optimization.

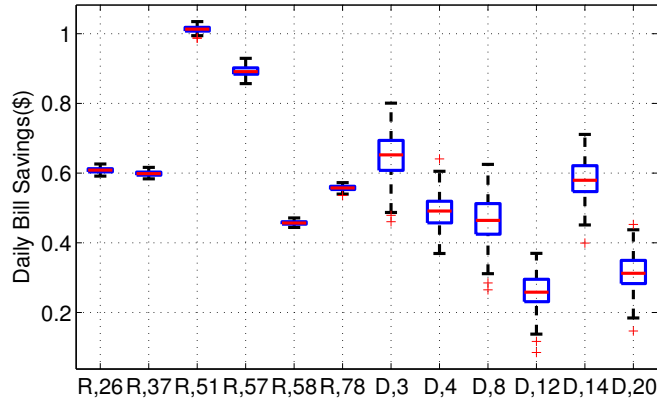


Figure 4.6: Reduction in daily bill for randomly selected users before and after optimization. Values calculated based on 200 realizations of randomly drawn DG generation data.

load to off-peak hours. The flexible loads do not deviate from their flexible demand much, because the users make a trade-off between their cost in money and their satisfaction cost. In this simulation we set the parameters of the satisfaction cost function so that the flexible user demands were relatively inelastic. In practice, users can decide their own preferences, and their flexible load can significantly deviate from their flexible demand if they are more concerned with their electricity bills than their convenience or comfort. We notice that the load curve of each individual user is not necessarily balanced. The generation curve of the

utility, however, is well balanced, as shown in Fig. 4.5. The peak load reduces from 504.2 kW to 407.5 kW, and the peak-to-average (PAR) ratio reduces from 1.37 to 1.10. This benefits the utility because it reduces the use of expensive secondary generators to satisfy peak-hour user load, and the reduced peak load is beneficial for the stability and efficiency of the power grid. The total daily cost to the utility decreases from \$891.2 to \$841.9. This saving is based on the specific generation cost function (4.25) and the specific parameters we selected for this simulation. In practice, the savings could be more because the fast-response secondary generators are often extremely inefficient and costly.

In Fig. 4.6 we show box plots of daily savings in electricity bills for six regular users and six users with DGs, all randomly selected. The savings are calculated based on 200 realizations of actual DG generations, which are randomly drawn based on (4.7). The savings for users with DGs have a higher uncertainty due to the prediction error in DG generation. Note that the comparisons are made between the bills before and after optimization. The absolute savings of users with DGs is much higher than regular users without DGs, as their energy needs are partially provided by DGs.

4.5.3 Convergence with a fixed penalty coefficient

In this section, we study the effect of the fixed penalty coefficients on the results. In Section 4.4 we briefly discussed the convergence of the distributed optimization for different choices of the penalty coefficient. Here, we show a series of numerical simulations with different choices of this parameter in Fig. 4.7. We select the base λ according to (4.26), and perform a series of simulations with the parameter λ_{sim} selected as $\lambda, \frac{1}{2}\lambda, \frac{1}{4}\lambda, \frac{1}{8}\lambda, \frac{1}{16}\lambda$, and $\frac{1}{32}\lambda$. For comparison, we also perform simulations with sequential optimization, similar to the algorithm in [79]. Since the centralized optimization problem is intractable, we use the final optimization result from the sequential optimization as the benchmark, which is indicated in Fig. 4.7 and Fig. 4.8 as “optimal”.

We observe that for this 100-user grid, the convergence of our algorithm can be achieved in as few as 14 iterations, and the number of iterations depends on the selected λ . The distributed optimization is done in parallel, and thus the runtime for each iteration is approximately $\frac{1}{100}$ that of sequential optimization. For properly chosen λ_{sim} , the parallel optimization converges

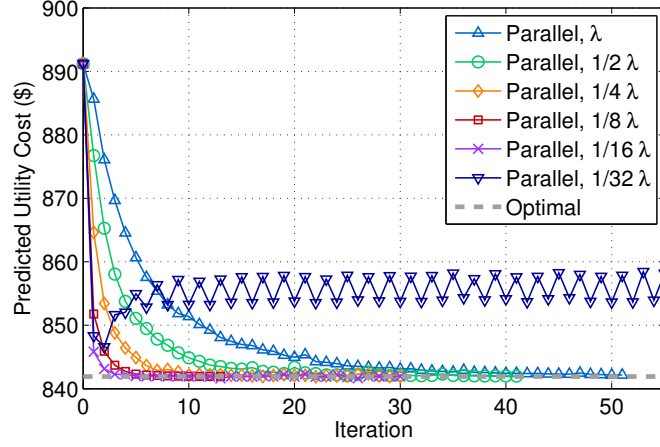


Figure 4.7: Convergence of parallel optimization with fixed penalty coefficients.

Table 4.3: Comparison of computation time.

	Per Iteration (s)	# of Iteration	Total (s)
Parallel, λ	2.11	52	109.75
Parallel, $1/2\lambda$	2.00	42	83.91
Parallel, $1/4\lambda$	1.90	31	58.96
Parallel, $1/8\lambda$	1.71	14	23.99
Parallel, $1/16\lambda$	1.76	31	54.63
Sequential, 0	168.93	44	7442.98
Sequential, $1/8\lambda$	175.32	17	2980.41

to the same optimal point as the sequential optimization. The number of iterations also decreases as λ_{sim} decreases. At $\lambda_{\text{sim}} = \frac{1}{8}\lambda$, only 14 iterations are required for convergence. As λ_{sim} continues to decrease, the number of iterations increases. If λ_{sim} is too small, for example when $\lambda_{\text{sim}} = \frac{1}{32}\lambda$ in this case, the algorithm no longer converges, because there is little constraint on each user. The users always choose the optimal load schedule for their own benefit, with little consideration for the aggregate impact of their personal behavior on the retail price. As a result, the users shift their load between peak-hours and off-peak hours in consecutive iterations, resulting in oscillations.

In Table 4.3, we compare the computation time and number of iterations of our method with the sequential algorithm. For the sequential algorithm, we consider the original case without a stepsize penalty, and the case with a stepsize penalty $\lambda_{\text{sim}} = \frac{1}{8}\lambda$. Note that since

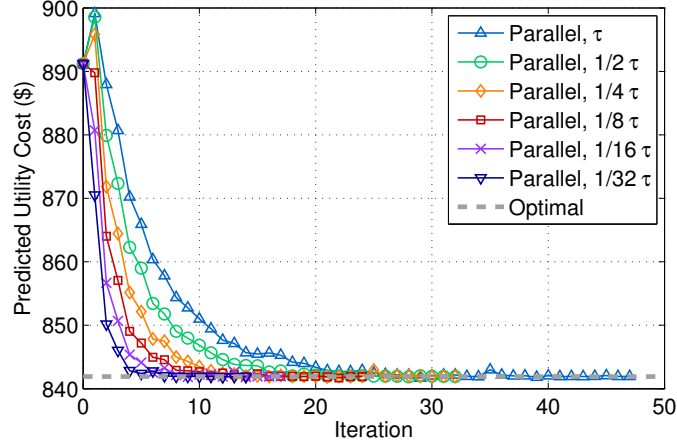


Figure 4.8: Convergence of parallel optimization with adaptive penalty coefficients.

the simulation platform is a quad-core computer, only four optimizations can be executed in parallel. The time per iteration for the parallel algorithm is calculated using the actual CPU time divided by 25. We observe that for each individual optimization, the runtime is approximately the same. However, due to the parallelization, the runtime of our algorithm is significantly reduced compared to the runtime of the sequential algorithm. We also observe that adding a stepsize penalty significantly decreases the number of iterations required by sequential algorithms.

4.5.4 Convergence with an adaptive penalty coefficient

In this subsection, we show numerical simulations employing an adaptive penalty coefficient, calculated using (4.23). We select different τ for the first term, similar to the way we selected different penalty coefficients in the fixed parameter scenario. Namely, we select the τ calculated in Section 4.5.1 as the base τ , and select $\tau_{\text{sim}} = \tau, \frac{1}{2}\tau, \frac{1}{4}\tau, \frac{1}{8}\tau, \frac{1}{16}\tau$, and $\frac{1}{32}\tau$. The parameter κ for the second term is calculated using (4.27), and is fixed in all the simulations. The results are plotted in Fig. 4.8.

We observe that using an adaptive penalty coefficient makes the convergence not sensitive to the selection of the penalty coefficient, because the coefficient changes with the stepsize and the iteration number. As the iteration number increases, the penalty on the stepsize

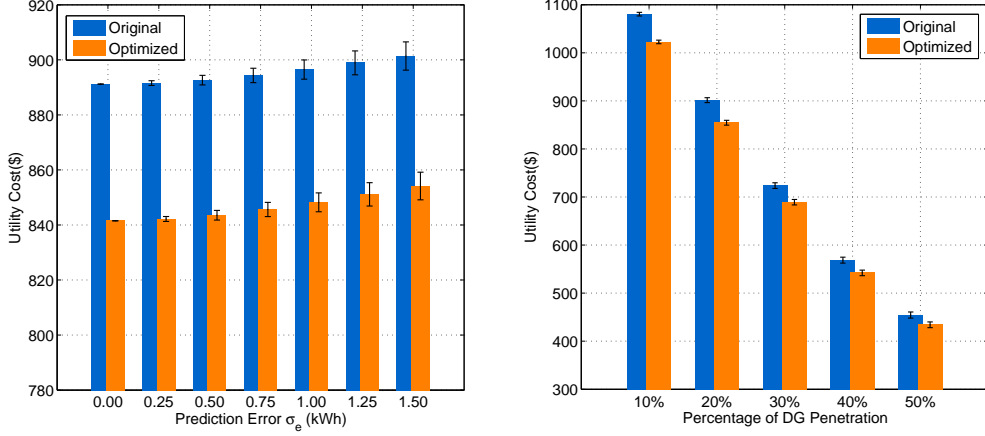


Figure 4.9: Effect of prediction error and DG penetration percentage on utility costs.

increases, and thus the algorithm converges for all the tested choices of parameters. A good choice of the parameter τ would, however, expedite the convergence.

4.5.5 Effect of distributed generators

In this subsection we show numerical examples of the effects of prediction error in DG generation, and the percentage of users with DGs. In Fig. 4.9a we observe that as the prediction error increases, both the original and optimized costs to the utility increase, and the uncertainty in the costs also increases. Therefore an accurate DG generation prediction is important when integrating such generators. In Fig. 4.9b we show how the utility cost changes as the percentage of DG penetration increases with a fixed σ_e . Since distributed generators partially satisfy user demands, the total generation and cost of the utility decreases. However, as more DGs are integrated, the uncertainties in the cost also increase. In this numerical example we assume the prediction error has a zero mean, and is uncorrelated among multiple users. In practice, if the prediction error suffers from an estimation bias, or is highly correlated, its impact on the total load and cost to the utility would be more significant.

4.6 Summary

In this chapter, we introduced a preliminary framework for parallel distributed optimization of demand response in a power system with renewable distributed generators. We divided the total user load into base load, flexible load, and schedulable loads. Some of the users are equipped with DGs, and can sell extra electricity to the utility company. In this framework, subscribed users optimize their own schedules of electricity consumption with limited information from the utility company. In order to coordinate the distributed optimization, we added a soft constraint on the user load schedule change (stepsize) between two consecutive iterations. We proposed and analyzed two methods to select the penalty coefficient – the fixed and the adaptive. Numerical results showed that our algorithm helps reduce costs to the utility and lowers electricity bills for users. The optimal scheduling and integration of DGs significantly reduces the peak-hour generation of the utility. The results of the parallel distributed optimization equal the optimal solution from sequential optimization within numerical precision, with reduced information exchange, lower communication costs, and less control complexity. Due to the parallelization of the algorithm, the computation time is significantly reduced compared to the commonly employed sequential algorithms. We also illustrated the effect of DGs on the algorithm from two perspectives - prediction errors in DG generation and the DG penetration level. We concluded that a more precise prediction of DG generation potentially reduces the costs to the utility. An increased ratio of users with DGs will reduce the total load and cost to the utility, but the prediction uncertainties for these generators need to be carefully investigated.

Chapter 5

Joint Storage and Renewable Generation Capacity Planning

In this chapter, we consider joint storage and renewable generation capacity planning, in the scenario of micro-grids with large renewable penetration.¹²

5.1 Introduction

Renewable energy sources, which currently provide only about 3% of the electricity in the United States, are expected by the National Renewable Energy Laboratory (NREL) to be able to support about 80% of the total electricity consumption in the U.S. in 2050 [8]. The high penetration of renewable energy is especially common in (remote) isolated grids, or micro-grids with small carbon footprints [90]. Micro-grids have many advantages, including increased robustness to power outages (which occur rarely but cause significant losses), increased use of onsite renewable energy sources, reduced loss from long-distance transmission, and potential economic benefits. This is especially true with the increase of fuel costs, environmental taxes, and incentives for renewable energy. Although it is not necessary for a micro-grid to operate in isolated mode on a regular basis, they are often designed to be self-sustained most of the time.

¹²This chapter is based on P. Yang, and A. Nehorai, “Joint optimization of hybrid energy storage and generation capacity with renewable energy,” to appear in *IEEE Trans. on Smart Grid*. © IEEE 2014.

