Optimizing for the Future Smart Grid: Efficient Methods for Nonconvex AC Power Flow Problems

By

Elizabeth Glista

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Mechanical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Assistant Professor Somayeh Sojoudi, Chair
Associate Professor Javad Lavaei
Professor Kameshwar Poolla

Spring 2023

Optimizing for the Future Smart Grid:
Efficient Methods for Nonconvex AC Power Flow Problems

Abstract

Optimizing for the Future Smart Grid:
Efficient Methods for Nonconvex AC Power Flow Problems

by

Elizabeth Glista

Doctor of Philosophy in Engineering – Mechanical Engineering

University of California, Berkeley

Assistant Professor Somayeh Sojoudi, Chair

The increased electrification of society and calls for improvements in electrical grid reliability lend renewed attention to the domain of power systems optimization. For the future smart grid, it will be crucial both to find better solutions to classical power grid optimization problems such as optimal power flow (OPF) and state estimation (SE) as well as to develop new optimization methods for the updated physical systems. Due to the nonlinearity of alternating current (AC), many optimization problems in power systems are nonlinear and nonconvex and thus are hard to solve with existing fast convex methods. In this dissertation, we propose some efficient methods related to the OPF, post-contingency OPF, SE, and power flow (PF) problems. These methods provide new optimization and machine learning frameworks for ensuring the reliable and efficient operation of the power grid.

The OPF problem is an important tool for power systems operation that aims to minimize the cost of electric power generation subject to consumer demand, the physics of power flow, and technological constraints. OPF is a nonconvex problem that is known to have many local minima in realistic test cases. Inspired by the frequent use of local search methods to solve OPF, which may result in sub-optimal solutions and provide no guarantee on the quality of the solution, we propose a convex method to bound the worst-case local minimum that can be obtained by a local solver. In order to formulate the problem of finding the worst-case local minimum of OPF, we start with a version of OPF modeled as a quadratically-constrained quadratic program (QCQP) and characterize the local minima of the QCQP using first- and second-order necessary optimality conditions. Because this derived worst-case local minimum problem is also nonconvex, we implement a convex relaxation of the problem into a semidefinite program (SDP) and show that it is exact for certain cases. Using some test cases which are known to have multiple local minima, we demonstrate the effectiveness of the proposed relaxation to bound the worst-case local minimum. We compare the obtained upper bound on local minima to the lower bound provided by the standard SDP relaxation

of the OPF problem to understand how much SDP outperforms local search for a given problem.

In anticipation of component failures, such as transmission line or generator outages, it is also important to incorporate contingency scenarios into OPF, yielding the security-constrained OPF (SCOPF) problem and the related post-contingency OPF problem. However, finding an optimal, or even feasible, post-contingency solution to the OPF problem is challenging due to the nonconvexity of the power flow equations and the large number of possible contingency cases in practice. With the goal of finding a global solution to the post-contingency OPF problem of a stressed network, e.g. a network with a line or generator outage, we apply a new homotopy method to the problem. By parametrizing the constraint set, we define a series of optimization problems to represent a gradual outage and iteratively solve these problems using local search. Using this homotopy method, we can find solutions to stressed problems that did not initially converge or had a suboptimal cost. Under the condition that the global minimum of the OPF problem for the base case is attainable, we find theoretical guarantees to ensure that the post-contingency OPF problem will also converge to its global minimum. With simulations on Polish and other European networks, we demonstrate that the effectiveness of the proposed homotopy method is dependent on the choice of the homotopy path and that homotopy yields an improved solution in many cases. For at least 5% of the test cases, bad local minima were identified, and the homotopy method yielded a solution that was significantly better than state-of-the-art interior point methods in terms of reducing the violation cost during a catastrophic contingency scenario.

After considering the classical OPF and post-contingency OPF problems, we focus on new SE- and PF-related problems that aim to make use of the large amounts of data that are increasingly available for power networks. The reliability of the electric power grid is linked to the reliability of measured data which is used to understand the current state of the system. Determining the current state of the electric grid is the basis for decision-making related both to the normal operation of the grid as well as to operations in the case of an emergency scenario. When some of this data is corrupted in the case of a cyberattack, it is important that we can recover the true state of the system via SE. Inspired by recent works in two-stage power systems SE, we propose a novel method using a notion in machine learning to optimize the choice of measurements in a given power network, formulating the problem as a mixed-integer linear program (MILP). Using this MILP, we study some test cases and show that it is impossible to certify that the network is fully robust in the case of bad data. However, we propose a method to optimally place the sensors in order to make the network more robust in the case of cyberattacks.

Understanding an electric power system's topology, including both its nodal connectivity and physical parameters, is critically important to the reliable operation and control of the power grid. In cases where this power system topology may be unavailable, due to data collection deficiencies, real-time line switching, or intentional cyberattacks, it is important to be able to estimate the real power system topology with high accuracy. We propose a new data-driven

constrained support vector regression (SVR) method that aims to map voltage data collected from phasor measurement units (PMUs) to data collected by Supervisory Data Acquisition and Control (SCADA) systems. We show that the dual of the constrained SVR model can be formulated as a quadratic program (QP) and solved efficiently with off-the-shelf solvers. Testing our method on standard IEEE test cases, we demonstrate that our proposed method significantly outperforms existing state-of-the-art SVR methods in learning the true network topology, even in the presence of measurement noise, outliers, and missing data.

Contents

List of Figures

List of Tables

Acknowledgments

I am very grateful to UC Berkeley for the fruitful years of learning and research during my doctoral degree. Above all, I would like to thank my advisor, Prof. Somayeh Sojoudi, for supporting my research and providing guidance along the winding path to this dissertation. Thank you for encouraging me to explore various topics in optimization and power systems and for fostering a sense of academic freedom in my pursuits. Thank you for supporting my ideas and providing needed criticism of their deficiencies.

I would also like to thank Prof. Javad Lavaei for his guidance and mentorship in our collaborative project on contingency-OPF. Along with Prof. Sojoudi, Prof. Lavaei was also instrumental in sparking my deep interest in the power systems domain. For that, I am incredibly grateful.

Furthermore, I am thankful for the opportunity to spend two summers at Lawrence Livermore National Laboratory (LLNL) working under the mentorship of Jean-Paul Watson. Thank you for providing a glimpse of national laboratory work in the power systems area and for supporting my collaboration on the EGRET software tool.

Thank you to all the professors at Berkeley and MIT who have encouraged my learning over the years. The Mechanical Engineering and Statistics professors at MIT led me on the path towards advanced study in controls and optimization, and my Berkeley professors provided new insights into the optimization, machine learning, and power systems domains. I would also like to explicitly thank Prof. Kameshwar Poolla for serving on my dissertation committee.

The work in this dissertation benefited from collaborations with SangWoo Park, a graduate of Prof. Lavaei's research group, and Salar Fattahi and Ming Jin, graduates of both Prof. Sojoudi's and Prof. Lavaei's research groups. Thank you to the rest of the current and graduated members of Prof. Sojoudi's and Prof. Lavaei's groups, including Brendon Anderson, Cedric Josz, Richard Y. Zhang, Nan Tian, Ziye Ma, Tanmay Gautam, Jingqi Li, Yatong Bai, Alec Zhou, Samuel Pfrommer, Eli Brock, and Lucas Hyunin Lee, for the interesting group discussions and cross-pollination of ideas.

Lastly, I would like to thank my family and friends for their incredible support. Thank you for encouraging me to pursue my scattered passions and for keeping me balanced in my graduate studies.

Chapter 1

Background & Motivation

The power grid in the continental U.S. connects 145 million customers to over 7,300 power plants with around 160,000 miles of high-voltage power lines and millions of low-voltage power lines and transformers [1]. The traditional grid infrastructure follows a centralized, tree-like approach, wherein power is generated at a few large power plants and then transmitted to substations that distribute power to consumers [2]. In the power systems community, the phrase "smart grid" is used to describe a vision of a future grid which integrates an advanced metering infrastructure with renewable energy sources and automated building controls [2, 3]. In this future decentralized grid, consumers actively participate by sending energy from renewable energy sources back into the grid [2]. We are on the path to realize this vision as renewable energy sources have become increasingly important in the operation of the U.S. power system. Renewable energy generation doubled between 2008 and 2018, accounting for 17.6% of electricity generation in 2018 [4]. However, after the recent power outages in California and Texas, which resulted in hundreds of deaths and billions of dollars in property damage [5–7], it appears that the "smart grid" is still a distant concept. In the following sections, we discuss some challenges and opportunities faced by the power systems community in dealing with both legacy and smart grid systems.

1.1 Challenges with Scalability

The enormous scale of this U.S. power system means that it is computationally intractable to solve many critical power flow problems, such as optimal power flow (OPF), over the complete network with existing techniques. For example, a realistic case of OPF has around 10^4 variables and 10^6 constraints and must be solved within a five-minute operating window [8]. Thus, in practice, a large network such as the U.S. power system is divided into smaller regions, but this strategy fails to take into account the global effects of local decisions. Poor local decisions can result in cascading failures in the interconnected network, as was the case for the Northeast blackout of 2003 which affected around 50 million people [9].

To achieve reliability in the U.S. power system, the system is divided into three in-

terconnections — Eastern, Western, and Texas (ERCOT) — which operate primarily as independent power distribution and transmission networks [1]. The operation of the electric system within these interconnections is controlled by independent systems operators (ISOs) and certain utilities companies which function as "balancing authorities" [1]. The continental U.S. has 66 of these balancing authorities, which are responsible for ensuring the balance between power supply and consumer demand over the network [1]. Ideally, the OPF problem would be solved over each of the three U.S. interconnections; however, this is computationally intractable within the required operating window. To solve the problem in practice, each of the 66 balancing authorities solves the OPF problem, which reduces the size of the network for each problem.

1.2 Challenges with Optimality

Due to the nonlinear nature of the physics of alternating current (AC), many fundamental power systems optimization problems such as OPF and state estimation (SE) are nonconvex. The nonconvex nature of these problems means that modern algorithms, which are efficient for solving convex optimization problems to global optimality, may not be able to find the global minimum of these nonconvex problems that may have many local minima.

The current industry practice for solving nonconvex power systems optimization problems within desired time limits is to simplify the AC power flow equations to a linearized (DC) version for the optimization stage and then use heuristics to generate a feasible AC power flow solution [10]. This method leads to an operating point which may not be a global let alone local solution to the original problem. For example, the DC-OPF approximation is only valid if voltage magnitudes are within tight bounds (per unit voltage magnitude is approximately 1) and voltage phase angle differences are small between buses. The DC-OPF approximation ignores important physics in AC power flow, such as reactive power and voltage magnitudes, and may lead to a suboptimal solution. Based on the approximate cost of electric power production, a 1% difference between a local and a global minimum of the OPF problem would equate to between 1 and 5 billion dollars annually in the U.S. [8].