Most renewable energy sources, including wind and solar, are highly intermittent. The availability of such energy sources varies significantly in different geographical locations. In the same location, the amount of generation also fluctuates depending on the time of day, season, and weather conditions. A grid with high renewable energy penetration needs to build sufficient energy storage to ensure an uninterrupted supply to end users and make the best use of generated energy [91, 92]. There are different types of energy storage, including super-capacitors, flywheels, chemical batteries, pumped hydro, hydrogen, and compressed air [93, 94, 95, 96, 97, 98]. These types differ in round-trip energy efficiency, maximum capacity/power rating, self-discharging, and investment/operational costs. For example, flywheel energy storage has high energy efficiency and charge/discharge rates, but the rate of self-discharging is relatively high. Chemical batteries have relatively high energy efficiency and low self-discharge rate, however their maintenance cost is high due to their low durability, which is quantified by cycling capacity¹³. Pumped hydro and hydrogen energy storages have relatively low energy efficiency (large scale pumped hydro has higher efficiency), but their self-discharging rate is quite small. In addition, the capital cost per MWh of pumped hydro storage is low. Therefore it is often used for longer-term energy storage.

Although there has been research on planning and/or operating a specific type of energy storage system for isolated electricity grids [99, 100, 101, 102], few works consider exploiting the different characteristics of multiple types of energy storage and the different availabilities of multiple types of renewable energy sources, forming a hybrid energy generation and storage system. Nevertheless, jointly planning for energy storage together with renewable generation capacity potentially results in a more economical and efficient energy system.

Since the future grid is becoming decentralized, we consider the scenario of an isolated grid, or a micro-grid with a small carbon footprint, whose energy is generated mainly from renewable energy sources. To make the scenario more practical, we assume the grid also has traditional diesel generators. The diesel generator on its own is insufficient to supply the demand of the grid, as its generation capacity is significantly less than the peak load. We formulate an optimization problem with the objective of minimizing the investment cost and operational/maintenance cost of energy storage and generators. To do this, we find an optimal combination of different energy storages and generators (which we refer to as design parameters) and optimize their operations.

¹³The maximum number of charging cycles (full charge and discharge).

The renewable generation and user demands change with time, and have different characteristics at different times of day and different days of the year. It is often difficult to obtain an accurate probability density function to reflect these complex characteristics. Therefore, several years of historical data may be needed to obtain better optimization results. As the size of the historical database increases, the design horizon of the optimization problem increases, and the problem becomes increasingly difficult to solve. To resolve this problem, we reformulate the original problem as a consensus problem. The entire design horizon is divided into multiple shorter horizons, and thus the design parameters become the consensus parameters, which should be consistent across all sub-problems. This framework can also be extended to the case of solving chance-constrained optimization using scenario approximations, as we will elaborate later. We propose to solve the consensus problem in a parallel distributed manner based on the alternating direction method of multipliers (ADMM) [103], which mitigates the curse of dimensionality due to increased number of scenarios.

The rest of this chapter is organized as follows. In Section 5.2 we briefly review some related work. In Section 5.3 we describe the system model, including the energy storage and generators. In Section 5.4 we formulate the optimization problem and solve it in a distributed manner. We provide numerical examples in Section 5.5, and summarize this chapter in Section 5.6.

5.2 Related work

Here we briefly review several works on optimization with energy storage and renewable generation. In [93], the authors investigated the combined optimization of a wind farm and a pumped storage facility from the perspective of a generation company, using a two-step stochastic optimization approach. The optimization produces optimal bids for the day-ahead spot market, and optimal operation strategies of the facilities. The optimal planning of generation and energy storage capacity was not considered. Zhou *et al.* [95] proposed a composite energy storage system that contains both high energy density storage and high power density storage. The proposed power converter configuration enables actively distributing demands among different energy storages. Brown *et al.* [100] provided an economical analysis of the benefits of having pumped storage in a small island system with abundant renewable energy,

and proposed to find the optimal pumped storage capacity through linear programming. In [101], the authors considered optimizing the capacity of energy storage in a wind-diesel isolated grid, and demonstrated that high wind penetration potentially results in significant cost savings in terms of fuel and operating costs.

The main contributions of our work in this chapter are twofold. First, instead of a single type of energy storage or renewable energy source, we consider a hybrid system with multiple types of energy storage and renewable energy sources, and jointly optimize their capacities and operation. This joint optimization exploits the benefits from each individual element, and therefore is more cost efficient. Second, we propose a distributed optimization framework, so that the capacity design problem becomes scalable when the number of scenarios increases.

5.3 System model

5.3.1 Energy storage model

Assume there is a set \mathcal{S} of different types of energy storages. We use the superscript $s \in \mathcal{S}$ to denote the type of the storage. Each type of energy storage is characterized by a group of parameters. We use η^s to denote the one-way energy efficiency of energy storage type s ; δ^s to denote the ratio between the rated power and rated energy; and ξ^s to denote the self-discharging rate per unit time period. The cost of energy storage includes the initial investment cost c_{inv}^s and operational/maintenance cost $c_{\text{o/m}}^s$. We use a^s to denote the amortization factor.

Let S_t^s denote the energy in storage s at the beginning of time period t , satisfying the following equation:

$$S_{t+1}^s = \begin{cases} S_t^s - \frac{1}{\eta^s} P_t^s - \xi^s S_t^s & \text{if } P_t^s \geq 0, \\ S_t^s - \eta^s P_t^s - \xi^s S_t^s & \text{if } P_t^s < 0, \end{cases} \quad (5.1)$$

where positive P_t^s denotes discharge from storage s during time period t , and negative P_t^s denotes charge to the storage. Make the following substitution:

$$P_t^s = P_t^{s,+} - P_t^{s,-}, \quad P_t^{s,+} \geq 0, \quad P_t^{s,-} \geq 0, \quad (5.2)$$

and we can then rewrite (5.1) as

$$S_{t+1}^s = S_t^s - \frac{1}{\eta^s} P_t^{s,+} + \eta^s P_t^{s,-} - \xi^s S_t^s. \quad (5.3)$$

An interpretation of (5.3) is that the energy stored in a specific energy storage type equals the stored energy at the beginning of the previous time point, minus (plus) the discharge (charge) during the previous time period, minus the energy loss due to the nature of the storage.

The amount of stored energy and the charge/discharge power is constrained by the capacity of the storage, i.e.,

$$0 \leq S_t^s \leq S_{\max}^s, \quad (5.4)$$

$$0 \leq P_t^{s,+} \leq P_{\max}^{s,+}, \quad 0 \leq P_t^{s,-} \leq P_{\max}^{s,-}. \quad (5.5)$$

In this work we use δ^s to denote the ratio between the rated power and the rated storage capacity. Therefore $P_{\max}^{s,+} = \eta^s \delta^s S_{\max}^s$ and $P_{\max}^{s,-} = \frac{1}{\eta^s} \delta^s S_{\max}^s$. If the ratio is not fixed, we can introduce another design variable for the rated power, and modify the investment cost so that it depends on both S_{\max}^s and $\{P_{\max}^{s,+}, P_{\max}^{s,-}\}$.

The cost of each type of energy storage during time period t , denoted by C_t^s , includes the amortized investment cost and the operational/maintenance cost, i.e.,

$$C_t^s = a^s c_{\text{inv}}^s S_{\max}^s + c_{\text{o/m}}^s(P_t^{s,+}, P_t^{s,-}). \quad (5.6)$$

In this equation, the operational/maintenance cost depends on the amount of charge and discharge, and any fixed cost can be included as a constant term in this cost function.

Note that we made substitution (5.2), and therefore constraints (5.1) and (5.3) are equivalent if only one element of each pair $\{P_t^{s,+}, P_t^{s,-}\}$ is non zero for all s, t . Based on the problem setup, we have the following theorem.

Theorem 5.1. *In a cost minimization context, given an increasing positive operational cost function for charging and discharging, and by making the substitution (5.2), we have that $P_t^{s,+} P_t^{s,-} = 0$ for all t ; i.e., only one of $P_t^{s,+}$ and $P_t^{s,-}$ can be non-zero for any given time period t .*

Proof. Let $\{P_t^{s,+}, P_t^{s,-}\}$ and $\{\hat{P}_t^{s,+}, \hat{P}_t^{s,-}\}$ denote two charge/discharge pairs satisfying

$$\begin{aligned} P_t^{s,+} - P_t^{s,-} &= P_t^s, & P_t^{s,+} P_t^{s,-} &= 0, \\ \hat{P}_t^{s,+} - \hat{P}_t^{s,-} &= P_t^s, & \hat{P}_t^{s,+} \hat{P}_t^{s,-} &> 0. \end{aligned} \tag{5.7}$$

We then have

$$\begin{aligned} (P_t^{s,+} + P_t^{s,-})^2 &= (P_t^{s,+} - P_t^{s,-})^2 + 4P_t^{s,+} P_t^{s,-} \\ &< (\hat{P}_t^{s,+} - \hat{P}_t^{s,-})^2 + 4\hat{P}_t^{s,+} \hat{P}_t^{s,-} \\ &= (\hat{P}_t^{s,+} + \hat{P}_t^{s,-})^2. \end{aligned} \tag{5.8}$$

Since $P_t^{s,+} - P_t^{s,-} = \hat{P}_t^{s,+} - \hat{P}_t^{s,-}$, we then have $P_t^{s,+} < P_t^{s,+}$, and $P_t^{s,-} < P_t^{s,-}$. Because the operational cost is an increasing function of $\{P_t^{s,+}, P_t^{s,-}\}$, we obtain that

$$c_{o/m}(P_t^{s,+}, P_t^{s,-}) < c_{o/m}(\hat{P}_t^{s,+}, \hat{P}_t^{s,-}). \tag{5.9}$$

Therefore the optimal pair $\{P_t^{s,+}, P_t^{s,-}\}$ must satisfy $P_t^{s,+} P_t^{s,-} = 0$; i.e., only one of $P_t^{s,+}, P_t^{s,-}$ can be non-zero. \square

Therefore we prove that the two constraints (5.1) and (5.3) are indeed equivalent.

5.3.2 Generator model

The generators are classified into traditional diesel generators and renewable generators. For diesel generators, the constraints include the generation capacity and generator ramp constraints. Let \mathcal{H} denote the set of all diesel generators, and H_t^h denote the generation of

generator type $h \in \mathcal{H}$ during time period t . We then have

$$0 \leq H_t^h \leq H_{\max}^h, \quad (5.10)$$

$$H_{\text{ramp}}^{h,-} \leq H_{t+1}^h - H_t^h \leq H_{\text{ramp}}^{h,+}, \quad (5.11)$$

where H_{\max}^h denotes the maximum generation capacity, and $H_{\text{ramp}}^{h,-}$ and $H_{\text{ramp}}^{h,+}$ denote ramp down and ramp up constraints, respectively. The cost of diesel generators consists of the amortized investment cost and the operational/maintenance cost, denoted by

$$C_t^h = a^h c_{\text{inv}}^h H_{\max}^h + c_{\text{o/m}}^h(H_t^h). \quad (5.12)$$

Usually a second-order quadratic function or piece-wise linear function is used for $c_{\text{o/m}}^h(H_t^h)$. Any environmental tax can also be included in this cost function.

We employ multiple types of renewable generators, including wind and solar, which are considered as non-dispatchable generations. Let R_t^r denote the renewable generation from type $r \in \mathcal{R}$ generator during time period t , and R_{\max}^r denote the installed capacity. Then the generation can be written as

$$R_t^r = r_t^r R_{\max}^r, \quad (5.13)$$

where r_t^r is a random variable denoting the renewable generation per unit generation capacity. The cost for renewable energy during time period t is then

$$C_t^r = a^r c_{\text{inv}}^r R_{\max}^r + c_{\text{o/m}}^r(R_t^r). \quad (5.14)$$

In addition to the generator types we discuss here, other types of generators, e.g., hydro and nuclear generators, can also be modeled similarly and included in the planning problem.

5.3.3 Load balance constraint

The total generation should equal the total demand in a power grid at all times. Let G_t denote the energy shortage for an isolated grid, or the energy drawn from the main grid for a micro-grid. The total generation and discharge from the energy storages should be equal

to the total consumption and charge to the energy storages. We can then write the load balance constraint as follows:

$$D_t = \sum_{r \in \mathcal{R}} R_t^r + \sum_{h \in \mathcal{H}} H_t^h + \sum_{s \in \mathcal{S}} (P_t^{s,+} - P_t^{s,-}) + G_t, \forall t, \quad (5.15)$$

where D_t denotes the demand from users. Note that G_t can be negative, which denotes energy injection to the main grid from a micro-grid, or dumped energy in an isolated grid.

5.4 Storage and renewable generation planning

5.4.1 Optimal planning problem

The planning goal is to find the optimal portfolio of different types of energy storage and generators, so that the total cost (including investment and operational/maintenance) is minimized, while the needs of the grid can be satisfied. Let \mathcal{T} denote the planning horizon, and the objective function can then be written as

$$f(\mathbf{S}_{\max}, \mathbf{R}_{\max}, \mathbf{H}_{\max}) = \sum_{t \in \mathcal{T}} \left(\sum_{s \in \mathcal{S}} C_t^s + \sum_{r \in \mathcal{R}} C_t^r + \sum_{h \in \mathcal{H}} C_t^h \right). \quad (5.16)$$

Due to the intermittency of renewable energy sources, it is possible that in extreme cases, the total local generation will not meet the total demand. We write the grid reliance constraint (for micro-grids) or the energy shortage constraint (for isolated grids) as

$$G_t \leq G_{\text{th}}, \quad (5.17)$$

where G_{th} is a threshold which can be a function of current time and demand. There are also constraints on the minimum and maximum capacity for each type of storage and generator,

which are denoted as

$$\begin{aligned}
S_{\min}^{\text{s,cap}} &\leq S_{\max}^{\text{h}} \leq S_{\max}^{\text{s,cap}}, \forall \text{s}, \\
R_{\min}^{\text{r,cap}} &\leq R_{\max}^{\text{h}} \leq R_{\max}^{\text{r,cap}}, \forall \text{r}, \\
H_{\min}^{\text{h,cap}} &\leq H_{\max}^{\text{h}} \leq H_{\min}^{\text{h,cap}}, \forall \text{h}.
\end{aligned} \tag{5.18}$$

We then formulate the optimization problem for energy planning as

$$\begin{aligned}
&\min_{\mathbf{S}_{\max}, \mathbf{R}_{\max}, \mathbf{H}_{\max}} f(\mathbf{S}_{\max}, \mathbf{R}_{\max}, \mathbf{H}_{\max}) \\
\text{subject to} \quad &S_{t+1}^{\text{s}} = S_t^{\text{s}} - \frac{1}{\eta^{\text{s}}} P_t^{\text{s},+} + \eta^{\text{s}} P_t^{\text{s},-} - \xi^{\text{s}} S_t^{\text{s}}, \quad \forall \text{s}, t, \\
&0 \leq S_t^{\text{s}} \leq S_{\max}^{\text{s}}, \quad \forall \text{s}, t, \\
&0 \leq P_t^{\text{s},+} \leq P_{\max}^{\text{s},+}, 0 \leq P_t^{\text{s},-} \leq P_{\max}^{\text{s},-}, \quad \forall \text{s}, t, \\
&0 \leq H_t^{\text{h}} \leq H_{\max}^{\text{h}}, \quad \forall \text{h}, t, \\
&H_{\text{ramp}}^{\text{h},-} \leq H_{t+1}^{\text{h}} - H_t^{\text{h}} \leq H_{\text{ramp}}^{\text{h},+}, \quad \forall \text{h}, t, \\
&R_t^{\text{r}} = r_t^{\text{r}} R_{\max}^{\text{r}}, \quad \forall \text{r}, t, \\
&D_t = \sum_{\text{r} \in \mathcal{R}} R_t^{\text{r}} + \sum_{\text{h} \in \mathcal{H}} H_t^{\text{h}} + \sum_{\text{s} \in \mathcal{S}} (P_t^{\text{s},+} - P_t^{\text{s},-}) + G_t, \quad \forall t, \\
&G_t \leq G_{\text{th}}, \quad \forall t \\
&S_{\min}^{\text{s,cap}} \leq S_{\max}^{\text{h}} \leq S_{\max}^{\text{s,cap}}, \quad \forall \text{s}, \\
&R_{\min}^{\text{r,cap}} \leq R_{\max}^{\text{h}} \leq R_{\max}^{\text{r,cap}}, \quad \forall \text{r}, \\
&H_{\min}^{\text{h,cap}} \leq H_{\max}^{\text{h}} \leq H_{\min}^{\text{h,cap}}, \quad \forall \text{h}.
\end{aligned} \tag{5.19}$$

Variables such as the charging and discharging schedule of different energy storages, generation of diesel generators, etc. are all optimization variables. Since they are not design parameters of interest, we omit them in (5.19) for notational simplicity.

Additional costs and constraints can also be easily included in this formulation. For example, an environmental tax for traditional diesel generators, as well as government incentives for renewable generations, can be included in the corresponding cost functions. The maximum allowed diesel generation capacity specified by certain energy policies can be included in the generator constraints.

Remark 5.1. *The problem formulation can be slightly modified into a chance-constrained problem. Instead of the deterministic constraint (5.17), we can use the following probabilistic constraint:*

$$\Pr(G_t \geq G_{\text{th}}) \leq \alpha, \quad (5.20)$$

where $\alpha \in [0, 1]$ is the maximal energy shortage probability allowed. Constraint (5.20) means that local generators and storages have a probability less than or equal to α to be short of energy greater than G_{th} . In this case, using the results from [104], [105], the probabilistic constraint can be approximated by a set of deterministic constraints, sampled from the probability distribution of the random parameters from the probabilistic constraint. To be more specific, let each scenario be a random realization of load, renewable generation, and initial conditions of the energy storages. The number of required scenarios $J = \text{card}(\mathcal{J})$ is determined by the number of design parameters and the probability measure. Let N denote the number of design parameters. According to [104], if the number of scenarios J is no less than $\lceil 2N\alpha^{-1} \ln(2\alpha^{-1}) + 2\alpha^{-1} \ln(\epsilon^{-1}) + 2N \rceil$, then the solution to the scenario approximation problem has a probability of at least $1 - \epsilon$ to satisfy the original chance constraint. The problem formulation and method of solving the problem are very similar to (5.19). We will point out the difference in Remark 5.2. For examples using scenario approximation to solve chance constrained optimization, please refer to [106] and [107], where the authors employed this framework to solve the problem of optimizing distributed renewable energy source management.

5.4.2 Formulation of consensus problem

The renewable generation and user loads in (5.19) are all random. In practice, historical data is used in the problem formulation. With a large number of realizations of the random parameters from historic data, the problem becomes increasingly difficult to solve due to the increase of dimensionality. In the rest of this section, we will reformulate the original problem (5.19) as a consensus problem, which can be solved in a distributed manner.

We divide the entire planning horizon \mathcal{T} into sub planning horizons \mathcal{T}^j , which we call scenarios for simplicity. Let \mathcal{J} denote the set of all sub horizons, and we have that $\mathcal{T} = \cup_{j \in \mathcal{J}} \mathcal{T}^j$. For convenience, we assume \mathcal{T}^j 's are arranged in the order of time. Let $\mathbf{x}_d^j = [\mathbf{S}_{\text{max}}^j; \mathbf{R}_{\text{max}}^j; \mathbf{H}_{\text{max}}^j]$

denote the design parameters for the j th scenario, $\mathbf{z}_d = [\mathbf{S}_{\max}; \mathbf{R}_{\max}; \mathbf{H}_{\max}]$ denote the global design parameters, and \mathcal{C}^j denote the feasible set for the design variables of the j th scenario, with $\mathcal{C} = \cap_{j \in \mathcal{J}} \mathcal{C}^j$.

In practice, the energy in storages at the beginning of each time period is not random, but rather depends on the energy from the previous time period. Assuming the energy stored at the beginning of a scenario should be equal to the energy stored at the end of the previous scenario, we need additional constraints to ensure this condition is satisfied. Let \mathbf{S}_0^j denote the energy storage at the beginning of the j th scenario, and \mathbf{S}_T^j denote the energy storage at the end of the j th scenario. In [108], we followed the approach in [100], and imposed an additional assumption that the energy in each energy storage at the end of the optimization horizon should be equal to that at the beginning of the optimization horizon, i.e., $\mathbf{S}_0^j = \mathbf{S}_T^j, j \in \mathcal{J}$. However, this assumption makes the solution suboptimal. In this work, we eliminate this assumption and add additional consensus constraints across scenarios.

Let $\mathbf{x}_b^j = [\mathbf{S}_0^j; \mathbf{S}_T^j]$ denote the boundary parameters for the j th scenario, and \mathbf{z}_b denote the global boundary parameters. Let $\mathcal{B}(j)$ denote the mapping for the indices of the boundary conditions for the j th scenario. To be specific, $\mathbf{z}_{b, \mathcal{B}(j)}$ denotes the global boundary parameters corresponding to $[\mathbf{S}_T^{j-1}; \mathbf{S}_0^{j+1}]$. We also use the scalar function $B(j, i)$ to denote element-wise index mapping, i.e., $(\mathbf{x}_b^j)_i$ corresponds to $(\mathbf{z}_b)_{B(j, i)}$. The constraints $\mathbf{S}_0^j = \mathbf{S}_T^{j-1}, j \in \mathcal{J}$ can then be written as $\mathbf{x}_b^j = \mathbf{z}_{b, \mathcal{B}(j)}, j \in \mathcal{J}$.

Using the notations $\mathbf{x}^j = [\mathbf{x}_d^j; \mathbf{x}_b^j]$, $\mathbf{z} = [\mathbf{z}_d; \mathbf{z}_b]$, and $\tilde{\mathbf{z}}^j = [\mathbf{z}_d; \mathbf{z}_{b, \mathcal{B}(j)}]$, we then formulate the original optimization problem (5.19) as follows:

$$\begin{aligned} \min_{\mathbf{x}_d^j \in \mathcal{C}^j, \mathbf{x}_b^j} \quad & \sum_{j \in \mathcal{J}} f^j(\mathbf{x}^j) \\ \text{subject to} \quad & \mathbf{x}^j = \tilde{\mathbf{z}}^j, \quad j \in \mathcal{J}. \end{aligned} \tag{5.21}$$

The global design parameters \mathbf{z}_d from solving (5.21) will satisfy $\mathbf{z}_d \in \mathcal{C}$. Note that, similar to (5.19), we omit optimization variables that are not design variables, for notational simplicity.

Remark 5.2. *If the probabilistic constraint is considered, and the scenario approximation approach is used, the formulation has to be slightly revised. According to [104], the random samples for each scenario have to be generated from independent identical distributions. Note*

that the starting energy in the storages also has to be drawn from certain probability distributions. The consensus formulation for the energy storage boundary conditions can then be removed. The number of generated scenarios has to be greater than or equal to the minimum number described in Remark 5.1.

5.4.3 Distributed optimization

The challenge in solving (5.21) is that as the number of scenarios increases, the problem becomes increasingly difficult due to high time complexity. We propose to solve the problem in a distributed manner, based on the alternating direction method of multipliers (ADMM) [103], which mitigates the time complexity issue and makes the problem scalable.

To enforce the equality (consensus) constraint in (5.21), an additional quadratic term is added to the original Lagrangian, forming an augmented Lagrangian, which can be written as

$$L_\rho(\{\mathbf{x}^j\}, \mathbf{z}, \{\mathbf{v}^j\}) = \sum_{j \in \mathcal{J}} \left(f^j(\mathbf{x}^j) + \mathbf{v}^{j\top}(\mathbf{x}^j - \tilde{\mathbf{z}}^j) + \frac{\rho}{2} \|\mathbf{x}^j - \tilde{\mathbf{z}}^j\|_2^2 \right), \quad (5.22)$$

where $\{\mathbf{v}^j\}$ denote the dual variables, and ρ is a pre-defined parameter which is the dual variable update step size. The quadratic term penalizes the difference between the local variables $\{\mathbf{x}^j\}$ and corresponding entries of the global variable \mathbf{z} , denoted by $\tilde{\mathbf{z}}^j$.

The ADMM algorithm iterates among the following steps, with the subscript k denoting the iteration number.

x-minimization step

For each $j \in \mathcal{J}$, the following local minimization problems are solved in parallel:

$$\mathbf{x}_{k+1}^j = \underset{\mathbf{x}_d^j \in \mathcal{C}^j, \mathbf{x}_b^j}{\operatorname{argmin}} f(\mathbf{x}^j) + \mathbf{v}^{j\top}(\mathbf{x}^j - \tilde{\mathbf{z}}_k^j) + \frac{\rho}{2} \|\mathbf{x}^j - \tilde{\mathbf{z}}_k^j\|_2^2. \quad (5.23)$$

z-minimization step

After the **x-minimization** step, the following problem is solved

$$\mathbf{z}_{k+1} = \underset{\mathbf{z}}{\operatorname{argmin}} \sum_{j \in \mathcal{J}} \left(\mathbf{v}^{j\top} (\mathbf{x}_{k+1}^j - \tilde{\mathbf{z}}^j) + \frac{\rho}{2} \|\mathbf{x}_{k+1}^j - \tilde{\mathbf{z}}^j\|_2^2 \right). \quad (5.24)$$

To solve for the **z-minimization** step, we consider \mathbf{z}_b and \mathbf{z}_d separately. Decompose $\mathbf{v}^j = [\mathbf{v}_d^j; \mathbf{v}_b^j]$, and we then rewrite (5.24) as

$$\begin{aligned} \mathbf{z}_{k+1} = \underset{\mathbf{z}}{\operatorname{argmin}} \sum_{j \in \mathcal{J}} & \left(\mathbf{v}_{d,k}^{j\top} (\mathbf{x}_{d,k+1}^j - \mathbf{z}_d) + \frac{\rho}{2} \|\mathbf{x}_{d,k+1}^j - \mathbf{z}_d\|_2^2 + \right. \\ & \left. \mathbf{v}_{b,k}^{j\top} (\mathbf{x}_{b,k+1}^j - \mathbf{z}_{b,B(j)}) + \frac{\rho}{2} \|\mathbf{x}_{b,k+1}^j - \mathbf{z}_{b,B(j)}\|_2^2 \right). \end{aligned} \quad (5.25)$$

Solving (5.25), we obtain that

$$\mathbf{z}_{d,k+1} = \frac{1}{J} \sum_{j \in \mathcal{J}} \left(\mathbf{x}_{k+1}^j + \frac{1}{\rho} \mathbf{v}_{d,k}^j \right), \quad (5.26)$$

$$(\mathbf{z}_{b,k+1})_g = \frac{\sum_{B(j,i)=g} ((\mathbf{x}_{b,k+1}^j)_i + (1/\rho)(\mathbf{v}_{b,k}^j)_i)}{\sum_{B(j,i)=g} 1}. \quad (5.27)$$

When the algorithm converges, the resulting global design variable \mathbf{z}_d has to satisfy the constraints of each sub-problem, i.e., $\mathbf{z}_d \in \mathcal{C}^j, \forall j \in \mathcal{J}$. Therefore we have that $\mathbf{z}_d \in \mathcal{C}$.