Modern interior point methods have been shown to efficiently solve AC-OPF and other nonconvex PF-related problems such as SE but provide no guarantee on the quality of the solution and may find a local minimum [11–13]. The development of optimization methods with global guarantees, such as the relaxation of a nonconvex program into a convex program, is promising for finding globally optimal solutions to these problems [14]. If one is able to find a solution to a convex relaxation that is feasible for the original problem, it is guaranteed that the point is globally optimal for the original problem. In recent years, relaxations of OPF into semidefinite programs (SDPs) and second-order cone programs (SOCPs) have gained attention [15, 16]. However, solving a convexified problem still requires a high level of computational complexity, and its solution may not be feasible for the original problem, although it provides a lower bound on the original solution. While the solution to the SDP relaxation of OPF has been shown to be "exact," i.e. feasible for the original problem and

achieving global optimality, in certain real-world cases [17–22], these conditions are not met in general for power networks. It has also been shown that several realistic OPF test cases have many local minima due to cyclic network structures, excess real or reactive power, large phase angle differences across lines, or other reasons [23]. Some work has been done to address the non-zero duality gap by adding valid inequalities, by using branch-and-cut approaches, or by implementing Lasserre hierarchies [24–26]. However, none of these convex relaxation methods are able to guarantee the recovery of a feasible solution to the original nonconvex problem for large-scale networks in the desired timeframe. In Chapter 2, we leverage SDP relaxation techniques to study the problem of finding the worst-case local minimum of the OPF problem.[1]

1.3 Impact of Renewable and Distributed Energy Generation

Renewable and distributed energy sources have become increasingly important in the operation of the U.S. power system [4]. Adding renewable sources to the power system creates new problems for finding solutions to classical problems such as OPF that make use of the AC power flow equations. Consumers with solar panels on their homes may add power back into the power network, making the network structure more cyclic and leading to a greater number of local optima in the OPF problem [23]. Replacing thousands of generators with tens of thousands of solar panels increases the scale of classical problems, and dependency on wind power adds uncertainty to the power generation model. As renewable energy sources generate an increasing share of U.S. electrical power, the importance of solving PF-based problems on a mega-scale will grow.

1.4 Reliability in Case of Outages

The security of a power system in case of a contingency, such as a line or generator outage, is a critical problem to ensuring grid reliability, classically formulated as the security-constrained OPF (SCOPF) problem [28, 29]. While solving the SCOPF problem is difficult due to the large number of contingencies, a variety of decomposition and approximation methods have been effectively used on the problem [30–34]. The SCOPF problem solves for the optimal operating point under normal conditions, i.e. a version of the base-OPF problem, provided that some contingencies are possible. The related problem of contingency-OPF aims to find the optimal operating point under the contingency scenario.

In addition to the existing difficulties solving OPF-based problems to global optimality, the contingency-OPF problem faces further complications: a contingency creates an abrupt change in the system, causing a disconnection between the base-OPF and the contingency-OPF. The disconnection suggests that the initial point that is either provided to or obtained

[1] Chapter 2 includes materials from [27] that were previously published.

from the base-OPF may not be desirable for the contingency-OPF and can mislead any local search algorithms to become stuck at poor local solutions. In Chapter 3, we present a homotopy-based method for finding better solutions to contingency-OPF problems.[2]

1.5 Prevalence of Cyberattacks

On December 23, 2015, a coordinated cyberattack hit three Ukrainian regional electric power distribution companies, resulting in outages for approximately 225,000 different customers [38]. It was determined that the outages were due to the hacking and hijack of the Supervisory Control and Data Acquisition (SCADA) systems [38]. The 2015 Ukrainian attack was the first publicly acknowledged grid attack, and the risk of cyberattacks on SCADA systems continues to pose national security and grid reliability concerns [39, 40].

The 2015 Ukrainian attack demonstrated that many of the academic assumptions in the false data injection attack (FDIA) literature regarding knowledge of cyberattackers are plausible, including that attackers are capable of manipulating SCADA measurements and may be aware of the network topology [41]. An FDIA corresponds to modifying stored or transmitted data in the power system and can affect any process that uses measured data, including SE [42, 43]. SE is a fundamental problem in power systems operation and is performed every few minutes to monitor the state of a transmission or distribution network [44]. FDIAs can lead the state estimator to find a perceived state that is wrong, leading to bad decisions for real-time grid operation and control. Even if the attack is detected, part of the system may become unobservable due to the measurement corruption [43]. Due to the interconnected nature of the grid, a local physical attack may have effects that propagate to a larger region, impacting overall system observability. System observability and SE robustness against cyberattacks are directly related to the choice of sensors in a power system [45, 46]. In Chapter 4, we propose a method to optimally place sensors in a power system in order to ensure the robustness of a state estimator in case of cyberattack.[3]

1.6 Abundance of System Data

One component of the smart grid vision is an advanced sensing infrastructure, which includes smart meters and phasor measurement units (PMUs) [3, 48]. With an increase in the number of physical sensors as well as an increase in data sampling rates, the amount of data available and the opportunities for big data methods in the power systems domain continues to grow [49]. Several machine learning (ML) methods have been proposed to exploit these large datasets in order to learn power flow and OPF mappings [50–54]. While some recent papers use ML techniques to effectively map the power flow equations when a network's topology

[2]Chapter 3 includes materials from [35, 36] that were previously published. Some of this material also appears in [37] due to the collaborative nature of this work with SangWoo Park.

[3]Chapter 4 includes materials from [47] that were previously published.

or parameters are uncertain, these ML-based methods fail to account for the fundamental physics of power networks [55, 56]. In Chapter 5, we develop a new data-driven and physics-based method for learning a power system's true topology and parameters that outperforms existing methods.[4]

1.7 Bibliography

[1] "U.S. electric system is made up of interconnections and balancing authorities," *U.S. Energy Information Administration*, July 2016.

[2] H. Farhangi, "The path of the smart grid," *IEEE Power and Energy Magazine*, vol. 8, no. 1, pp. 18–28, 2010.

[3] X. Fang, S. Misra, G. Xue, and D. Yang, "Smart grid — the new and improved power grid: A survey," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 4, pp. 944–980, 2012.

[4] "U.S. renewable electricity generation has doubled since 2008," *U.S. Energy Information Administration*, March 2019.

[5] California Independent System Operator (CAISO), California Public Utilities Commission (CPUC), and California Energy Commission (CEC), "Root cause analysis: Mid-august 2020 extreme heat wave," January 2021.

[6] City of Austin & Travis County, "2021 winter storm Uri after-action review: Findings report," November 2021.

[7] The University of Texas at Ausin Energy Institute, "The timeline and events of the February 2021 Texas electric grid blackouts," July 2021.

[8] M. B. Cain, R. P. O'Neill, and A. Castillo, "History of optimal power flow and formulations," *Federal Energy Regulatory Commision*, December 2012.

[9] U.S.-Canada Power System Outage Task Force, *Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations.* U.S. Department of Energy, 2004.

[10] A. J. Wood, B. F. Wollenberg, and G. B. Sheblé, *Power generation, operation and control.* Wiley, 2014.

[11] G. Torres and V. Quintana, "An interior-point method for nonlinear optimal power flow using voltage rectangular coordinates," *IEEE Transactions on Power Systems*, vol. 13, no. 4, pp. 1211–1218, November 1998.

[4] Chapter 5 includes materials from [57].

[12] R. A. Jabr, "A primal-dual interior-point method to solve the optimal power flow dispatching problem," *Optimization and Engineering*, vol. 4, pp. 309–336, 2003.

[13] E. J. Oliveira, L. W. Oliveira, J. Pereira, L. M. Honório, I. C. Junior, and A. Marcato, "An optimal power flow based on safety barrier interior point method," *International Journal of Electrical Power & Energy Systems*, vol. 64, pp. 977–985, 2015.

[14] F. Zohrizadeh, C. Josz, M. Jin, R. Madani, J. Lavaei, and S. Sojoudi, "A survey on conic relaxations of optimal power flow problem," *European Journal of Operational Research*, vol. 287, no. 2, pp. 391–409, 2020.

[15] X. Bai, H. Wei, K. Fujisawa, and Y. Wang, "Semidefinite programming for optimal power flow problems," *International Journal of Electrical Power & Energy Systems*, vol. 30, no. 6-7, pp. 383–392, July 2008.

[16] R. Jabr, "Radial distribution load flow using conic programming," *IEEE Transactions on Power Systems*, vol. 21, no. 3, pp. 1458 – 1459, August 2006.

[17] S. Sojoudi and J. Lavaei, "Physics of power networks makes hard optimization problems easy to solve," *2012 IEEE Power and Energy Society General Meeting*, pp. 1–8, 2012.

[18] J. Lavaei, D. Tse, and B. Zhang, "Geometry of power flows and optimization in distribution networks," *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 572–583, October 2013.

[19] J. Lavaei, "Zero duality gap for classical OPF problem convexifies fundamental nonlinear power problems," *American Control Conference*, 2011.

[20] J. Lavaei and S. H. Low, "Zero duality gap in optimal power flow problem," *IEEE Transactions on Power Systems*, vol. 27, no. 1, February 2012.

[21] B. Zhang and D. Tse, "Geometry of feasible injection region of power networks," *Allerton Conference*, September 2011.

[22] S. Sojoudi and J. Lavaei, "Exactness of semidefinite relaxations for nonlinear optimization problems with underlying graph structure," *SIAM Journal on Optimization*, vol. 24, no. 4, pp. 1746–1778, 2014.

[23] W. Bukhsh, A. Grothey, K. McKinnon, and P. Trodden, "Local solutions of the optimal power flow problem," *IEEE Transactions on Power Systems*, November 2013.

[24] B. Kocuk, S. S. Dey, and X. A. Sun, "Strong SOCP relaxations for the optimal power flow problem," *Operations Research*, vol. 64, no. 6, pp. 1177–1196, dec 2016.

[25] C. Coffrin, H. Hijazi, and P. Van Hentenryck, "Strengthening the SDP relaxation of AC power flows with convex envelopes, bound tightening, and valid inequalities," *IEEE Transactions on Power Systems*, vol. 32, no. 5, September 2017.

[26] C. Josz, M. Jin, R. Madani, S. Sojoudi, and J. Lavaei, "Applications of conic relaxation in power system optimization: Theory, algorithms, and case studies," April 2019.

[27] E. Glista and S. Sojoudi, "Convex model to evaluate worst-case performance of local search in the optimal power flow problem," in *2020 59th IEEE Conference on Decision and Control (CDC)*, 2020, pp. 1691–1697.

[28] O. Alsac and B. Stott, "Optimal load flow with steady-state security," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-93, no. 3, pp. 745–751, 1974.

[29] A. Monticelli, M. V. F. Pereira, and S. Granville, "Security-constrained optimal power flow with post-contingency corrective rescheduling," *IEEE Transactions on Power Systems*, vol. 2, no. 1, pp. 175–180, 1987.

[30] J. F. Benders, "Partitioning procedures for solving mixed-variables programming problems," *Numerische Mathematik*, vol. 4, no. 1, pp. 238–252, 1962.

[31] J. Martinez-Crespo, J. Usaola, and J. Fernandez, "Security-constrained optimal generation scheduling in large-scale power systems," *IEEE Transactions on Power Systems*, vol. 21, no. 1, pp. 321–332, 2006.

[32] M. Shahidehopour and Y. Fu, "Benders decomposition: Applying Benders decomposition to power systems," *IEEE Power and Energy Magazine*, vol. 3, no. 2, pp. 20–21, 2005.

[33] F. Capitanescu, M. Glavic, D. Ernst, and L. Wehenkel, "Contingency filtering techniques for preventive security-constrained optimal power flow," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 1690–1697, 2007.

[34] R. Madani, M. Ashraphijuo, and J. Lavaei, "Promises of conic relaxation for contingency-constrained optimal power flow problem," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 1297–1307, 2016.

[35] S. Park, E. Glista, J. Lavaei, and S. Sojoudi, "Homotopy method for finding the global solution of post-contingency optimal power flow," in *2020 American Control Conference (ACC)*, 2020, pp. 3126–3133.

[36] ——, "An efficient homotopy method for solving the post-contingency optimal power flow to global optimality," *IEEE Access*, vol. 10, pp. 124 960–124 978, 2022.

[37] S. Park, "Computational methods for the design and operations of electric power systems: Towards resiliency and security," Ph.D. dissertation, UC Berkeley, 2022.

[38] Electricity Information Sharing and Analysis Center (E-ISAC) and SANS Industrial Control Systems, "Analysis of the cyber attack on the Ukrainian power grid," March 2016.

[39] Y. Xu, "A review of cyber security risks of power systems: From static to dynamic false data attacks," *Protection and Control of Modern Power Systems*, vol. 5, no. 1, p. 19, 2020.

[40] M. Alanazi, A. Mahmood, and M. J. M. Chowdhury, "SCADA vulnerabilities and attacks: A review of the state-of-the-art and open issues," *Computers & Security*, vol. 125, p. 103028, 2023.

[41] G. Liang, S. R. Weller, J. Zhao, F. Luo, and Z. Y. Dong, "The 2015 Ukraine blackout: Implications for false data injection attacks," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3317–3318, 2017.

[42] G. Liang, J. Zhao, F. Luo, S. R. Weller, and Z. Y. Dong, "A review of false data injection attacks against modern power systems," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1630–1638, 2017.

[43] M. Jin, J. Lavaei, and K. H. Johansson, "Power grid AC-based state estimation: Vulnerability analysis against cyber attacks," *IEEE Transactions on Automatic Control*, vol. 64, no. 5, pp. 1784–1799, 2019.

[44] F. F. Wu, "Power system state estimation: A survey," *International Journal of Electrical Power & Energy Systems*, vol. 12, no. 2, pp. 80–87, 1990.

[45] B. Xu and A. Abur, "Observability analysis and measurement placement for systems with PMUs," in *IEEE PES Power Systems Conference and Exposition, 2004.*, 2004, pp. 943–946 vol.2.

[46] M. Jin, J. Lavaei, S. Sojoudi, and R. Baldick, "Boundary defense against cyber threat for power system state estimation," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 1752–1767, 2021.

[47] E. Glista and S. Sojoudi, "A MILP for optimal measurement choice in robust power grid state estimation," in *2022 IEEE Power & Energy Society General Meeting (PESGM)*, 2022, pp. 1–5.

[48] U.S. Department of Energy, "Grid modernization and the smart grid." [Online]. Available: https://www.energy.gov/oe/grid-modernization-and-smart-grid

[49] Y. Zhang, T. Huang, and E. F. Bompard, "Big data analytics in smart grids: A review," *Energy Informatics*, vol. 1, no. 1, p. 8, 2018.

[50] X. Hu, H. Hu, S. Verma, and Z.-L. Zhang, "Physics-guided deep neural networks for power flow analysis," *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 2082–2092, 2021.

[51] S. B. Efe and M. Cebeci, "Power flow analysis by artificial neural network," *International Journal of Energy and Power Engineering*, vol. 2, p. 204, 2013.

[52] L. H. Hassan, M. Moghavvemi, H. A. Almurib, and O. Steinmayer, "Current state of neural networks applications in power system monitoring and control," *International Journal of Electrical Power & Energy Systems*, vol. 51, pp. 134–144, 2013.

[53] L. Imen, L. Djamel, S. Hassiba, D. Abdellah, and F. Selwa, "Optimal power flow study using conventional and neural networks methods," in *2015 International Conference on Renewable Energy Research and Applications (ICRERA)*, 2015, pp. 1422–1427.

[54] K. Baker, "Emulating AC OPF solvers with neural networks," *IEEE Transactions on Power Systems*, vol. 37, no. 6, pp. 4950–4953, 2022.

[55] J. Yu, Y. Weng, and R. Rajagopal, "Robust mapping rule estimation for power flow analysis in distribution grids," in *2017 North American Power Symposium (NAPS)*, 2017, pp. 1–6.

[56] J. Yuan and Y. Weng, "Support matrix regression for learning power flow in distribution grid with unobservability," *IEEE Transactions on Power Systems*, vol. 37, no. 2, pp. 1151–1161, 2022.

[57] E. Glista and S. Sojoudi, "Leveraging the physics of AC power flow in support vector regression to identify power system topology," in *submitted to the 2023 Conference on Decision and Control (CDC)*, 2023.

Chapter 2

The Worst-case Local Minimum of Optimal Power Flow

In this chapter, we present a convex model to evaluate the worst-case performance of local search in the optimal power flow problem.[1]

2.1 Introduction

Optimal power flow (OPF) is a fundamental problem in power systems analysis that is solved approximately every five minutes to find the steady-state operating point of power flow over a network [2, 3]. The goal of this problem is to minimize the cost of electric power generation subject to consumer demand, the physical constraints of power flow, and technological bounds on voltage magnitude, power generation, and transmission line flow. The nonlinearity of alternating current (AC) and the bounds on voltage magnitude create nonconvexity in the feasible region of the OPF problem [4, 5].

The nonconvex nature of OPF makes it NP-hard to find a globally optimal solution to the problem in general [6, 7]. However, finding a globally optimal solution to OPF via scalable, fast methods remains a critical problem in power systems [3]. OPF is highly important to ensure the reliable and efficient operation of the U.S. power grid. By finding a better solution to the OPF problem over the U.S. power grid, it is likely that we would save billions of dollars each year [3]. Additionally, by finding a more optimal solution to OPF, we could reduce the amount of energy generation needed to satisfy the same amount of consumer demand, thereby reducing dependence on fossil fuels without lowering the standard of living.

The OPF problem is related to a variety of other problems, such as security-constrained optimal power flow (SCOPF) and unit commitment (UC) [8–11]. SCOPF is the version of OPF typically solved in industry practice, as it ensures that the solution to OPF is robust to contingency scenarios, such as transmission line or generator outages [12]. In this chapter, we examine the OPF problem, noting that the results developed here could be applied to any

[1] Chapter 2 includes materials from [1] that were previously published.

of these OPF-based problems. Additionally, since we develop results for the OPF problem formulated as a canonical quadratically-constrained quadratic program (QCQP), it is clear that many of the following results can apply to the broader class of nonconvex problems that can be formulated as QCQPs, which are relevant to many domains including signal processing and financial engineering such as in [13–15].

2.1.1 Existing Methods to Solve OPF

Two categories of well-studied techniques for solving the OPF problem are local search methods and conic relaxations [16, 17]. Because the functions in the OPF problem are typically continuous and twice differentiable, iterative local search methods that make use of second-order derivatives are often implemented [16]. Two classical iterative search methods for solving OPF are the Newton-Raphson and Gauss-Seidel methods, which are variations of gradient descent [16, 18].

Barrier and interior point methods are the most commonly used local search methods for solving OPF, and there exist a variety of algorithms based on these approaches [19–23]. These methods have been shown to efficiently find a solution to the OPF problem if a feasible point is available. More recent work has been done to apply homotopy methods to solving OPF [24, 25]. While these local search methods have been shown to efficiently find solutions to OPF problems for many cases, they cannot be guaranteed to find a globally optimal, let alone feasible, solution. They are also highly dependent on a good choice of the initial point, which can be hard to find in general.

Another area of research concerns optimization methods with global guarantees, typically in the form of conic relaxations of the original OPF problem. The paper [26] first developed the semidefinite programming (SDP) relaxation for the OPF problem in the real domain. In the following years, the SDP relaxation of the OPF problem has gained attention due to several papers which show that the relaxation has a zero duality gap in certain cases [27], implying that the SDP relaxation yields an exact solution to the original problem for many real-world networks [28–31]. Nonetheless, there exist other realistic cases where there is a non-zero duality gap, either due to the physics of the problem or the problem formulation itself [32–34]. In these cases, the SDP relaxation can fail to find a feasible and globally optimal solution to the OPF problem.

Other types of conic relaxations, such as quadratic programming and second-order cone programming relaxations, of the OPF problem have been proposed [35], but these are typically dominated by the SDP relaxation [29]. More recent work has aimed to strengthen existing conic relaxations by adding valid inequalities or using branch-and-cut approaches [36, 37]. However, these methods cannot promise an exact solution to the original problem in general.

2.1.2 Existence of Local Optima

In practice, OPF solutions have been shown to be unique, i.e. there is typically only one globally optimal solution to the problem [4]. Due to the nonconvex nature of the OPF problem, there also may be many local optima. The possible existence of local optima in the OPF problem is due to the intersection of nonlinear constraints from the physics of power flow and the rest of the convex constraints. If the intersection of constraints results in a disconnected feasible region for the problem, local optima will exist. Since each disconnected feasible region has an optimal point for that region, there will be at least as many local optima as there are disconnected feasible regions.