Dual-variable update

For each $j \in \mathcal{J}$, the dual variables are updated in parallel after the **x-minimization** and the **z-minimization** steps are finished:

$$\mathbf{v}_{k+1}^j = \mathbf{v}_k^j + \rho (\mathbf{x}_{k+1}^j - \tilde{\mathbf{z}}_{k+1}). \quad (5.28)$$

Since (5.23) and (5.28) can be parallelized, the problem is scalable as the number of scenarios increases. The convergence of this approach is guaranteed, as proved in [103]. For faster convergence, we use an adaptive dual update stepsize ρ . The primal residual \mathbf{r}_k^p and dual residual \mathbf{r}_k^d are defined as

$$\|\mathbf{r}_{k+1}^p\|^2 = \frac{1}{J} \sum_{j \in \mathcal{J}} \|\mathbf{x}_{k+1}^j - \tilde{\mathbf{z}}_{k+1}^j\|^2, \quad (5.29)$$

$$\|\mathbf{r}_{k+1}^d\|^2 = \rho^2 \|\mathbf{z}_{k+1} - \mathbf{z}_k\|^2. \quad (5.30)$$

As a larger ρ penalizes more on the primal residual, and a smaller ρ penalizes on the dual residual, the parameter ρ is updated following the rule below:

$$\rho_{k+1} = \begin{cases} \tau \rho_k & \text{if } \|\mathbf{r}_k^p\| > \mu \|\mathbf{r}_k^d\|, \\ \rho_k / \tau & \text{if } \|\mathbf{r}_k^d\| > \mu \|\mathbf{r}_k^p\|, \\ \rho_k & \text{otherwise,} \end{cases} \quad (5.31)$$

where $\tau > 1$, $\mu > 1$. The algorithm converges when both the primal and dual residual are less than a certain threshold.

5.5 Numerical examples

In this section, we provide a series of numerical examples using real data from online databases, to showcase how the proposed framework can help in making decisions on renewable generation and energy storage planning.

5.5.1 Data and parameters

Renewable generation data

We consider two types of renewable generation, wind and solar, and simulate renewable generation data using the National Solar Radiation Data Base (NSRDB) from NREL [109], [110]. The database provides hourly solar radiation data as well as wind speed data. The solar generation is calculated using the hourly “modeled global horizontal radiation”. Denote the global horizontal radiation as R_g^s . The power received on panel R_s is then calculated by

$$R_s = \min(\mu^s R_g^s, R_r^s), \quad (5.32)$$

where μ_s is the solar panel efficiency, and R_r^s is the rated power output.

The wind power output is calculated using the following equation:

$$R_w = \begin{cases} \frac{1}{2} \mu_r \rho_{\text{air}} V^3 \frac{\pi d^2}{4} & \text{if } V_{\text{in}} \leq V \leq V_{\text{rated}}, \\ \frac{1}{2} \mu_r \rho_{\text{air}} V_{\text{rated}}^3 \frac{\pi d^2}{4} & \text{if } V_{\text{rated}} \leq V \leq V_{\text{out}}, \\ 0 & \text{otherwise,} \end{cases} \quad (5.33)$$

where ρ_{air} is the density of air, d is the diameter of the wind turbine, and μ_r is the wind turbine efficiency. The generation output is zero when the wind speed is lower than the cut-in speed V_{in} or higher than the cut-out speed V_{out} .

Both the solar and wind generations are normalized by the rated power outputs. The typical capital costs and life span for renewable [111] and diesel generators are included in Table 5.1. We use linear functions of the generations to model operational and management costs for solar and wind generators, and a quadratic function for the diesel generator.

Load data

We use the ERCOT hourly load data from [112], normalized by the average hourly demand.

Table 5.1: Parameters for energy generators.

Type	Solar panel (R^1)	Wind turbine (R^2)	Diesel generator (H^1)
Investment cost (M\$/MWh)	5.284	2.414	0.400
Life span (years)	30	20	5
Operation/management cost	linear	linear	quadratic

Table 5.2: Parameters for energy storage.

Type	Advanced lead acid (S^1)	Li-ion battery (S^2)	Pumped storage (S^3)
Round-trip efficiency	0.85	0.85	0.70
Full charge time (hours)	1.00	4.00	10.00
Self-discharging rate [†]	0.01	0.01	0.00
Investment cost (M\$/MWh)	3.200	1.425	0.450
Life span (years)	15	15	50
Operation/management cost	linear	linear	linear

[†] This parameter was not provided in [113].

Energy storage parameters

We select three types of energy storage as prototypes for our simulations, including advanced lead acid batteries, Li-ion batteries, and pumped storage. The corresponding parameters are determined based on [113] and included in Tables 5.1 and 5.2. Due to the limitation of available load and renewable generation data, the time scale considered is limited to one hour. With finer time-scale data available, other types of energy storage, for example flywheel storage, can be considered. Note that the minimum and maximum capacity allowed for each type of energy storage are selected to be 0 and 10 MWh, respectively, for illustrative purposes. The choice of such constraints does not affect the proposed framework. In practice, the minimum and maximum capacity requirements can be determined by various aspects, including space and area availability, cost limitations, operating constraints, etc.

Other parameters

In the test case, we use three-year data, spanning 1095 days. We set the sub planning horizon to be 5 days, and thus $T = 24 \times 5 = 120$. There are $J = 219$ sub planning horizons in total. For simplicity, we use linear amortization for the investment cost, where the investment cost for each year is the total investment cost divided by the entire life span of the technology.

Definitions

To quantify G_{th} and the maximum capacity of the diesel generator $H_{\text{max}}^{\text{cap}}$, we define two quantities: the *shortfall-to-demand ratio*, r_{SD} , and the *diesel generation capacity ratio*, r_{DC} . We define the threshold G_{th} at time t as

$$G_{\text{th}} = r_{\text{SD}} D_t, \quad (5.34)$$

and therefore the shortfall-to-demand ratio is the ratio between the threshold G_{th} and the current demand. The maximum diesel generator capacity $H_{\text{max}}^{\text{cap}}$ is determined by

$$H_{\text{max}}^{\text{cap}} = r_{\text{DC}} \max(D_t), \quad (5.35)$$

and therefore the diesel generation capacity ratio is the ratio between the maximum diesel generator capacity and the peak demand.

5.5.2 Results of storage and generation planning

In this subsection we perform a case study using the setup described in Section 5.5.1. Solar panel efficiency is set to be 20%, with a rated power output of 150W/m². Wind turbine cut-in and cut-out wind speeds are set to be 3m/s and 20m/s, with rated power output achieved at 10m/s. The wind turbine efficiency is set to be 50%. In the following simulations, we consider the average hourly load to be unit megawatt (1MW) for illustrative purposes, while a micro-grid is usually on the scale of 5-10MW. We use the ERCOT hourly load data from the years 2008-2010 to generate the load data, and the NSRDB data to generate the renewable

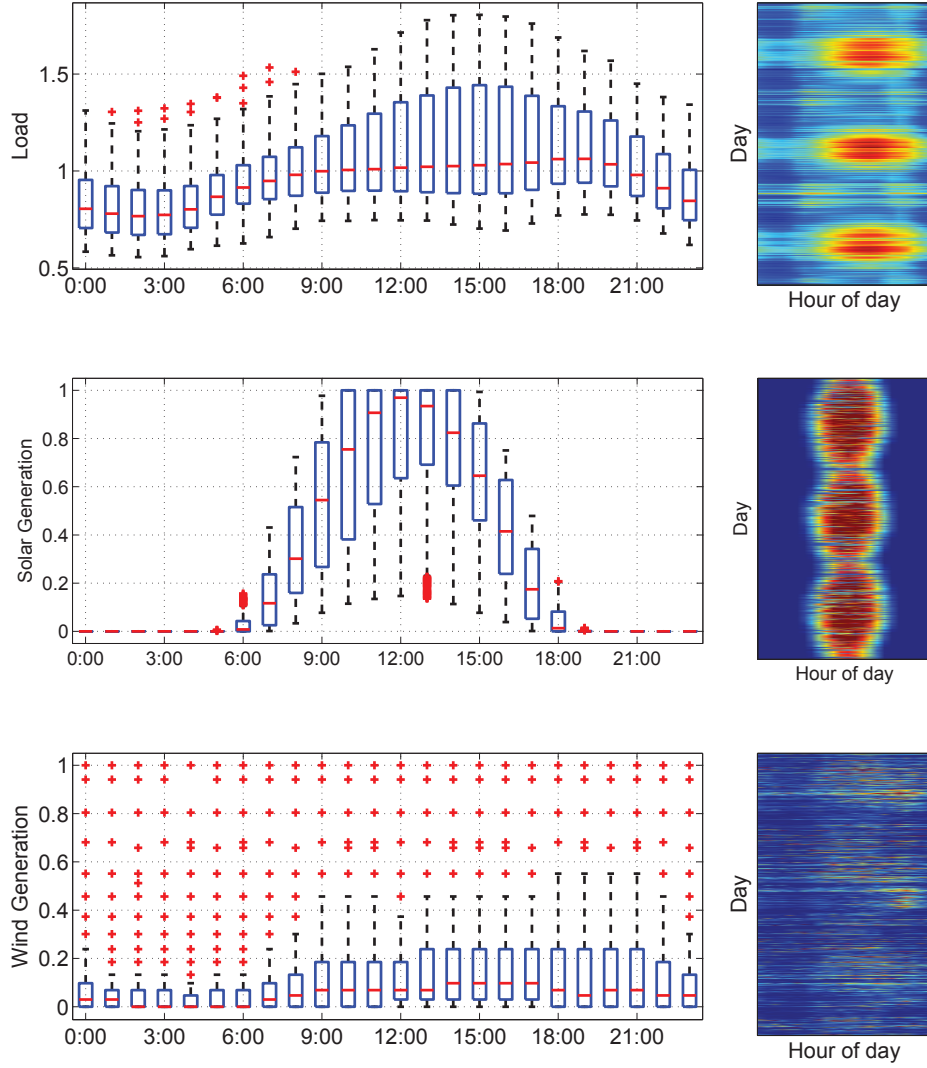


Figure 5.1: Plots for normalized load, solar generation, and wind generation data. On the left are box plots, and on the right are raw data heat maps for three years.

generation data. The maximum hourly load is 1.8050 MW, and the minimum hourly load is 0.5557 MW. The average hourly generations from each unit MW of wind turbines and solar panels are 0.1217 MWh and 0.2534 MWh, respectively. The normalized data (in heatmap) and corresponding box plots are shown in Fig. 5.1.

We set $r_{DC} = 0.50$ and $r_{SD} = 0.05$, meaning that the maximum allowed diesel generation capacity is half of the peak load, and the maximum energy shortfall to demand from local

Table 5.3: Results for $r_{DC} = 0.50$, $r_{SD} = 0.05$ over a three-year design horizon.

Type	Planned capacity (MWh)	Investment cost (M\$)	O./M. cost (M\$)	Total cost (M\$)
S^1	0.3432	0.2196	0.0068	0.2265
S^2	2.0375	0.5807	0.0221	0.6028
S^3	4.5203	0.1220	0.0925	0.2145
R^1	2.0855	1.1020	0.6946	1.7966
R^2	2.6010	0.9418	0.0416	0.9834
H^1	0.9025	0.2166	1.5596	1.7762
Total	—	3.1828	2.4171	5.5999

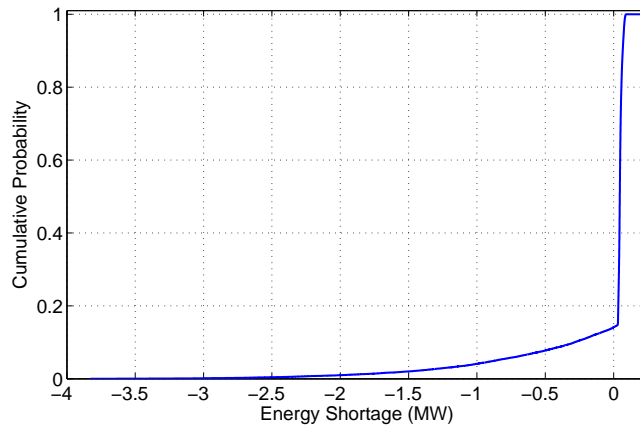


Figure 5.2: Hourly energy shortage distribution. Note that a negative value denotes that generation is greater than demand.

generators is 5% of the demand. The CVX toolbox [114] is used to solve for the **x-updates**. The optimization results are available in Table 5.3. All costs are for a three-year horizon.

Although the planned diesel generation capacity is high (equivalent to the upper limit, which is half of the peak load), the actual average generation from this generator is relatively low. Overall, only 27.36% of the consumed energy is from the diesel generator, i.e., renewable energy constitutes 72.64% of the total consumed energy¹⁴.

¹⁴We assume the renewable generators are generating as much energy as they can, and thus there is excess generation on certain occasions. Only the actual energy consumed by the end users is counted here.

Table 5.4: Results for different diesel generation costs

Diesel generation cost factor	0.50	1.00	1.50	2.00
Diesel capacity (MW)	1.4690	1.2933	1.1539	1.1653
Renewable percentage (%)	25.16	48.65	61.37	68.65
Optimal total cost (M\$)	3.3294	5.1291	6.3126	7.2419
Diesel only total cost (M\$)	3.7463	7.0594	10.3725	13.6856

The empirical cumulative probability function of hourly energy shortage is plotted in Fig. 5.2. For most of the time, the energy shortage G_t is close to zero, meaning that the total generation (including storage discharge) is close to the total demand (including storage charge). There are also occasions when there is excess generation, which is either dumped in isolated grids or injected to the main grid in grid-connected micro-grids.

5.5.3 Results with different diesel generation costs

The diesel generation cost affects the trade-off between renewable energy and traditional fossil energy. In Table 5.4 we show the results with different diesel generation cost factors (based on the cost in Section 5.5.2). We do not place constraints on the diesel capacity ratio r_{DC} in the simulations. It can be observed that as the diesel generation cost increases, the percentage of renewable energy significantly increases. The planned diesel capacity, however, does not decrease significantly, because high diesel capacity is needed for hours with extremely high demand or low renewable generation. This analysis is helpful in designing environmental taxes on diesel generation or incentives for renewable generation, to achieve a certain renewable percentage goal.