It has been shown that several realistic OPF test cases have many local minima [4]. Typical networks with many local minima include cyclic networks with losses or networks with large phase angle differences. A few possible causes of local minima are:

- **Loop flow with losses**: Since voltage phases must sum to zero around a loop but any increment of 2π (i.e. $2\pi k, k = 0, 1, 2, ...$) is valid, there are multiple feasible, disconnected operating regions. The circulating flow results in extra line losses, yielding local optima in the problem.

- **Reduced loads or excess reactive power**: Reducing the loads in the network leads to excess reactive power since loads absorb reactive power. If a line has phase angle difference θ_{ij}, the reactive power absorbed and real power lost in the line are both proportional to $1 - \cos(\theta_{ij})$, so increasing the reactive power absorbed in a line would increase the real power lost when θ_{ij} is held constant. When there is excess reactive power in the network, there is an advantage to increasing the reactive power absorbed in the network by choosing a high value for $1 - \cos(\theta_{ij})$. This expression takes a high value at either small or large θ_{ij}, which leads to local optima.

- **Excess real power**: When there is excess real power in the network, there is also an advantage to increasing $1 - \cos(\theta_{ij})$, which increases the real power lost in a line. This leads to local optima as described for excess reactive power. Excess real power in the network can yield negative real power marginal prices, which are a documented case of problems with local optima for which the SDP method fails [33].

- **Large phase angle differences**: When the voltage phase angle differences are too large across lines, the feasible region may be widened such that it includes discontinuities, resulting in a nonconvex feasible region.

While conic relaxation methods may find the global solution to many cases with local minima, they fail in some of these cases, such as when there is excess real power in the network or when the system is under stress [33, 38]. Additionally, there is no efficient way to characterize all the local minima of the OPF problem. (Note that if there were, the original problem would be easy to solve to global optimality.) To find a collection of local minima, Monte-Carlo simulations with a randomization of the initial point used in the local search

method could be implemented. With this procedure, one may find many local minima but cannot guarantee that all the local minima of the problem have been detected.

2.1.3 Contributions

In this chapter, we propose a new method to characterize the worst-case local minimum of the OPF problem. This allows us to quantify the quality of the solution obtained from a local search method, independent of the initial point. In order to find the worst-case local minimum of the OPF problem, we formulate a new maximization problem based on the first-order and second-order optimality conditions applied to a QCQP model of the OPF. Since this problem is also nonconvex, we propose an SDP relaxation to find an upper bound on the worst-case local minimum of the original OPF problem. We prove that this SDP relaxation is exact in a special case. While this relaxation is not tight in general, we show that high-quality solutions can be obtained with the introduction of a penalty term. We test the proposed SDP relaxation of the worst-case local minimum on benchmark networks and compare its solution to known local minima for these networks. By comparing this upper bound on the unknown local minima with the lower bound on the global minimum obtained from the SDP relaxation of the original problem, one can bound the range of solutions obtained from a local search method. We interpret this distance as a measure of the hardness of the problem, i.e. an estimate of how far apart the global minimum is from the worst-case local minimum, and thus as a measure for the usefulness of convex relaxation techniques to improve solution quality.

2.1.4 Notations

The symbols $\mathbb{R}$ and $\mathbb{C}$ denote the sets of real and complex numbers, respectively. $\mathbb{R}^N$ and $\mathbb{C}^N$ denote the spaces of N-dimensional real and complex vectors, respectively. The symbol $\mathbb{S}^N$ denotes the space of $N \times N$ symmetric real matrices. The symbol $\mathbf{0}^N$ denotes an N-dimensional vector of zeros. The symbols $(\cdot)^T$ and $(\cdot)^*$ denote the transpose and conjugate transpose of a vector or matrix. $\mathrm{Re}\{\cdot\}$ and $\mathrm{Im}\{\cdot\}$ denote the real and imaginary part of a given scalar or matrix. The symbol $|\cdot|$ is the absolute value operator if the argument is a scalar, vector, or matrix; otherwise, it is the cardinality of a measurable set. The imaginary unit is denoted by $\mathbf{j} = \sqrt{-1}$. Given a function $f(x, \cdot)$, $\nabla_x f(x, \cdot)$ and $\nabla_x^2 f(x, \cdot)$ denote the Jacobian and Hessian of f with respect to x, respectively.

2.2 Problem Formulation

In this section, we present the mathematical formulations for the OPF problem and the new problem of finding the worst-case local minimum of the OPF problem. In order to understand the intuition behind the problem of finding the worst-case local minimum, Figure 2.1 presents the problem for a one-dimensional minimization problem.

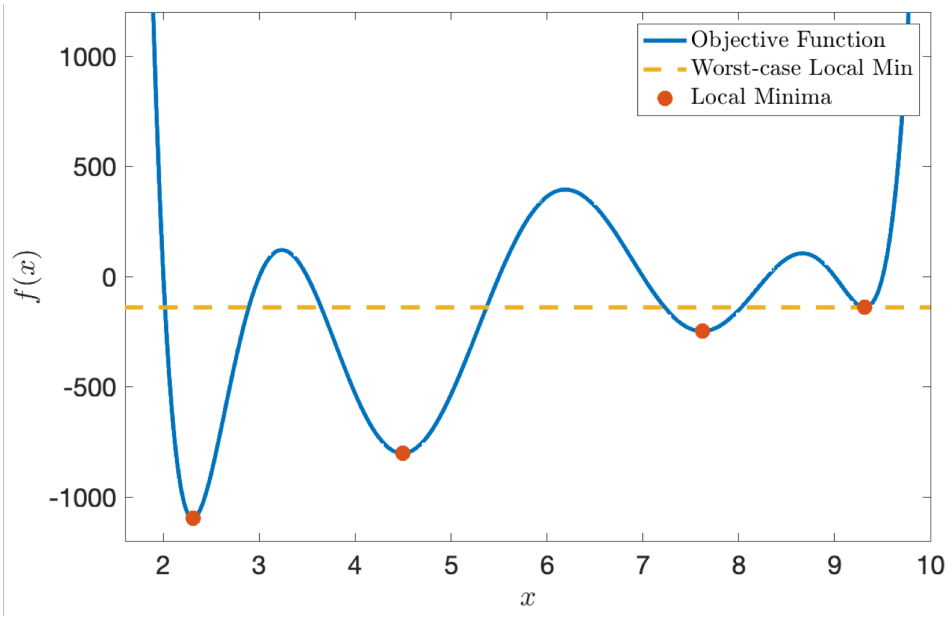

Figure 2.1: The worst-case local minimum problem for a one-dimensional unconstrained optimization problem, i.e. $\min_{x\in\mathbb{R}} f(x)$. The feasible set of the worst-case local minimum problem is the collection of local minima, shown as orange dots. The objective of the worst-case local minimum problem is the maximum of these minima, highlighted by a dashed yellow line.

2.2.1 Classical AC-OPF Problem

Let the power network be defined by a graph $\mathcal{N}(\mathcal{V},\mathcal{E})$, where $\mathcal{V}$ is the set of buses and $\mathcal{E}$ is the set of transmission or distribution lines. Let $\mathcal{G}\subseteq\mathcal{V}$ be the set of buses that are attached to generators.

The classical OPF problem can be written as:

$$\min_{v\in\mathbb{C}^{|\mathcal{V}|}} \sum_{i\in\mathcal{G}} f_i(p_i^g) \tag{2.1a}$$

$$\text{s.t. } \underline{P_i^g} \le p_i^g \le \overline{P_i^g}, \qquad \forall i\in\mathcal{G} \tag{2.1b}$$

$$\underline{Q_i^g} \le q_i^g \le \overline{Q_i^g}, \qquad \forall i\in\mathcal{G} \tag{2.1c}$$

$$\underline{V_i} \le |v_i| \le \overline{V_i}, \qquad \forall i\in\mathcal{V} \tag{2.1d}$$

$$p_i^g - P_i^d = \sum_{(i,j)\in\mathcal{E}} \mathrm{Re}\{v_i(v_i-v_j)^* Y_{ij}^*\}, \qquad \forall i\in\mathcal{G} \tag{2.1e}$$

$$q_i^g - Q_i^d = \sum_{(i,j)\in\mathcal{E}} \mathrm{Im}\{v_i(v_i - v_j)^* Y_{ij}^*\}, \qquad \forall i \in \mathcal{G} \tag{2.1f}$$

$$- P_i^d = \sum_{(i,j)\in\mathcal{E}} \mathrm{Re}\{v_i(v_i - v_j)^* Y_{ij}^*\}, \qquad \forall i \in \mathcal{V} \setminus \mathcal{G} \tag{2.1g}$$

$$- Q_i^d = \sum_{(i,j)\in\mathcal{E}} \mathrm{Im}\{v_i(v_i - v_j)^* Y_{ij}^*\}, \qquad \forall i \in \mathcal{V} \setminus \mathcal{G} \tag{2.1h}$$

where $f_i(\cdot)$ is the power generation cost at bus i, typically a convex polynomial or piecewise linear function. The decision variable v is a vector of complex voltages, where v_i is the complex voltage at bus i. The variables p_i^g and q_i^g are the real and reactive power generated at bus i and can be derived from the vector v. The fixed quantities P_i^d and Q_i^d are the real and reactive power demanded at bus i. The parameters $\underline{P_i^g}$, $\overline{P_i^g}$, $\underline{Q_i^g}$, $\overline{Q_i^g}$, $\underline{V_i}$, $\overline{V_i}$ are respectively the minimum real power generated, maximum real power generated, minimum reactive power generated, maximum reactive power generated, minimum voltage magnitude, and maximum voltage magnitude at bus i. The network parameters Y_{ij}, G_{ij}, and B_{ij} are respectively the complex admittance, conductance, and susceptance for the transmission line between buses i and j, where $Y_{ij} = G_{ij} + \mathbf{j}B_{ij}$. Equations (2.1b) through (2.1d) represent the technological constraints on the power network: the real and reactive power bounds and bounds on voltage magnitudes. Equations (2.1e) through (2.1h) represent Kirchoff's current and voltage laws that dictate the flow of energy in the power network based on nodal voltages, branch admittances, and power injections.

Note that we can also add constraints on line flow capacity to Problem (2.1). To be consistent with the following QCQP formulation, these constraints should have the form $\underline{P_{ij}} \le p_{ij} \le \overline{P_{ij}}$, where p_{ij} is the real power flow over line $(i,j) \in \mathcal{E}$ and $\underline{P_{ij}}$ and $\overline{P_{ij}}$ are its upper and lower bounds.

2.2.2 QCQP Formulation of OPF with Linear Costs

It is well-known that the OPF problem has a QCQP formulation [27]. In order to formulate the OPF problem (2.1) as a QCQP and then find its corresponding worst-case local minimum problem, we will look at power generation costs $f_i(\cdot)$ of a particular form. First, we consider the case where power generation cost functions $f_i(\cdot)$ are linear in p_i^g for all $i \in \mathcal{G}$ and thus are quadratic in terms of the decision vector $v \in \mathbb{C}^{|\mathcal{V}|}$, i.e. generation costs of the form:

$$f_i(p_i^g) = c_{i1}p_i^g + c_{i0}, \qquad \forall i \in \mathcal{G} \tag{2.2}$$

where c_{i1}, and c_{i0} are fixed cost coefficients for each generator $i \in \mathcal{G}$.

By considering linear costs in p_i^g, both the objective function and all constraints can be written as quadratic functions of the decision vector $u \in \mathbb{R}^{2|\mathcal{V}|}$ defined as:

$$u \triangleq \begin{bmatrix} \mathrm{Re}\{v_1\} & \dots \mathrm{Re}\{v_{|\mathcal{V}|}\} & \mathrm{Im}\{v_1\} & \dots & \mathrm{Im}\{v_{|\mathcal{V}|}\} \end{bmatrix}^T \tag{2.3}$$

We also introduce a vector of slack variables $z \in \mathbb{R}^{4|\mathcal{G}|+2|\mathcal{V}|}$ whose entries are associated with the inequality constraints in the OPF formulation (Equations (2.1b) through (2.1d)). Using the slack variables, one can convert inequality constraints to equality constraints (by adding z_i^2 terms) and rewrite the OPF problem (2.1) as the general nonconvex QCQP:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \ & \frac{1}{2} x^T M_0 x + a_0 \\ \text{subject to: } & \frac{1}{2} x^T M_i x + a_i = 0, \qquad \forall i = 1, \ldots, k \end{aligned} \tag{2.4}$$

where $x^T \triangleq \begin{bmatrix} u^T & z^T \end{bmatrix}$, $n \triangleq 4|\mathcal{V}|+4|\mathcal{G}|$ and $k \triangleq 4|\mathcal{V}|+2|\mathcal{G}|$. The matrices $M_0, \ldots, M_k \in \mathbb{R}^{n \times n}$ and the scalars $a_0, a_1, \ldots, a_k \in \mathbb{R}$ can be easily derived from the power flow equations in the OPF formulation (see Appendix 2.A). Note that the matrices $M_0, \ldots, M_k$ are symmetric by construction.

2.2.3 QCQP Formulation of OPF with Quadratic Costs

Next, we broaden the formulation to include cost functions $f_i(\cdot)$ that are quadratic in p_i^g, which is more representative of real-world power systems. We consider costs of the form:

$$f_i(p_i^g) = c_{i2}(p_i^g)^2 + c_{i1} p_i^g + c_{i0}, \qquad \forall i \in \mathcal{G} \tag{2.5}$$

where c_{i2}, c_{i1}, and c_{i0} are fixed cost coefficients for each generator $i \in \mathcal{G}$.

By considering costs of this form, the objective function and all constraints can be written as quadratic functions of the decision vector $u \in \mathbb{R}^{2|\mathcal{V}|+2|\mathcal{G}|}$ defined as:

$$u \triangleq \begin{bmatrix} \text{Re}\{v\}^T & \text{Im}\{v\}^T & (p^g)^T & (q^g)^T \end{bmatrix}^T \tag{2.6}$$

where $p^g \in \mathbb{R}^{|\mathcal{G}|}$ and $q^g \in \mathbb{R}^{|\mathcal{G}|}$ are the vectors formed by concatenating the p_i^g's and q_i^g's respectively.

Just as in the formulation in the preceding section, we also introduce a vector of slack variables $z \in \mathbb{R}^{4|\mathcal{G}|+2|\mathcal{V}|}$ whose entries are associated with the inequality constraints in the OPF formulation (Equations (2.1b) through (2.1d)). Using the slack variables, one can convert inequality constraints to equality constraints (by adding z_i^2 terms) and rewrite the OPF problem (2.1) as the general nonconvex QCQP:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \ & \frac{1}{2} x^T M_0 x + b_0^T x + a_0 \\ \text{subject to: } & \frac{1}{2} x^T M_i x + b_i^T x + a_i = 0, \qquad \forall i = 1, \ldots, k \end{aligned} \tag{2.7}$$

where $x^T \triangleq \begin{bmatrix} u^T & z^T \end{bmatrix}$, $n \triangleq 4|\mathcal{V}| + 6|\mathcal{G}|$, and $k \triangleq 4|\mathcal{V}| + 4|\mathcal{G}|$. The matrices $M_0, \ldots, M_k \in \mathbb{S}^n$, the vectors $b_0, \ldots, b_k \in \mathbb{R}^n$, and the scalars $a_0, \ldots, a_k \in \mathbb{R}$ can be easily derived from the equations in the OPF formulation (2.1) (see Appendix 2.B). Note that the matrices $M_0, \ldots, M_k$ are symmetric by construction.

We use this canonical QCQP (2.7) as the baseline problem throughout the following sections, noting that (2.4) can be written as (2.7) by considering $b_i = \mathbf{0}^n$ for all $i \in \{0, \ldots, k\}$ and that n is defined differently whether we consider the voltage decision vector (2.3) or the voltage and power decision vector (2.6).

Note that any arbitrary QCQP can be reformulated in this form (2.7). Therefore, the results developed in the following sections extend beyond the OPF problem to a wide variety of nonconvex problems that can be formulated as QCQPs. The QCQP in (2.7) may have many local minima, local maxima, and saddle points and is NP-hard to solve in general.

2.2.4 Formulation of the Worst-case Local Minimum Problem

In order to formulate the problem of finding the worst-local minimum of the QCQP in (2.7), we will define a problem whose feasible set is the set of local minima of (2.7) and whose objective is to maximize the objective function of (2.7). We make the mild assumption that all local minima of the OPF problem are regular points, which was shown to be true for a generic OPF problem in [39] and can be extended to a valid QCQP formulation. A local minimum x^* of (2.7) is said to be a regular point if the gradients of the constraints evaluated at x^*, given by $M_1x^* + b_1, \ldots, M_kx^* + b_k$, are linearly independent. Any regular local minimum of (2.7) will satisfy the Karush-Kuhn-Tucker (KKT) conditions. These KKT conditions will be used to define the feasible set of the worst-case local minimum problem.

We introduce Lagrange multipliers $\lambda \in \mathbb{R}^k$ associated with the equality constraints and write the Lagrangian of (2.7) as:

$$L(x, \lambda) = \frac{1}{2}x^T M_0 x + b_0^T x + a_0 + \sum_{i=1}^{k} \lambda_i \left(\frac{1}{2}x^T M_i x + b_i^T x + a_i \right) \tag{2.8}$$

Then, the KKT conditions are given by the equations:

$$0 = \nabla_x L(x, \lambda) = M_0 x + b_0 + \sum_{i=1}^{k} \lambda_i (M_i x + b_i) \tag{2.9a}$$

$$0 = \frac{\partial L(x, \lambda)}{\partial \lambda_i} = \frac{1}{2}x^T M_i x + b_i^T x + a_i, \qquad \forall i = 1, \ldots, k \tag{2.9b}$$

The points $x \in \mathbb{R}^n$ that satisfy both KKT conditions (2.9a) and (2.9b) may be local minima, local maxima, or saddle points. In order to narrow the feasible set of our worst-case problem, we add a second-order optimality condition. Any point that satisfies this condition is called a second-order critical point. Second-order critical points can include local minima and certain saddle points. From [40], we have the second-order necessary condition:

$$\begin{aligned} &y^T(\nabla_x^2 L(x, \lambda))y \geq 0 \\ &\quad \text{for all } y \text{ such that } y^T(M_i x + b_i) = 0, \ \ \forall i = 1, \ldots, k \end{aligned} \tag{2.10}$$

where $\nabla_x^2 L(x,\lambda) = M_0 + \sum_{i=1}^k \lambda_i M_i$.

At first glance, this second-order condition (2.10) involves a possibly infinite number of constraints. Thus, it is more useful to reformulate the second-order necessary condition as a finite-dimensional constraint. To do this, we define a new matrix $\mathbf{M}(x) \in \mathbb{R}^{n\times k}$ whose columns are $M_i x + b_i$ for $i = 1,\ldots,k$, so we can rewrite (2.10) as:

$$y^T(\nabla_x^2 L(x,\lambda))y \geq 0 \text{ for all } y \in \text{null}\left(\mathbf{M}(x)^T\right) \tag{2.11}$$

In order to incorporate the second-order necessary condition (2.11) into constraints of the worst-case local minimum problem, we would have to solve for the nullspace of the matrix $\mathbf{M}(x)^T$, which is a nonconvex operation and has no straightforward convex relaxation. Thus, it is more useful to reformulate (2.11) into a different second-order constraint that is given in the following lemma.

Lemma 1. *An equivalent second-order condition to (2.10) or (2.11) is given as:*

$$M_0 + \sum_{i=1}^k \lambda_i M_i + c\sum_{i=1}^k (M_i x + b_i)(M_i x + b_i)^T \succeq 0 \tag{2.12}$$

which is satisfied for every second-order critical point x, for all values of c above some threshold, i.e. $c > \bar{c}$.

Proof. First, we note that by definition of $\mathbf{M}(x)$, (2.10) and (2.11) are equivalent and $\sum_{i=1}^k (M_i x + b_i)(M_i x + b_i)^T = \mathbf{M}(x)\mathbf{M}(x)^T$.

Next, we show that Condition (2.10) implies Condition (2.12). Define $P \triangleq M_0 + \sum_{i=1}^k \lambda_i M_i$ and $Q \triangleq \mathbf{M}(x)\mathbf{M}(x)^T$, where P is positive semidefinite on the nullspace of Q. It follows from Lemma 4.2.1 in [40] that (2.10) and (2.12) are equivalent for all large values of c.

Lastly, we show that Condition (2.12) implies Condition (2.10). We have that (2.12) is equivalent to:

$$y^T\left(M_0 + \sum_{i=1}^k \lambda_i M_i + c\cdot\mathbf{M}(x)\mathbf{M}(x)^T\right)y \geq 0, \qquad \forall y \in \mathbb{R}^n \tag{2.13}$$

If y is selected to satisfy the equations $y^T(M_i x + b_i) = 0$ for all $i = 1,\ldots,k$, or equivalently $y \in \text{null}(\mathbf{M}(x)^T)$, then the above inequality (2.13) reduces to:

$$y^T\left(M_0 + \sum_{i=1}^k \lambda_i M_i\right)y \geq 0 \tag{2.14}$$

which combined with the equations $y^T(M_i x + b_i) = 0$ for all $i = 1,\ldots,k$ is equivalent to (2.10). □

By combining the KKT first-order necessary conditions (2.9a) and (2.9b) with the second-order necessary condition of the form (2.12), we can define the feasible set of the worst-case local minimum problem. In this worst-case problem, the objective is to maximize the objective of the original QCQP (2.7). The worst-case local minimum problem is formally stated below.