Compared to the pure diesel generation case (i.e., a case where all the energy comes from diesel generators), we observe that the optimal cost considering renewable energy sources is always lower, which provides economic incentives to include renewable energy.

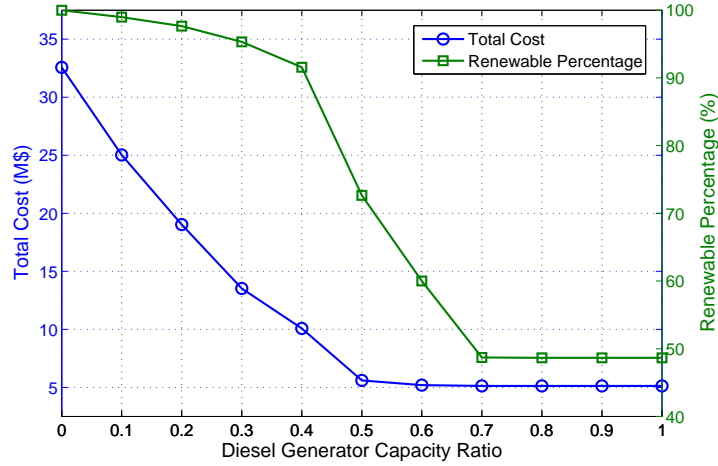


Figure 5.3: Total cost and percentage of renewable generation as a function of diesel generation capacity ratio.

5.5.4 Results with different diesel generation capacities

The maximum allowed diesel generation capacity affects the planning for renewable generation and energy storages. Intuitively, the lower the allowed diesel generation capacity, the more renewable generators and energy storages are required to ensure an uninterrupted energy supply. In this section, we consider different diesel capacity ratios r_{DC} while keeping other parameters fixed. The diesel capacity ratio r_{DC} is varied from 0.0 to 1.0, in increments of 0.1. The resulting total costs and average renewable generation percentages over the three planning years are illustrated in Fig. 5.3.

It can be observed that when r_{DC} is smaller than 0.5 and decreases, the total cost increases rapidly. However, the increase of the renewable percentage is less significant when r_{DC} is less than 0.4. When r_{DC} is greater than 0.7, both the cost and renewable percentage do not change since the unconstrained optimal diesel capacity ratio is approximately 0.7165. This analysis potentially provides guidelines for determining the capacity of diesel generators in a micro-grid, when balancing financial cost and environmental cost.

Table 5.5: Geographic locations and climates.

City	GPS Coordinates	Climate type
San Antonio	29°25'N 98°30'W	humid subtropical / hot semi-arid
St. Louis	38°37'N 90°11'W	humid continental / subtropical
San Francisco	37°47'N 122°25'W	cool-summer Mediterranean



Figure 5.4: Geographical locations of the three cities for comparison.

5.5.5 Results with data from different geographic locations

The geographic location has a significant impact on the availability of different renewable energy sources and the load structures. We consider the case of microgrids in three cities in the United States, described in Table 5.5. The hourly load data for St. Louis and San Francisco are based on the MISO daily report [88] and CAISO daily report [115], respectively. As in Section 5.5.3, we do not limit the diesel generation capacity that can be planned. It can be observed from Table 5.6 that the availability of different resources and the differences in load structures significantly affect the planned capacity of different renewable generators and energy storage. For example, relatively lower solar generation capacity is planned for San Francisco, most likely because the summer daytime air-conditioning load is much smaller, and less solar generation is needed to offset the peak hour loads.

Remark 5.3. *To make better use of renewable generation, it would be more efficient to consider an interconnected network of multiple micro-grids, with inter-grid energy transmission. In this way, less renewable generation and energy storage capacity might be needed to satisfy*

Table 5.6: Comparison of optimization results for different geographic locations.

	San Antonio	St. Louis	San Francisco
S^1 (MWh)	0.0539	0.0338	0.0242
S^2 (MWh)	0.1611	0.0342	0.2696
S^3 (MWh)	1.1535	1.2295	1.2829
R^1 (MWh)	1.1593	1.1054	0.7312
R^2 (MWh)	1.6273	1.6243	1.9418
H^1 (MWh)	1.2933	1.2660	1.3393
Total cost (M\$)	5.1291	5.2185	4.7830
Renewable (%)	48.65	44.78	50.61

the energy demands. The optimization problem will more complicated, and new issues, e.g., long-distance power transmission, need to be considered.

5.6 Summary

In this chapter, we considered the problem of jointly optimizing the capacities of multiple energy storages, renewable generators, and diesel generators, in the context of an isolated grid, or a micro-grid with a small carbon footprint. The joint optimization exploits the different characteristics of multiple energy storage types, as well as the availability of different sources of renewable energy. To mitigate the large dimensionality of the optimization problem due to the use of large volumes of historical data, we formulated the original optimization problem as a consensus problem, which can be solved in a parallel distributed manner. We provided a series of numerical examples to illustrate how the proposed framework can be used for planning purposes in practice. To be more specific, we considered scenarios with different maximum diesel generation capacities and different diesel generation costs, and also compared the different planning results in different geographic regions. The proposed work will be helpful in designing renewable generation and energy storage systems for future decentralized power grids with large renewable penetration, and will help policy makers make decisions related to renewable energy and sustainability. The proposed framework also solves the problem of optimally operating a given hybrid energy storage and generation system, under the assumption of perfect load and renewable generation forecasting.

Chapter 6

Conclusions and Future Work

6.1 Summary and conclusions

In this dissertation we studied state estimation and renewable energy optimization in smart grids. We first considered power system state estimation using PMUs, when phase angle mismatch exists in the measurements. In particular, we built a measurement model that takes into account the phase angle mismatch in PMU measurements. We then developed multiple algorithms to estimate the system state under this model, including the alternating minimization method for the static case, and the parallel Kalman filtering method for the dynamic case. Using numerical examples, we showed that our methods significantly improve state estimation accuracy over traditional methods. We also showed that when a sufficient number of PMUs are installed, the phase mismatch can be largely compensated for by using signal processing techniques.

We next considered the problem of PMU placement in a power grid. In order to understand the effect of PMU location and specifications on the estimation performance, we derived the posterior Cramér-Rao bound on the estimation error, using the proposed measurement model. Based on the bound, we formulated an optimal PMU placement problem, and then solved the problem using a greedy algorithm. We showed that under certain design criteria, the objective function can be written as a non-decreasing submodular function, and thus a performance bound exists for the greedy algorithm. We compared the greedy algorithm with multiple other algorithms and heuristics, and the results demonstrated the effectiveness of the greedy algorithm. Numerical examples also demonstrated that the optimized PMU placement and the consideration of PMU phase mismatch improve state estimation accuracy.

On the topic of optimization for renewable energy integration, we first considered the scenario of individual energy users with on-site renewable generators. We developed a framework for parallel distributed load scheduling for these users, with the goals of reducing peak hour energy consumption and reducing costs to both the utility and the users. We used a soft constraint on the stepsize between consecutive iterations, to coordinate the distributed optimization. Numerical examples demonstrated the reduction in the peak hour load as well as the costs. The results of the parallel distributed optimization equaled the optimal solution from sequential optimization within numerical precision, with reduced information exchange, lower communication costs, and less control complexity. Due to the parallelization of the algorithm, the computation time was significantly less than needed by the commonly employed sequential algorithms.

Finally, we considered the problem of renewable energy integration in a micro-grid. We jointly optimized the capacity of multiple types of renewable generators and energy storages, in order to make the best use of the different availability of renewable energy sources, as well as the different characteristics of the energy storage devices. To mitigate the large dimensionality of the optimization problem due to the use of large volumes of historical data, we formulated the original optimization problem as a consensus problem, which could be solved in a parallel distributed manner. The proposed work will aid designers of renewable generation and energy storage systems in future decentralized power grids. It will also guide policy-making decisions related to renewable energy and sustainability.

6.2 Future directions

In this section, we point out several potential future research directions.

Power system state estimation: In this dissertation we focused on centralized state estimation, but decentralized state estimation and more effective approaches to hybrid state estimation using both the SCADA and PMU measurements are also interesting topics to explore. In addition, state estimators that are robust to sensor failure, transmission delay, and/or malicious data attack are also important in practice and deserve research.

Detection of system abnormality and attack: One of the important tasks of real-time state monitoring is to detect system abnormalities, in order to predict system failure. Because this is crucial for avoiding large scale system failure which leads to blackouts, another potential research area is developing algorithms to detect system abnormalities and cyber attacks using measurements from advanced sensors.

Demand response with renewable energy: In this dissertation we proposed a parallel load scheduling framework in a scenario with renewable distributed generators. In reality, the intermittency of renewable generation and the uncertainty of load are among the most challenging problems. More sophisticated stochastic approaches can be employed to improve the real-time demand response performance.

Renewable energy planning and dispatch: We have studied the generation and storage capacity design problem for micro-grids. Future extensions include the problem of optimally operating such a system, taking into account the stochastic nature of demand and renewable generation. It is also worthwhile to study the case of an interconnected network of multiple micro-grids, each with local energy generation and inter-grid energy transmission, which potentially makes better use of renewable energy.

Human-in-the-loop energy efficient building: Much of the existing literature neglects proper modeling of humans and their effects in increasing energy efficiency in buildings. By incorporating advanced heterogeneous sensors, wireless sensor networks, signal processing and optimal control techniques, and knowledge from psychology, it is possible to build a human-in-the-loop building automation system, to incentivize the use of renewable energy and increase overall building efficiency.

References

- [1] S. Collier, “Ten steps to a smarter grid,” *IEEE Ind. Appl. Mag.*, vol. 16, no. 2, pp. 62–68, 2010.
- [2] Electricity Advisory Committee, “Smart grid: Enabler of the new energy economy,” U.S. Department of Energy, Tech. Rep., Dec. 2008.
- [3] A. Abur and A. G. Expósito, *Power System State Estimation: Theory and Implementation*. New York: Marcel Dekker, Inc., 2004.
- [4] L. Jia, J. Kim, R. Thomas, and L. Tong, “Impact of data quality on real-time locational marginal price,” *IEEE Trans. Power Syst.*, vol. 29, no. 2, pp. 627–636, March 2014.
- [5] J. De La Ree, V. Centeno, J. Thorp, and A. Phadke, “Synchronized phasor measurement applications in power systems,” *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 20–27, Jun. 2010.
- [6] “Renewables: energy you can count on,” Union of Concerned Scientists, Tech. Rep., Apr. 2013. [Online]. Available: http://www.ucsusa.org/assets/documents/clean_energy/Ramping-Up-Renewables-Energy-You-Can-Count-On.pdf
- [7] J. A. Turner, “A realizable renewable energy future,” *Science*, vol. 285, no. 5428, pp. 687–689, 1999.
- [8] “Exploration of high-penetration renewable electricity futures,” National Renewable Energy Laboratory, Tech. Rep., 2012.
- [9] F. Schweppe and E. Handschin, “Static state estimation in electric power systems,” *Proc. IEEE*, vol. 62, no. 7, pp. 972–982, Aug. 1974.
- [10] O. Smith, “Power system state estimation,” *IEEE Trans. Power App. Syst.*, vol. 89, no. 3, pp. 363–379, Mar. 1970.
- [11] A. Debs and R. Larson, “A dynamic estimator for tracking the state of a power system,” *IEEE Trans. Power App. Syst.*, vol. 89, no. 7, pp. 1670–1678, Sep. 1970.
- [12] A. Monticelli, *State Estimation in Electric Power Systems: a Generalized Approach*. New York: Springer, 1999.

- [13] K.-R. Shih and S.-J. Huang, "Application of a robust algorithm for dynamic state estimation of a power system," *IEEE Trans. Power Syst.*, vol. 17, no. 1, pp. 141–147, Feb. 2002.
- [14] M. Zhou, V. Centeno, J. Thorp, and A. Phadke, "An alternative for including phasor measurements in state estimators," *IEEE Trans. Power Syst.*, vol. 21, no. 4, pp. 1930–1937, Nov. 2006.
- [15] A. Jain and N. Shivakumar, "Impact of PMU in dynamic state estimation of power systems," in *40th North American Power Symposium*, Sep. 2008, pp. 1–8.
- [16] S. Chakrabarti and E. Kyriakides, "PMU measurement uncertainty considerations in WLS state estimation," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 1062–1071, Apr. 2012.
- [17] E. Farantatos, G. Stefopoulos, G. Cokkinides, and A. Meliopoulos, "PMU-based dynamic state estimation for electric power systems," in *Power Energy Society General Meeting*, Jul. 2009, pp. 1–8.
- [18] L. Vanfretti, J. H. Chow, S. Sarawgi, and B. Fardanesh, "A phasor-data based state estimation incorporating phase bias correction," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 111–119, Feb. 2011.
- [19] V. Kekatos and G. B. Giannakis, "Distributed robust power system state estimation," *arXiv:1204.0991v2*, Apr. 2012.
- [20] Y.-F. Huang, S. Werner, J. Huang, N. Kashyap, and V. Gupta, "State estimation in electric power grids: Meeting new challenges presented by the requirements of the future grid," *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 33–43, Sep. 2012.
- [21] D. G. Hart, D. Uy, V. Gharpure, D. Novosel, D. Karlsson, and M. Kaba, *PMUs - A new approach to power network monitoring*. ABB Technical Report, 2001.
- [22] A. P. Meliopoulos, V. Madani, and D. N. *et. al*, "Synchrophasor measurement accuracy characterization," North American SynchroPhasor Initiative Performance & Standards Task Team, Tech. Rep., Aug. 2007.
- [23] G. Giorgi and C. Narduzzi, "Performance analysis of Kalman-filter-based clock synchronization in IEEE 1588 networks," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 8, pp. 2902–2909, Aug. 2011.
- [24] M. Kezunovic, A. Sprintson, J. Ren, and Y. Guan, "Signal processing, communication, and networking requirements for synchrophasor systems," in *13th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jun. 2012, pp. 464–468.