Theorem 1. *Let the objective function of the QCQP (2.7) be denoted as* $g(x) \triangleq \frac{1}{2}x^T M_0 x + b_0^T x + a_0$ *and the sorted second-order critical points of the same QCQP be given as* $x_{(1)}, \ldots, x_{(\ell)}$, *where* $g(x_{(1)}) \geq g(x_{(2)}) \geq \cdots \geq g(x_{(\ell)})$. *We take* $x^* = x_{(i)}$ *to be any second-order critical point with the highest objective value, i.e.* $g(x_{(i)}) = g(x_{(1)})$. *Then, it follows that* x^* *is a globally optimal solution of the following optimization problem (2.15) if* c *is selected to be greater than a certain threshold* $\bar{c}$ *and there exists some point* (x^0, λ^0) *that satisfies (2.15b) through (2.15d):*

$$\max_{x \in \mathbb{R}^n, \lambda \in \mathbb{R}^k} \quad \frac{1}{2}x^T M_0 x + b_0^T x + a_0 \tag{2.15a}$$

$$\textit{subject to:} \quad \frac{1}{2}x^T M_i x + b_i^T x + a_i = 0, \quad \forall i = 1, \ldots, k \tag{2.15b}$$

$$M_0 x + b_0 + \textstyle\sum_{i=1}^k \lambda_i (M_i x + b_i) = 0 \tag{2.15c}$$

$$M_0 + \textstyle\sum_{i=1}^k \lambda_i M_i + c \cdot \boldsymbol{M}(x)\boldsymbol{M}(x)^T \succeq 0 \tag{2.15d}$$

Proof. Let (x^*, λ^*) be a globally optimal solution to optimization problem (2.15). If there exists some feasible point (x^0, λ^0) to (2.15), then we know that any globally optimal solution (x^*, λ^*) satisfies (2.15b) through (2.15d) since the problem (2.15) is deemed to be feasible. Constraint (2.15b) ensures that x^* is a feasible point of (2.7), constraint (2.15c) ensures that x^* is first-order critical point of (2.7), and constraint (2.15d) ensures that x^* is a second-order critical point of (2.7) (see Lemma 1). Since the objective (2.15a) is the same as that in (2.7) and the goal is of (2.15) is to maximize the objective, we have that $x^* = x_{(i)}$ for some second-order critical point $x_{(i)}$ with the highest objective value, i.e. $g(x_{(i)}) = g(x_{(1)})$. □

Note that Problem (2.15), which is nonconvex, finds the worst-case second-order critical point. As a result, its optimal objective value serves as an upper bound on the objective value at the worst-case local minimum.

2.2.5 Decision Version of Worst-case Local Minimum Problem

In the study of the worst-case local minimum problem, it is useful to also examine the related feasibility problem for some given value $\alpha \in \mathbb{R}$:

$$\max_{x \in \mathbb{R}^n, \lambda \in \mathbb{R}^k} \quad 0 \tag{2.16a}$$

$$\text{subject to:} \quad \frac{1}{2}x^T M_i x + b_i^T x + a_i = 0, \quad \forall i = 1, \ldots, k \tag{2.16b}$$

$$M_0 x + b_0 + \textstyle\sum_{i=1}^k \lambda_i (M_i x + b_i) = 0 \tag{2.16c}$$

$$M_0 + \textstyle\sum_{i=1}^k \lambda_i M_i + c \cdot \mathbf{M}(x)\mathbf{M}(x)^T \succeq 0 \tag{2.16d}$$

$$\frac{1}{2}x^T M_0 x + b_0^T x + a_0 \geq \alpha \tag{2.16e}$$

If there is a local minimum to the OPF problem (2.7) whose corresponding cost is greater than or equal to α, then the optimal value of the above problem will be 0. Otherwise, the optimal value of this problem will be $-\infty$. Thus, the problem (2.16) asks whether or not there exists any local minima to the OPF problem above some threshold α. The interpretation of this problem is to certify that any local search solution is a "near-global" solution, i.e. below a given threshold.

For the rest of this chapter, we will focus on the optimization version (2.15) of the worst-case local minimum problem since the decision version (2.16) can be easily deduced from the result of the optimization version. However, we remark that notions from algebraic geometry such as sum of squares and Positivstellensatz could be used to certify that there is no solution to the decision problem. If there is no solution to the above problem, then there exists a certificate that no real solution exists. However, the degree of this certificate may be arbitrarily large, thus these techniques are not efficient in general. See [41–43] for more details on these methods.

2.3 SDP Relaxation of Worst-case Local Min Problem

To bound the worst-case local minimum, Theorem 1 requires solving the nonconvex problem (2.15) to global optimality, which cannot be achieved using local search methods. However, any upper bound on the optimal objective value will still serve the same purpose, and this can be accomplished using convex relaxations. In this chapter, we develop a tightened SDP relaxation of the worst-case local minimum problem in (2.15).

We define a matrix $W \in \mathbb{S}^{n+k+1}$ based on $x \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}^k$ as follows:

$$W \triangleq \begin{bmatrix} 1 \\ x \\ \lambda \end{bmatrix} \begin{bmatrix} 1 & x^T & \lambda^T \end{bmatrix} = \begin{bmatrix} 1 & x^T & \lambda^T \\ x & xx^T & x\lambda^T \\ \lambda & \lambda x^T & \lambda\lambda^T \end{bmatrix} \tag{2.17}$$

We regard W as a 3×3 block matrix with the block entries W_{ij} for all $i, j \in \{1, 2, 3\}$. By lifting the problem into a higher-dimensional matrix W, we move all the nonconvexity of Problem (2.15) to a constraint on the rank of W, which we drop to generate the SDP relaxation given in the theorem below.

Theorem 2. *Having selected a sufficiently large value for the parameter c, the optimal objective value of the following SDP provides an upper bound on the cost of the worst-case local minimum of the OPF problem (2.7):*

$$\max_{W \in \mathbb{S}^{n+k+1}} \quad 1/2 \cdot \mathit{trace}\{M_0 W_{22}\} + b_0^T W_{21} + a_0 \tag{2.18a}$$

$$\textit{subject to:} \quad 1/2 \cdot \mathit{trace}\{M_i W_{22}\} + b_i^T W_{21} + a_i = 0, \quad \forall i = 1, \ldots, k \tag{2.18b}$$

$$M_0 W_{21} + b_0 + \sum_{i=1}^{k} M_i (W_{23})_i + b_i (W_{31})_i = 0 \tag{2.18c}$$

$$M_0 + \sum_{i=1}^{k} M_i (W_{31})_i + c \sum_{i=1}^{k} (M_i W_{22} M_i + b_i W_{12} M_i + M_i W_{21} b_i^T + b_i b_i^T) \succeq 0 \tag{2.18d}$$

$$trace\{M_0 W_{22}\} + b_0^T W_{21} - \sum_{i=1}^{k} b_i^T (W_{23})_i + 2a_i (W_{31})_i = 0 \tag{2.18e}$$

$$W_{11} = 1 \tag{2.18f}$$

$$W \succeq 0 \tag{2.18g}$$

where $(W_{23})_i$ *is the* i^{th} *column of* W_{23} *and* $(W_{31})_i$ *is the* i^{th} *entry of* W_{31} *for all* $i \in \{1, \ldots, k\}$.

Proof. The objective (2.18a) and constraints (2.18b), (2.18c) and (2.18d) follow directly from the objective (2.15a) and constraints (2.15b), (2.15c) and (2.15d), along with the definition of W (2.17). To control the structure of matrix W, we have the constraints (2.18f) and (2.18g), as well as the nonconvex constraint $\text{rank}(W) = 1$. We drop the constraint $\text{rank}(W) = 1$ to obtain a convex feasible region for the problem. Thus, the obtained relaxation (2.18) is an upper bound on (2.15), which is an upper bound on the worst-case local minimum. Finally, to strengthen the SDP relaxation of the worst-case local minimum problem, we add a valid constraint, obtained by multiplying (2.15c) by x^T. Combining this with the constraint (2.15b), we have:

$$\text{trace}\{M_0 x x^T\} + b_0^T x + \sum_{i=1}^{k} \lambda_i (-b_i^T x - 2a_i) = 0 \tag{2.19}$$

By substituting the entries of W for the terms that depend on x and λ, this constraint becomes constraint (2.18e) in the strengthened SDP relaxation. □

2.4 Analysis of SDP Relaxation

Because the worst-case local minimum problem is a nonconvex problem, there could be a non-zero gap between the optimal objective values of the nonconvex problem (2.15) and the tightened convex relaxation (2.18). However, it is desirable to show that the gap is zero in a fundamental class of QCQPs, and therefore the SDP relaxation yields an exact solution to the worst-case local minimum problem for this class.

2.4.1 Particular Case with Exact SDP Relaxation

Consider a particular case of the canonical QCQP in (2.7):

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \ & x^T M_0 x \\ \text{subject to: } & x^T x = 1 \end{aligned} \tag{2.20}$$

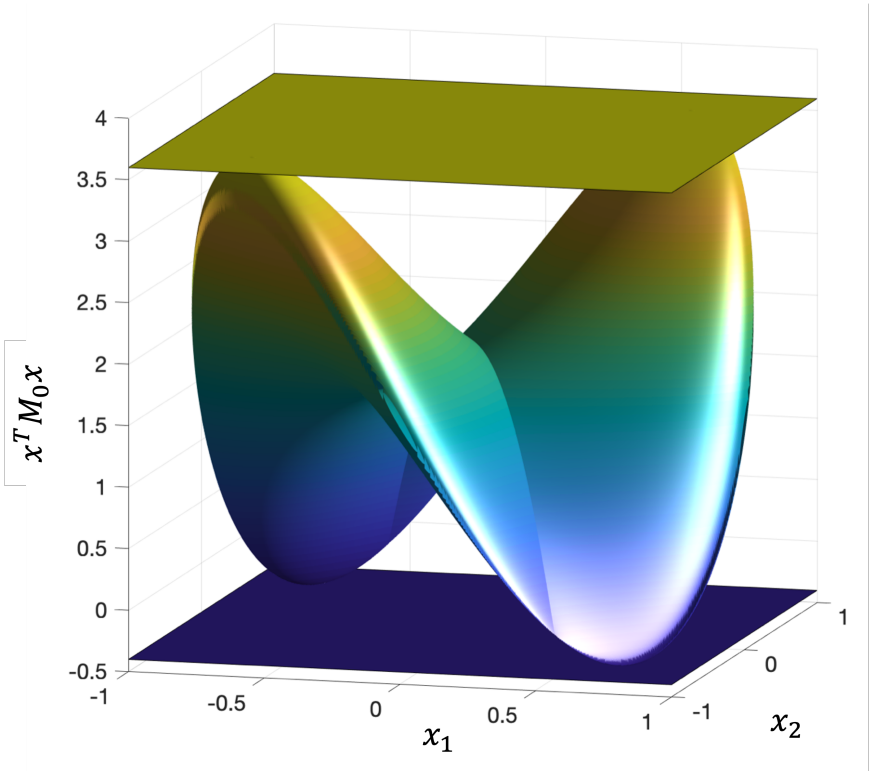

Figure 2.2: The three-dimensional curve shows the nonconvexity of a particular case of the canonical QCQP, given by Equation (2.20), for a 3×3 symmetric matrix M_0 with the eigenvalues -0.4, 1.6, and 3.6. The maximum and minimum eigenvalues are plotted as surfaces in the plane. There are also saddle points in between. The objective value of the worst-case local minimum problem is the same as the minimum eigenvalue, i.e. -0.4.

where M_0 is an arbitrary symmetric matrix with n distinct eigenvalues, ordered as $\mu_1 < \cdots < \mu_n$. associated with the normalized eigenvectors $y_1, \ldots, y_n \in \mathbb{R}^n$. For this problem, one can analytically compute all of the points that satisfy the KKT conditions of (2.20). It can be shown that there are $2n$ KKT points: $\pm y_1$ are local minima, $\pm y_n$ are local maxima, and $\pm y_2, ..., \pm y_{n-1}$ are saddle points. Thus, the cost corresponding to the worst-case local minimum problem is equal to $y_1^T M_0 y_1$. The optimization problem (2.15) provided in Theorem 1 can be written as:

$$\max_{x \in \mathbb{R}^n, \lambda \in \mathbb{R}} \quad x^T M_0 x \tag{2.21a}$$

$$\text{subject to:} \quad x^T x = 1 \tag{2.21b}$$

$$(M_0 + \lambda I)x = 0 \tag{2.21c}$$

$$M_0 + \lambda I + cxx^T \succeq 0 \tag{2.21d}$$

In light of (2.21b) and (2.21c), the only possible solutions are $x = \pm y_i$ and $\lambda = -\mu_i$ for $i = 1, 2, ..., n$. However, the only solution satisfying (2.21d) is $\lambda = -\mu_1$ and $x = \pm y_1$. In addition, (2.21d) is satisfied for any c greater than or equal to zero. This leads to the following result.

Theorem 3. *The SDP relaxation (2.18) returns the cost corresponding to the worst-case local minimum problem (2.21) for $c = 0$, and provides an upper bound for $c > 0$.*

Proof. Using the same methodology as described above, the SDP relaxation of (2.21) is:

$$\begin{aligned}
\max_{W \in \mathbb{S}^{n+2}} \quad & \text{trace}\{M_0 W_{22}\} && \text{(2.22a)}\\
\text{subject to:} \quad & \text{trace}\{W_{22}\} = 1 && \text{(2.22b)}\\
& M_0 W_{21} + W_{23} = 0 && \text{(2.22c)}\\
& M_0 + W_{31} I + c W_{22} \succeq 0 && \text{(2.22d)}\\
& \text{trace}\{M_0 W_{22}\} + W_{31} = 0 && \text{(2.22e)}\\
& W_{11} = 1 && \text{(2.22f)}\\
& W \succeq 0 && \text{(2.22g)}
\end{aligned}$$

Since (2.22) is a relaxation of (2.21) we have that (2.22) upper bounds (2.21). For the constraint $M_0 + W_{31} I \succeq 0$ to hold ((2.22d) in the case where $c = 0$), W_{31} must be greater than or equal to $-\mu_1$. Since $\text{trace}\{M_0 W_{22}\} = -W_{31}$, the optimal objective value of (2.22) is less than or equal to μ_1. In addition, the optimal objective value of (2.21) is equal to μ_1. Combining these inequalities yields the fact that the SDP relaxation (2.22) meant to provide an upper bound on the optimal value of (2.21) also provides a lower bound. Thus, the relaxation is exact. □

Figure 2.2 exemplifies the nonconvexity of the QCQP (2.20) for $n = 3$. This problem has 6 KKT points, and an SDP relaxation that does not incorporate the second-order optimality condition will return the global maximum. However, in light of Theorem 3, the SDP relaxation (2.22) will be able to correctly eliminate the local maxima and saddle points with a negative curvature when $c = 0$.

2.4.2 Choice of Hyperparameter " c "

The hyperparameter c is used to convert the infinite-dimensional second-order optimality condition to a finite-dimensional one. The exact value of c is not needed in the nonconvex model (2.15) since Theorem 2 states that every sufficiently large c enables finding the worst-case second-order critical point. However, since (2.18) is a relaxation of (2.15), selecting an exorbitantly large value for c affects the quality of the solution to the SDP relaxation of the worst-case local minimum problem (2.18).

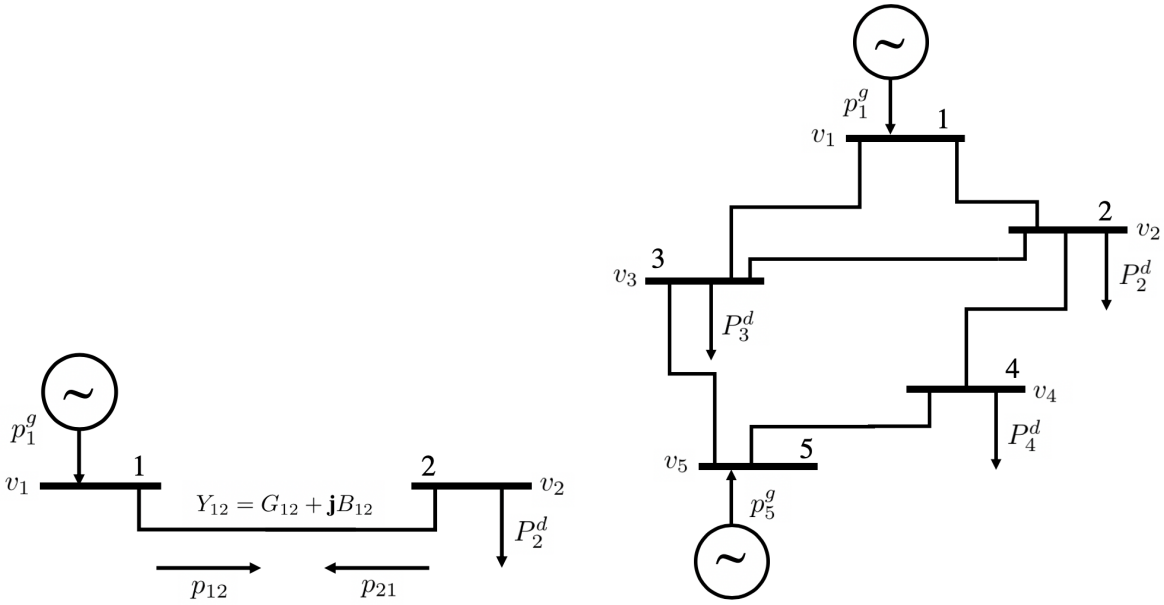

Figure 2.3: WB2 2-bus (left) and WB5 5-bus (right) networks.

For example, in the particular QCQP case described above, the relaxation is exact at $c = 0$ and gradually becomes loose as c increases. This is due to the fact that the second-order necessary condition (2.15d) holds for $c = 0$ at the local minima $\pm y_1$. In general, the smallest c needed in Theorems 2 or 3 coincides with the smallest number c satisfying the second-order condition (2.15d) at the worst-case second-order critical point x^*. Note that x^* is high-dimensional in general, whereas c is a single scalar. Finding a good upper bound on this scalar requires a careful analysis of the matrices $M_1, ..., M_k$ and is an open question.

2.5 Simulations

To test the tightened SDP relaxation of the worst-case local minimum problem (2.18) on benchmark networks, we use MATPOWER to compute the line admittance values and then formulate the matrices $M_0, \ldots, M_k$, the vectors $b_0, \ldots, b_k$, and the scalars $a_0, a_1, \ldots, a_k$ based on the given OPF constraints (see Section 2.2.1 for details). We solve the SDP relaxation of the worst-case local minimum problem using the SDPT3 solver [44]. Note that solving each of the SDP relaxations took less than 30 seconds on a standard laptop.