- [25] S. Ghiocel, J. Chow, G. Stefopoulos, B. Fardanesh, D. Maragal, B. Blanchard, M. Razanousky, and D. Bertagnolli, “Phasor-measurement-based state estimation for synchrophasor data quality improvement and power transfer interface monitoring,” *IEEE Trans. Power Syst.*, vol. 29, no. 2, pp. 881–888, Mar. 2014.
- [26] IEEE, “IEEE Std C37.118-2005: IEEE standard for synchrophasors for power systems,” 2005.
- [27] M. Begovic, D. Novosel, and B. Djokic, “Issues related to the implementation of synchrophasor measurements,” in *Hawaii International Conference on System Sciences, Proceedings of the 41st Annual*, 2008, pp. 164–164.
- [28] Q. Zhang, X. Luo, D. Bertagnolli, S. Maslennikov, and B. Nubile, “PMU data validation at ISO New England,” in *IEEE Power and Energy Society General Meeting*, Jul. 2013.
- [29] C. Na, D. Obradovic, and R. Scheiterer, “A probabilistic approach to clock synchronization of cascaded network elements,” in *ICASSP 2009*, Apr. 2009, pp. 1793–1796.
- [30] P. Top, M. Bell, E. Coyle, and O. Wasynczuk, “Observing the power grid: Working toward a more intelligent, efficient, and reliable smart grid with increasing user visibility,” *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 24–32, Sep. 2012.
- [31] L. Glielmo, P. Marino, R. Setola, and F. Vasca, “Parallel Kalman filter algorithm for state estimation in bilinear systems,” in *33rd Conference on Decision and Control*, Dec. 1994, pp. 1228–1229.
- [32] V. Kekatos, G. B. Giannakis, and B. Wollenberg, “Optimal placement of phasor measurement units via convex relaxation,” *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1521–1530, Aug. 2012.
- [33] A. Gomez-Exposito, C. Gomez-Quiles, and A. de la Villa Jaen, “Bilinear power system state estimation,” *IEEE Trans. Plasma Sci.*, vol. 27, no. 1, pp. 493–501, 2012.
- [34] C. Gomez-Quiles, H. Gil, A. de la Villa Jaen, and A. Gomez-Exposito, “Equality-constrained bilinear state estimation,” *IEEE Trans. Plasma Sci.*, vol. 28, no. 2, pp. 902–910, 2013.
- [35] J. Zhang, G. Welch, and G. Bishop, “Observability and estimation uncertainty analysis for PMU placement alternatives,” in *North American Power Symposium (NAPS), 2010*, Sep. 2010, pp. 1–8.
- [36] N. Shivakumar and A. Jain, “A review of power system dynamic state estimation techniques,” in *Power System Technology and IEEE Power India Conference, 2008. POWERCON 2008. Joint International Conference on*, Oct. 2008, pp. 1–6.

- [37] H. Beides and G. Heydt, “Dynamic state estimation of power system harmonics using Kalman filter methodology,” *IEEE Trans. Power Del.*, vol. 6, no. 4, pp. 1663–1670, Oct. 1991.
- [38] G. Valverde and V. Terzija, “Unscented kalman filter for power system dynamic state estimation,” *IET Generation, Transmission Distribution*, vol. 5, no. 1, pp. 29–37, Jan. 2011.
- [39] B. D. Moor, “Structured total least squares and L_2 approximation problems,” *Linear Algebra and its Applications*, vol. 188-189, no. 0, pp. 163–205, 1993.
- [40] G. H. Golub and C. F. van Loan, “An analysis of the total least squares problem,” *SIAM J. Numer. Anal.*, vol. 17, no. 6, pp. 883–893, 1980.
- [41] A. Wiesel, Y. Eldar, and A. Yeredor, “Linear regression with gaussian model uncertainty: Algorithms and bounds,” *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2194–2205, Jun. 2008.
- [42] B. Recht, M. Fazel, and P. Parrilo, “Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization,” *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.
- [43] B. Schaffrin and H. B. Iz, “Towards total kalman filtering for mobile mapping,” in *5th International Symposium on Mobile Mapping Technology*, Padua, Italy, Dec. 2007.
- [44] C. E. Davila, “Recursive total least squares algorithms for adaptive filtering,” in *International Conference on Acoustics, Speech, and Signal Processing*, Apr. 1991, pp. 1853–1856.
- [45] T. Basar and G. J. Oldser, *Dynamic Noncooperative Game Theory*. London: Academic Press, 1989.
- [46] L. E. Ghaoui and H. Lebrete, “Robust solutions to least-squares problems with uncertain data,” *SIAM Journal on Matrix Analysis and Applications*, vol. 18, no. 4, pp. 1035–1064, Oct. 1997.
- [47] S. Boyd, L. E. Ghaoui, E. Feron, and V. Balakrishnan, “Linear matrix inequalities in system and control theory,” *SIAM Studies in Applied Mathematics*, vol. 15, 1994.
- [48] B. Gou, “Generalized integer linear programming formulation for optimal PMU placement,” *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 1099–1104, aug. 2008.
- [49] N. Manousakis, G. Korres, and P. Georgilakis, “Taxonomy of PMU placement methodologies,” *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 1070–1077, 2012.

- [50] Q. Li, R. Negi, and M. Ilic, "Phasor measurement units placement for power system state estimation: A greedy approach," in *IEEE Power and Energy Society General Meeting*, Jul. 2011, pp. 1–8.
- [51] Q. Li, T. Cui, Y. Weng, R. Negi, F. Franchetti, and M. Ilic, "An information-theoretic approach to pmu placement in electric power systems," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 446–456, 2013.
- [52] M. Uddin, A. Kuh, A. Kavcic, and T. Tanaka, "Approximate solutions and performance bounds for the sensor placement problem," in *IEEE Smart Grid Communications (SmartGridComm)*, Tainan city, Taiwan, Nov. 2012.
- [53] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions," *Mathematical Programming*, vol. 14, pp. 265–294, 1978.
- [54] P. Yang, Z. Tan, A. Wiesel, and A. Nehorai, "Power system state estimation using PMUs with imperfect synchronization," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 884–892, May 2013.
- [55] P. Tichavsky, "Posterior Cramer-Rao bound for adaptive harmonic retrieval," *IEEE Trans. Signal Process.*, vol. 43, no. 5, pp. 1299–1302, 1995.
- [56] P. Tichavsky, C. Muravchik, and A. Nehorai, "Posterior Cramer-Rao bounds for discrete-time nonlinear filtering," *IEEE Trans. Signal Process.*, vol. 46, no. 5, pp. 1386–1396, 1998.
- [57] H. L. Van Trees, *Detection, Estimation and Modulation Theory I*. Wiley, 2002.
- [58] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Prentice Hall, 1993.
- [59] F. Pukelsheim, *Optimal Design of Experiments*. New York: Wiley, 1993.
- [60] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York: Cambridge Univ. Press, 2004.
- [61] X. Shen and P. K. Varshney, "Sensor selection based on generalized information gain for target tracking in large sensor networks," *arXiv:1302.1616v1*, 2013.
- [62] J. Xu, M. H. F. Wen, V. O. K. Li, and K.-C. Leung, "Optimal PMU placement for wide-area monitoring using chemical reaction optimization," in *Proceedings of Innovative Smart Grid Technologies Conference (IEEE ISGT)*, Feb. 2013.
- [63] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman and Company, 1979.

- [64] C. W. Gellings, “The concept of demand-side management for electric utilities,” *Proc. IEEE*, vol. 73, pp. 1468–1470, 1985.
- [65] M. Alizadeh, X. Li, Z. Wang, A. Scaglione, and R. Melton, “Demand-side management in the smart grid: Information processing for the power switch,” *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 55–67, Sep. 2012.
- [66] R. Hartway, S. Price, and C. K. Woo, “Smart meter, customer choice and profitable time-of-use rate option,” *Energy*, vol. 24, pp. 895–903, 1999.
- [67] Q. Wu, L. Wang, and H. Cheng, “Research of TOU power price based on multi-objective optimization of DSM and costs of power consumers,” in *International Conference on Electric Utility Deregulation, Restructuring and Power Technologies*, Hong Kong, Apr. 2004, pp. 343–348.
- [68] E. Çelebi and J. D. Fuller, “A model for efficient consumer pricing schemes in electricity markets,” *IEEE Trans. Power Syst.*, vol. 22, no. 1, pp. 60–67, 2007.
- [69] P. Yang, G. Tang, and A. Nehorai, “A game-theoretic approach for optimal time-of-use electricity pricing,” *IEEE Trans. Power Syst.*, vol. 28, no. 2, May 2013.
- [70] P. Samadi, A.-H. Mohsenian-Rad, R. Schober, V. W. Wong, and J. Jatskevich, “Optimal real-time pricing algorithm based on utility maximization for smart grid,” in *International Conference on Smart Grid Communications*, Gaithersburg, MD, Oct. 2010, pp. 415–420.
- [71] S. Karnouskos, O. Terzidis, and P. Karnouskos, “An advanced metering infrastructure for future energy networks,” in *NTMS 2007 Conference*, Paris, France, May 2007.
- [72] M. H. Bollen and F. Hassan, *Integration of Distributed Generation in the Power System*, Wiley-IEEE Press, Aug. 2011.
- [73] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Prentice Hall, 1989.
- [74] C. Chen, S. Kishore, and L. V. Snyder, “An innovative RTP-based residential power scheduling scheme for smart grids,” in *ICASSP*, Prague, Czech Republic, May 2011, pp. 5956–5959.
- [75] A.-H. Mohsenian-Rad and A. Leon-Garcia, “Optimal residential load control with price prediction in real-time electricity pricing environments,” *IEEE Trans. Smart Grid*, vol. 1, no. 2, pp. 120–133, Sep. 2010.
- [76] L. Jia, Z. Yu, M. C. Murphy-Hoye, E. G. Piccioli, A. Pratt, and L. Tong, “Multi-scale stochastic optimization for home energy management,” in *IEEE ICASSP*, San Juan, Puerto Rico, Dec. 2011.

- [77] W. Lin, J. Han, R. Xia, and D. He, “A scheme for building demand response based on a comprehensive load profile,” in *North American Power Symposium (NAPS)*, 2012, Sep. 2012, pp. 1–6.
- [78] S. Chan, K. Tsui, H. Wu, Y. Hou, Y.-C. Wu, and F. Wu, “Load/price forecasting and managing demand response for smart grids: Methodologies and challenges,” *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 68–85, Sep. 2012.
- [79] A.-H. Mohsenian-Rad, V. Wong, J. Jatskevich, and R. Schober, “Optimal and autonomous incentive-based energy consumption scheduling algorithm for smart grid,” in *Innovative Smart Grid Technologies (ISGT)*, 2010, Vancouver, Canada, Jan. 2010.
- [80] Q. Zhu, J. Zhang, P. Sauer, A. Dominguez-Garcia, and T. Basar, “A game-theoretic framework for control of distributed renewable-based energy resources in smart grids,” in *American Control Conference (ACC)*, Montreal, Canada, Jun. 2011.
- [81] 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.
- [82] N. Gatsis and G. Giannakis, “Residential load control: Distributed scheduling and convergence with lost ami messages,” *IEEE Trans. Smart Grid*, vol. 3, no. 2, pp. 770–786, Jun. 2012.
- [83] “Vehicle-to-grid power implementation: From stabilizing the grid to supporting large-scale renewable energy,” *Journal of Power Sources*, vol. 144, no. 1, pp. 280–294, 2005.
- [84] Z. Tan, P. Yang, and A. Nehorai, “An optimal and distributed demand response strategy with electric vehicles in the smart grid,” *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 861–869, Mar. 2014.
- [85] J. S. Netz, “Price regulation: a (non-technical) overview,” *Encyclopedia of law and economics*, pp. 1396–465, 2000.
- [86] E. Gilboa, P. S. LaRosa, and A. Nehorai, “Estimating electrical conductivity tensors of biological tissues using microelectrode arrays,” *Annals of Biomedical Engineering*, vol. 40, no. 10, pp. 2140–2155, 2012.
- [87] L. Gan, U. Topcu, and S. Low, “Optimal decentralized protocol for electric vehicle charging,” in *Decision and Control and European Control Conference (CDC-ECC)*, 2011 50th IEEE Conference on, 2011, pp. 5798–5804.

- [88] FERC, “MISO daily report,” 2011, Electric Power Markets: Midwest (MISO). [Online]. Available: <http://www.ferc.gov/market-oversight/mkt-electric/midwest/miso-archives.asp>
- [89] OPA, “Simulated wind generation data,” 2007, Ontario Power Authority. [Online]. Available: <http://www.powerauthority.on.ca/integrated-power-system-plan/simulated-wind-generation-data>
- [90] T. Wiedmann and J. Minx, *A Definition of ‘Carbon Footprint’*. Hauppauge NY, USA.: Nova Science Publishers, 2008.
- [91] J. Carrasco, L. Franquelo, J. Bialasiewicz, E. Galvan, R. Guisado, M. Prats, J. Leon, and N. Moreno-Alfonso, “Power-electronic systems for the grid integration of renewable energy sources: A survey,” *IEEE Trans. Ind. Electron.*, vol. 53, no. 4, pp. 1002–1016, 2006.
- [92] H. Ibrahim, A. Ilinca, and J. Perron, “Energy storage systems – characteristics and comparisons,” *Renewable and Sustainable Energy Reviews*, vol. 12, no. 5, pp. 1221 – 1250, 2008. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1364032107000238>
- [93] J. Garcia-Gonzalez, R. de la Muela, L. Santos, and A. Gonzalez, “Stochastic joint optimization of wind generation and pumped-storage units in an electricity market,” *IEEE Trans. Power Syst.*, vol. 23, no. 2, pp. 460–468, 2008.
- [94] T. D. Nguyen, K.-J. Tseng, S. Zhang, and T. D. Nguyen, “On the modeling and control of a novel flywheel energy storage system,” in *Industrial Electronics (ISIE), 2010 IEEE International Symposium on*, 2010, pp. 1395–1401.
- [95] H. Zhou, T. Bhattacharya, D. Tran, T. Siew, and A. Khambadkone, “Composite energy storage system involving battery and ultracapacitor with dynamic energy management in microgrid applications,” *IEEE Trans. Power Electron.*, vol. 26, no. 3, pp. 923–930, 2011.
- [96] S. G. Chalk and J. F. Miller, “Key challenges and recent progress in batteries, fuel cells, and hydrogen storage for clean energy systems,” *Journal of Power Sources*, vol. 159, no. 1, pp. 73 – 80, 2006.
- [97] J. Barton and D. Infield, “Energy storage and its use with intermittent renewable energy,” *Energy Conversion, IEEE Transactions on*, vol. 19, no. 2, pp. 441–448, 2004.
- [98] K. G. Vosburgh, “Compressed air energy storage,” *Journal of Energy*, vol. 2, no. 2, pp. 106–112, 1978.

- [99] C. Abbey and G. Joos, "Supercapacitor energy storage for wind energy applications," *IEEE Trans. Ind. Appl.*, vol. 43, no. 3, pp. 769–776, 2007.
- [100] P. Brown, J. Peas Lopes, and M. Matos, "Optimization of pumped storage capacity in an isolated power system with large renewable penetration," *IEEE Trans. Power Syst.*, vol. 23, no. 2, pp. 523–531, 2008.
- [101] C. Abbey and G. Joos, "A stochastic optimization approach to rating of energy storage systems in wind-diesel isolated grids," *IEEE Trans. Power Syst.*, vol. 24, no. 1, pp. 418–426, 2009.
- [102] Y. Zhang, N. Gatsis, and G. Giannakis, "Robust energy management for microgrids with high-penetration renewables," *IEEE Trans. Sust. Energy*, vol. PP, no. 99, pp. 1–10, 2013.
- [103] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.
- [104] G. Calafiore and M. Campi, "The scenario approach to robust control design," *IEEE Trans. Autom. Control*, vol. 51, no. 5, pp. 742–753, 2006.
- [105] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, 2009.
- [106] Y. Zhang, N. Gatsis, and G. Giannakis, "Risk-constrained energy management with multiple wind farms," in *Innovative Smart Grid Technologies (ISGT), 2013 IEEE PES*, Feb 2013, pp. 1–6.
- [107] Y. Zhang, N. Gatsis, V. Kekatos, and G. Giannakis, "Risk-aware management of distributed energy resources," in *Digital Signal Processing (DSP), 2013 18th International Conference on*, July 2013, pp. 1–5.
- [108] P. Yang and A. Nehorai, "Hybrid energy storage and generation planning with large renewable penetration," in *IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, Dec. 2013, pp. 1–4.
- [109] *National Solar Radiation Data Base*. [Online]. Available: http://rredc.nrel.gov/solar/old_data/nsrdb/
- [110] S. Wilcox, *National Solar Radiation Database 1991 - 2010 Update: Users Manual*, Aug. 2012.
- [111] EPRI, "Renewable energy technical assessment guide - TAG-RE:2006," EPRI, Palo Alto, Tech. Rep., 2007.

- [112] *ERCOT Hourly Load Data Archive*. [Online]. Available: http://www.ercot.com/gridinfo/load/load_hist/
- [113] EPRI, “Electricity energy storage technology options: A white paper primer on applications, costs, and benefits,” EPRI, Palo Alto, Tech. Rep., 2010.
- [114] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming, version 2.0 beta,” <http://cvxr.com/cvx>, Sep. 2012.
- [115] FERC, “CAISO daily report,” 2011, Electric Power Markets: California (CAISO). [Online]. Available: <http://www.ferc.gov/market-oversight/mkt-electric/california/caiso-archives.asp>
- [116] KEMA, Inc., “Substation communications: Enabler of automation / an assessment of communications technologies,” United Telecom Council, Tech. Rep., Nov. 2006.

Appendix A

Derivation of Condition (3.44)

The proof for the nondecreasing property of $g_A(\mathcal{X})$ is trivial, as $\mathbf{A}_{[y]} \succ \mathbf{A}_{[x]}$ indicates that $\mathbf{A}_{[y]}^{-1} \prec \mathbf{A}_{[x]}^{-1}$. Using the fact that the trace of a matrix equals the sum of its eigenvalues, we conclude that $g_A(\mathcal{X})$ is nondecreasing.

Let $\mathbf{A}_{[x]} = \mathcal{I}_{ss}(\mathcal{X})$, $\mathbf{A}_{[y]} = \mathcal{I}_{ss}(\mathcal{Y})$. Since $\mathcal{X} \subseteq \mathcal{Y}$, we can then write

$$\mathbf{A}_{[y]} = \mathbf{A}_{[x]} + \mathbf{C}_{[xy]} \quad (\text{A.1})$$

where

$$\mathbf{C}_{[xy]} = \sum_{n \in \mathcal{Y} \setminus \mathcal{X}} \mathbf{A}_n. \quad (\text{A.2})$$

The function $g_A(\mathcal{X})$ is constructed such that $g_A(\emptyset) = 0$. To show it is submodular, it suffices to show

$$g_A(\mathcal{X} \cup \{i\}) - g_A(\mathcal{X}) \geq g_A(\mathcal{Y} \cup \{i\}) - g_A(\mathcal{Y}) \quad (\text{A.3})$$

for all $\mathcal{X} \subseteq \mathcal{Y}$ and $i \notin \mathcal{Y}$ [A1]. Using the fact that if \mathbf{A}^{-1} and $(\mathbf{A} + \mathbf{B})^{-1}$ exist, we obtain from the matrix inversion lemma that

$$(\mathbf{A} + \mathbf{B})^{-1} = \mathbf{A}^{-1} - (\mathbf{I} + \mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}. \quad (\text{A.4})$$

Therefore, we obtain

$$\begin{aligned}
g_A(\mathcal{X} \cup \{i\}) - g_A(\mathcal{X}) &= -\text{tr}(\mathcal{I}_{ss}^{-1}(\mathcal{X} \cup \{i\})) + \text{tr}(\mathcal{I}_{ss}^{-1}(\mathcal{X})) \\
&= -\text{tr}\left((\mathbf{A}_{[x]} + \mathbf{A}_i)^{-1}\right) + \text{tr}\left((\mathbf{A}_{[x]})^{-1}\right) \\
&= \text{tr}\left(\left(\mathbf{I} + \mathbf{A}_{[x]}^{-1}\mathbf{A}_i\right)^{-1}\mathbf{A}_{[x]}^{-1}\mathbf{A}_i\mathbf{A}_{[x]}^{-1}\right) \\
&= \text{tr}\left(\mathbf{A}_{[x]}^{-1}(\mathbf{A}_{[x]} + \mathbf{A}_i)^{-1}\mathbf{A}_i\right).
\end{aligned} \tag{A.5}$$

Similarly, we have that

$$\begin{aligned}
g_A(\mathcal{Y} \cup \{i\}) - g_A(\mathcal{Y}) &= \text{tr}\left(\mathbf{A}_{[y]}^{-1}(\mathbf{A}_{[y]} + \mathbf{A}_i)^{-1}\mathbf{A}_i\right).
\end{aligned} \tag{A.6}$$

Therefore the condition (A.3) is equivalent to (3.44).

References

[A1] A. Schrijver, *Combinatorial Optimization*. Springer, 2003.

Appendix B

Proof of Submodularity

Proof. Let $\mathbf{A}_{[x]} = \mathcal{I}_{ss}(\mathcal{X})$, $\mathbf{A}_{[y]} = \mathcal{I}_{ss}(\mathcal{Y})$. Since $\mathcal{X} \subseteq \mathcal{Y}$, we can then write $\mathbf{A}_{[y]} = \mathbf{A}_{[x]} + \mathbf{C}_{[xy]}$, where $\mathbf{C}_{[xy]} = \sum_{n \in \mathcal{Y} \setminus \mathcal{X}} \mathbf{A}_n$. We already have that $g_D(\emptyset) = 0$, and need to show $g_D(\mathcal{X})$ satisfies

$$g_D(\mathcal{X} \cup \{i\}) - g_D(\mathcal{X}) \geq g_D(\mathcal{Y} \cup \{i\}) - g_D(\mathcal{Y}). \quad (\text{B.1})$$

Since \mathbf{A}_i is positive semidefinite, we can decompose it as

$$\mathbf{A}_i = \mathbf{V}_i \mathbf{\Lambda}_i \mathbf{V}_i^\top = (\mathbf{V}_i \mathbf{\Lambda}_i^{\frac{1}{2}})(\mathbf{V}_i \mathbf{\Lambda}_i^{\frac{1}{2}})^\top = \mathbf{U}_i \mathbf{U}_i^\top. \quad (\text{B.2})$$

Using the matrix determinant lemma [B1], we obtain that

$$\begin{aligned} g_D(\mathcal{X} \cup \{i\}) - g_D(\mathcal{X}) &= \log \det (\mathbf{A}_{[x]} + \mathbf{A}_i) - \log \det (\mathbf{A}_{[x]}) \\ &= \log \det \left(\mathbf{I} + \mathbf{U}_i^\top \mathbf{A}_{[x]}^{-1} \mathbf{U}_i \right) \end{aligned} \quad (\text{B.3})$$

$$\begin{aligned} g_D(\mathcal{Y} \cup \{i\}) - g_D(\mathcal{Y}) &= \log \det (\mathbf{A}_{[y]} + \mathbf{A}_i) - \log \det (\mathbf{A}_{[y]}) \\ &= \log \det \left(\mathbf{I} + \mathbf{U}_i^\top \mathbf{A}_{[y]}^{-1} \mathbf{U}_i \right). \end{aligned} \quad (\text{B.4})$$

Decompose $\mathbf{C}_{[\text{xy}]} = \mathbf{U}_{[\text{xy}]} \mathbf{U}_{[\text{xy}]}^\top$. Then, by following the Woodbury matrix identity [B2], we obtain

$$\begin{aligned}
& \mathbf{I} + \mathbf{U}_i^\top \mathbf{A}_{[\text{y}]}^{-1} \mathbf{U}_i \\
&= \mathbf{I} + \mathbf{U}_i^\top (\mathbf{A}_{[\text{x}]} + \mathbf{C}_{[\text{xy}]})^{-1} \mathbf{U}_i \\
&= \mathbf{I} + \mathbf{U}_i^\top \left(\mathbf{A}_{[\text{x}]}^{-1} - \mathbf{A}_{[\text{x}]}^{-1} \mathbf{U}_{[\text{xy}]} \left(\mathbf{I} + \mathbf{U}_{[\text{xy}]}^\top \mathbf{A}_{[\text{x}]}^{-1} \mathbf{U}_{[\text{xy}]} \right)^{-1} \right. \\
&\quad \left. \mathbf{U}_{[\text{xy}]}^\top \mathbf{A}_{[\text{x}]}^{-1} \right) \mathbf{U}_i \\
&= \mathbf{I} + \mathbf{U}_i^\top \mathbf{A}_{[\text{x}]}^{-1} \mathbf{U}_i \\
&\quad - \underbrace{\mathbf{U}_i^\top \mathbf{A}_{[\text{x}]}^{-1} \mathbf{U}_{[\text{xy}]} \left(\mathbf{I} + \mathbf{U}_{[\text{xy}]}^\top \mathbf{A}_{[\text{x}]}^{-1} \mathbf{U}_{[\text{xy}]} \right)^{-1} \mathbf{U}_{[\text{xy}]}^\top \mathbf{A}_{[\text{x}]}^{-1} \mathbf{U}_i}_{\mathbf{W}_{[\text{xy}]}}.
\end{aligned}$$

It is easy to verify that $\mathbf{W}_{[\text{xy}]}$ is positive semi-definite. From Corollary 2 in [B3], we then obtain that

$$\det \left(\mathbf{I} + \mathbf{U}_i^\top \mathbf{A}_{[\text{x}]}^{-1} \mathbf{U}_i \right) \geq \det \left(\mathbf{I} + \mathbf{U}_i^\top \mathbf{A}_{[\text{y}]}^{-1} \mathbf{U}_i \right) \geq 1, \quad (\text{B.5})$$

and therefore $g_{\text{D}}(\mathcal{X})$ is submodular and non-decreasing. \square

References

- [B1] D. A. Harville, *Matrix Algebra From a Statistician's Perspective*. Springer, 1997.
- [B2] M. A. Woodbury, “Inverting modified matrices,” *Statistical Research Group, Memo. Rep.*, no. 42, 1950.
- [B3] M. Gowda and J. Tao, “Some inequalities involving determinants, eigenvalues, and Schur complements in Euclidean Jordan algebras,” *Positivity*, vol. 15, no. 3, pp. 381–399, 2011.

Appendix C

Sketch of Convergence Proof

In this appendix we sketch a preliminary proof of convergence for our step-size regularized parallel optimization algorithm when λ is sufficiently large.

Proof. We use simplified notation in this proof. Let $f_i(\mathbf{x}_i, \mathbf{x}_{-i}^k)$ denote the cost function of the i th user at iteration k , where \mathbf{x}_i denotes the load of the i th user, and \mathbf{x}_{-i} denotes the load of other users. We use the superscript k to denote the k th iteration. The update can be written as

$$\mathbf{x}_i^{k+1} = \underset{\mathbf{x}_i \in \mathcal{X}_i}{\operatorname{argmin}} f_i(\mathbf{x}_i, \mathbf{x}_{-i}^k) + \lambda_i \|\mathbf{x}_i - \mathbf{x}_i^k\|_2^2. \quad (\text{C.1})$$

In this proof we use the following update as an approximation:

$$\mathbf{x}_i^{k+1} = T_i(\mathbf{x}_i^k) = \left[\mathbf{x}_i^k - \frac{1}{2\lambda_i} \nabla_i f_i(\mathbf{x}_i^k, \mathbf{x}_{-i}^k) \right]_{\mathcal{X}_i}, \quad (\text{C.2})$$

where $[\cdot]_{\mathcal{X}_i}$ denotes the projection onto the subspace \mathcal{X}_i defined by the constraints. From the projection theorem [C1], we obtain that

$$(\mathbf{x}_i^k - T_i(\mathbf{x}_i^k))^T \left(\mathbf{x}_i^k - \frac{1}{2\lambda_i} \nabla f_i(\mathbf{x}_i^k, \mathbf{x}_{-i}^k) - T_i(\mathbf{x}_i^k) \right) \leq 0 \quad (\text{C.3})$$

$$\Rightarrow \frac{1}{2\lambda_i} (T_i(\mathbf{x}_i^k) - \mathbf{x}_i^k)^T \nabla f_i(\mathbf{x}_i^k, \mathbf{x}_{-i}^k) \leq -\|T_i(\mathbf{x}_i^k) - \mathbf{x}_i^k\|_2^2. \quad (\text{C.4})$$

Assume that $\nabla f_i(\mathbf{x}_i, \mathbf{x}_{-i})$ is Lipschitz continuous, with

$$\|\nabla f_i(\mathbf{x}_i, \mathbf{x}_{-i}) - \nabla f_i(\mathbf{y}_i, \mathbf{x}_{-i})\| \leq K_1 \|\mathbf{x}_i - \mathbf{y}_i\|, \quad (\text{C.5})$$

Then from the descent lemma [C1],

$$\begin{aligned}
f_i(\mathbf{x}_i^{k+1}, \mathbf{x}_{-i}^{k+1}) &\leq f_i(\mathbf{x}_i^k, \mathbf{x}_{-i}^{k+1}) \\
&\quad + (T_i(\mathbf{x}_i^k) - \mathbf{x}_i^k)^T \nabla f_i(\mathbf{x}_i^k, \mathbf{x}_{-i}^{k+1}) \\
&\quad + \frac{K_1}{2} \|T_i(\mathbf{x}_i^k) - \mathbf{x}_i^k\|_2^2.
\end{aligned} \tag{C.6}$$

Assume that $\nabla f_{-i}(\mathbf{x}_i, \mathbf{x}_{-i}^k)$ is also Lipschitz continuous, with

$$\|\nabla f_{-i}(\mathbf{x}_i, \mathbf{y}_{-i}) - \nabla f_{-i}(\mathbf{x}_i, \mathbf{y}_{-i}')\| \leq K_2 \|\mathbf{y}_{-i} - \mathbf{y}_{-i}'\|, \tag{C.7}$$

and we can then further decompose the first term on the right side of (C.6),

$$f_i(\mathbf{x}_i^k, \mathbf{x}_{-i}^{k+1}) \leq f_i(\mathbf{x}_i^k, \mathbf{x}_{-i}^k) + \mathbf{d}_{-i}^k{}^T \nabla f_{-i}(\mathbf{x}_i^k, \mathbf{x}_{-i}^k) + \frac{K_2}{2} \|\mathbf{d}_{-i}^k\|_2^2, \tag{C.8}$$

where $\mathbf{d}_{-i}^k = \mathbf{x}_{-i}^{k+1} - \mathbf{x}_{-i}^k$. Using the projection theorem,

$$\begin{aligned}
&(\mathbf{x}_i^k - T_i(\mathbf{x}_i^k))^T \left(\mathbf{x}_i^k - \frac{1}{2\lambda_i} \nabla f_i(\mathbf{x}_i^k, \mathbf{x}_{-i}^k) + \right. \\
&\quad \left. \frac{1}{2\lambda_i} \nabla f_i(\mathbf{x}_i^k, \mathbf{x}_{-i}^{k+1}) - \frac{1}{2\lambda_i} \nabla f_i(\mathbf{x}_i^k, \mathbf{x}_{-i}^{k+1}) - T_i(\mathbf{x}_i^k) \right) \leq 0
\end{aligned} \tag{C.9}$$

$$\begin{aligned}
\Rightarrow \frac{1}{2\lambda_i} (T_i(\mathbf{x}_i^k) - \mathbf{x}_i^k)^T \nabla f_i(\mathbf{x}_i^k, \mathbf{x}_{-i}^{k+1}) &\leq -\|T_i(\mathbf{x}_i^k) - \mathbf{x}_i^k\|_2^2 \\
&\quad - \frac{1}{2\lambda_i} (\mathbf{x}_i^k - T_i(\mathbf{x}_i^k))^T (\nabla f_i(\mathbf{x}_i^k, \mathbf{x}_{-i}^{k+1}) - \nabla f_i(\mathbf{x}_i^k, \mathbf{x}_{-i}^k)).
\end{aligned} \tag{C.10}$$

Therefore we can expand (C.6) as

$$\begin{aligned}
f_i(\mathbf{x}_i^{k+1}, \mathbf{x}_{-i}^{k+1}) &\leq f_i(\mathbf{x}_i^k, \mathbf{x}_{-i}^k) \\
&\quad - 2\lambda_i \|T_i(\mathbf{x}_i^k) - \mathbf{x}_i^k\|_2^2 + \frac{K_1}{2} \|T_i(\mathbf{x}_i^k) - \mathbf{x}_i^k\|_2^2 \\
&\quad + \frac{K_2}{2} \|\mathbf{d}_{-i}^k\|_2^2 + \mathbf{d}_{-i}^k{}^T \nabla f_{-i}(\mathbf{x}_i^k, \mathbf{x}_{-i}^k) \\
&\quad - (\mathbf{x}_i^k - T_i(\mathbf{x}_i^k))^T (\nabla f_i(\mathbf{x}_i^k, \mathbf{x}_{-i}^{k+1}) - \nabla f_i(\mathbf{x}_i^k, \mathbf{x}_{-i}^k)).
\end{aligned} \tag{C.11}$$

For a large enough λ_i , the second term on the right side of (C.11) would dominate the third and fourth term on the same line. To show that $f_i(\mathbf{x}_i^{k+1}, \mathbf{x}_{-i}^{k+1}) \leq f_i(\mathbf{x}_i^k, \mathbf{x}_{-i}^k)$, we need to

further study the last two terms of (C.11), which depend on the function $f_i(\mathbf{x}_i, \mathbf{x}_{-i})$. For a specific cost generation cost function $C(g)$, which takes the quadratic form

$$C(g) = c_1 g + c_2 g^2, \quad (\text{C.12})$$

we can further simplify the last two terms of (C.11) as

$$(c_2 \mathbf{x}_i^{k+1})^T \sum_{j \neq i} (\mathbf{x}_j^{k+1} - \mathbf{x}_j^k). \quad (\text{C.13})$$

To analyze this term, we need to consider the assumption that the total user load does not change between iterations (in practice it may change slightly due to satisfaction cost, but we ignore the change here for simplicity of proof). In this case, we have that

$$\mathbf{1}^T \sum_{j \neq i} (\mathbf{x}_j^{k+1} - \mathbf{x}_j^k) = 0, \quad (\text{C.14})$$

where $\mathbf{1}$ denotes an all-one vector. Intuitively, when $\sum_{j \neq i} (x_j^{k+1}(t) - x_j^k(t)) > 0$, the users will shift loads to this time t because the price at this time is low, and thus the load at this time is also lower than average compared with other times. On the contrary, when $\sum_{j \neq i} (x_j^{k+1}(t) - x_j^k(t)) < 0$, the price is high at t , which means the load is higher than average at this time. Due to the condition (C.14), the sum of the change should equal zero. And recall that when λ_i is chosen to be big enough, the stepsize is small. Therefore the term (C.11) should be negative. To this point, we show that $f_i(\mathbf{x}_i^{k+1}, \mathbf{x}_{-i}^{k+1}) < f_i(\mathbf{x}_i^k, \mathbf{x}_{-i}^k)$, i.e., $f_i(\mathbf{x}_i, \mathbf{x}_{-i})$ monotonically decreases. Since it is also bounded below on a compact set, the algorithm converges. Note that the stepsize decreases as λ_i increases, and therefore the convergence will be slower with large λ_i . \square

The rigorous proof for the general case where $f(\mathbf{x}_i, \mathbf{x}_{-i})$ is an arbitrary convex function satisfying certain conditions is more mathematically involved, and will be an interesting topic to explore in our future work.

References

- [C1] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Prentice Hall, 1989.

Vita

Peng Yang

Degrees

Ph.D., Electrical Engineering, Washington University in St. Louis, Missouri, USA, May 2014

M.S., Electrical Engineering, Washington University in St. Louis, Missouri, USA, August 2011

B.S., Electrical Engineering, University of Science and Technology of China, Anhui, China, June 2009

Professional Societies

The Institute of Electrical and Electronics Engineers (IEEE)

IEEE Signal Processing Society

IEEE Power and Energy Society

Publications

Journal Publications:

P. Yang, Z. Tan, A. Wiesel, and A. Nehorai, “Placement of phasor measurement units considering imperfect time synchronization,” submitted.

Z. Tan, **P. Yang**, and A. Nehorai, “Joint sparse recovery method for compressed sensing with structured dictionary mismatch,” in revision.

P. Yang, and A. Nehorai, “Joint optimization of hybrid energy storage and generation capacity with renewable energy,” to appear in *IEEE Trans. on Smart Grid*.

Z. Tan, **P. Yang**, and A. Nehorai, “An optimal and distributed demand response strategy with electric vehicles in the smart grid,” *IEEE Trans. on Smart Grid*, Vol 5, No. 2, pp. 861-869, Mar. 2014.

P. Chavali, **P. Yang**, and A. Nehorai, “A distributed algorithm of appliance scheduling for home energy management system,” *IEEE Trans. on Smart Grid*, Vol 5, No. 1, pp. 282-290, Jan. 2014.

P. Yang, Z. Tan, A. Wiesel, and A. Nehorai, “Power system state estimation using PMUs with imperfect synchronization,” *IEEE Trans. on Power Systems*, Vol 28, No. 4, pp. 4162-4172, Nov. 2013.

P. Yang, P. Chavali, E. Gilboa, and A. Nehorai, “Parallel load schedule optimization with renewable distributed generators in smart grids,” *IEEE Trans. on Smart Grid*, Vol 4, No. 3, pp. 1431-1441, Sep. 2013.

P. Yang, G. Tang, and A. Nehorai, “A game-theoretic approach for optimal time-of-use electricity pricing,” *IEEE Trans. on Power Systems*, Vol. 28, No. 2, pp. 884-892, May 2013.

Conference Publications:

P. Yang, and A. Nehorai, “Hybrid energy storage and generation planning with large renewable penetration,” in Proc. *5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing(CAMSAP)*, Saint Martin, Dec. 2013.

Z. Tan, **P. Yang**, and A. Nehorai, “Distributed demand response for plug-in electrical vehicles in the smart grid,” in Proc. *5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing(CAMSAP)*, Saint Martin, Dec. 2013.

Z. Tan, **P. Yang**, and A. Nehorai, “Joint-sparse recovery in compressed sensing with dictionary mismatch,” in Proc. *5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing(CAMSAP)*, Saint Martin, Dec. 2013.

P. Yang, Z. Tan, A. Wiesel, and A. Nehorai, “Performance bounds and sensor placement for state estimation using PMUs with phase mismatch,” in Proc. *IEEE Power and Energy Society (PES) General Meeting*, Vancouver BC, Canada, July, 2013.

Z. Tan, **P. Yang**, and A. Nehorai, “Sparse MIMO radar with phase mismatch” in Proc. *38th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vancouver BC, Canada, May, 2013.

E. Gilboa, P. Chavali, **P. Yang**, and A. Nehorai, “Distributed optimization via adaptive regularization for large problems with separable constraints,” in Proc. *38th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vancouver BC, Canada, May, 2013.

P. Yang, Z. Tan, A. Wiesel, and A. Nehorai, “State estimation with consideration of PMU phase mismatch for smart grids,” in Proc. *IEEE Innovative Smart Grid Technologies Conference (ISGT)*, Washington D.C., Feb, 2013.

P. Yang, P. Chavali, and A. Nehorai, “Parallel autonomous optimization of demand response with renewable distributed generators,” in Proc. *IEEE Smart Grid Communications (SmartGridComm)*, Tainan City, Taiwan, Nov. 2012, pp. 55-60.

W. F. Zhang, **P. Yang**, D. Q. Dai, and A. Nehorai, “Reflectance estimation using local regression methods,” in Proc. *9th International Symposium on Neural Networks (ISNN)*, Shenyang, China, Jul. 2012, pp. 116-122.

P. Yang, G. Tang, and A. Nehorai, “Optimal time-of-use electricity pricing using game theory,” in Proc. *37th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 3081-3084.

P. Yang, G. Tang, and A. Nehorai, “Sparsity enforced regression based on over-complete dictionary,” in Proc. *4th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, San Juan, Puerto Rico, Dec. 2011, pp. 261-264.

May 2014