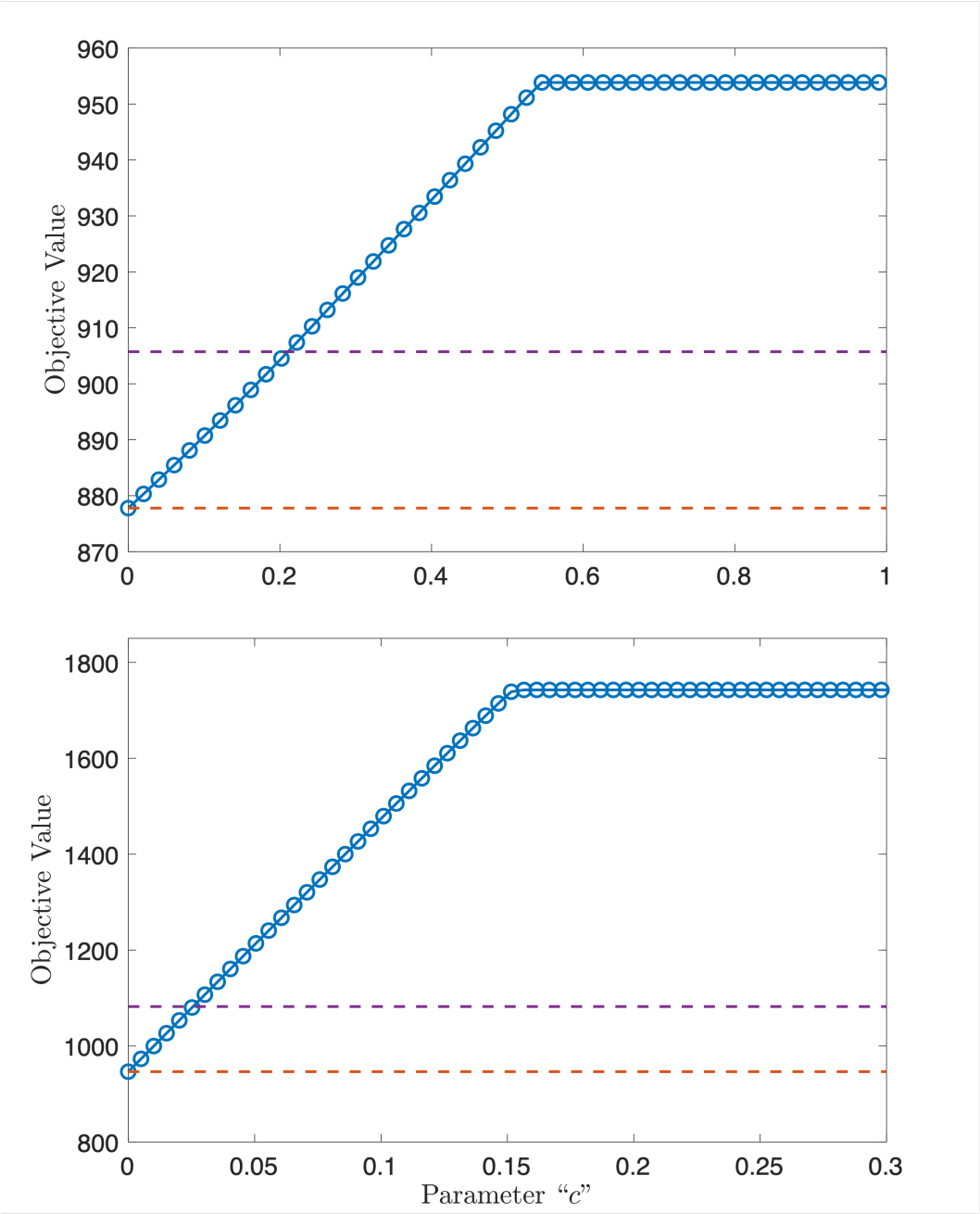

Figure 2.4: Objective values of the solutions to the strengthened SDP relaxation of the worst-case local minimum problem given in (2.18) on the WB2 network (top) and WB5 network (bottom) for varying values of the parameter c are shown by blue circles. The simulation results are compared to the objective values at two known local minima for these networks, shown as dashed lines.

2.5.1 Cases with Known Local Optima

We start by modifying the parameter c in the second-order necessary condition (2.18d) and observe how the choice of c affects the quality of the solution. First, we test the SDP relaxation of the worst-case local minimum problem (2.18) on small networks with known local minima. Using test networks from the online database [45], which are known to have multiple local minima, we run a series of simulations of the SDP relaxation of the worst-case local minimum problem, sweeping over a range of the parameter c (the networks are given in Figure 2.3). From the simulations (see Figure 2.4), it can be observed that the objective value of the relaxation increases with c, until a saturation point is reached. The objective value at this saturation point provides an upper bound on the worst-case local minimum, and it is not tight for the WB2 and WB5 networks. With an understanding of a good choice of c, one can tighten this bound.

2.5.2 IEEE Test Cases with Linear Generation Costs

Next, we test the SDP relaxation of the worst-case local minimum problem (2.18) on some IEEE test networks. Note that for these networks, we have removed the quadratic cost terms so that the costs are linear in terms of real power generation (see Equation (2.2)). We run a series of simulations of the SDP relaxation of the worst-case local minimum problem (2.18), sweeping over a range of the parameter c. We also run 200 simulations of local search on the canonical QCQP in Equation (2.4) with 200 random, feasible initial points. Out of these 200 simulations, we take the solutions of the simulations that converged as the "discovered" local minima. Note that some of these discovered local minima may in fact be saddle points, depending on solver performance. For these local search simulations, we use the FMINCON solver in MATLAB.

It can be observed from the simulation results in Figure 2.5 that the objective value increases with c until a saturation point is reached, at which point the value of c is too high and the relaxation is not exact.

2.5.3 Comparison of Worst-case SDP with Original SDP

We compare the objective value of the solution to the SDP relaxation of the worst-case local minimum problem (2.18) with that of the SDP relaxation of the original problem (the SDP relaxation of (2.7)). Table 2.1 shows the objective value of the solution to the SDP relaxation of the worst-case local minimum problem at the saturation point of c for each of the four test cases. These values provide an upper bound on the worst-case local minimum.

By comparing the worst-case local minimum SDP relaxation to the SDP relaxation of the original problem, we can compute a lower bound on the global optimality degree, which is defined as:

$$\text{Global optimality degree} \triangleq 100\% \times \left(1 - \frac{\text{upper bound - lower bound}}{|\text{upper bound}|}\right) \tag{2.23}$$

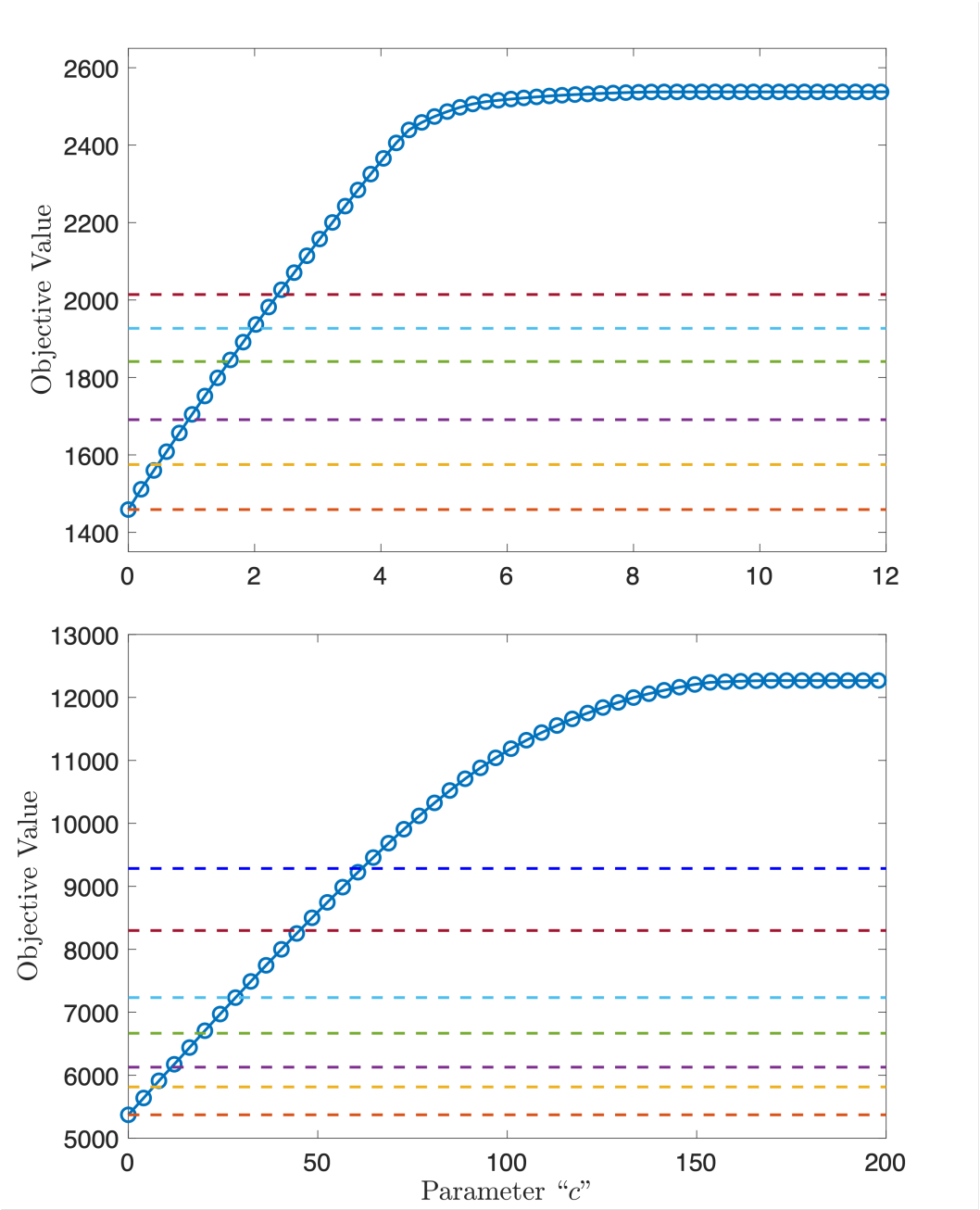

Figure 2.5: Objective values of solutions to the strengthened SDP relaxation of the worst-case local minimum problem (2.18) for the IEEE 9-bus (top) and 14-bus (bottom) networks for varying values of the parameter c are shown by blue circles. These solutions are compared to "discovered" local minima from randomized initializations of local search, shown as dashed lines.

Table 2.1: Objective value of solution to SDP relaxation of worst-case local minimum problem (2.18) (at saturation with respect to c) compared to the objective value of solution to SDP Relaxation of original QCQP (2.7)

Case	Worst-case SDP (2.18) Value of c	Worst-case SDP (2.18) Objective	SDP relaxation of (2.7) Objective	Optimality Degree Lower Bound
WB2	1	953.85	877.78	92.0%
WB5	0.3	1742.55	946.53	54.3%
case9	12	2537.59	1458.91	57.5%
case14	200	12268.8	5371.58	43.8%

The SDP relaxation of the original problem provides a lower bound on the optimal value of the original nonconvex problem (2.7) and is given in Table 2.1. The SDP relaxation of the worst-case local minimum problem (2.18) provides an upper bound on the objective value at any local minima. Thus, by computing the global optimality degree using these two bounds, we can find a lower bound on the global optimality degree for the problem. A summary of how the worst-case problems relate to the original QCQP is given in Figure 2.6.

This lower bound on optimality degree provides a metric of how useful the original SDP relaxation is for the given problem. If the lower bound on optimality degree is high, then any local search solution will be relatively close to the SDP solution. For these cases, such as the WB2 case, the more expensive SDP relaxation is less useful since local search solutions are of relatively high quality. For cases with a larger gap between the SDP solution and the worst-case local minimum, there is some benefit in using a convex relaxation instead of local search for solving the OPF problem.

2.5.4 IEEE Test Cases with Quadratic Generation Costs

In the above simulations, we consider the case where we have linear power generation costs and the voltage decision vector (2.3), such that the original QCQP is formulated as (2.4). In those simulations, the choice of c directly affected the tightness of the SDP relaxation. For that formulation, we see that the SDP relaxation is exact in certain cases and becomes less tight as c increases. Note that due to the structurally different formulations when we consider linear versus quadratic generation costs (due to the different decision vectors in the original QCQP), the parameter c from the above simulations does not relate to the parameter c from the simulations described below that consider quadratic power generation costs.

The situation changes with the QCQP formulation in (2.7) when we consider quadratic power generation costs and the expanded decision variable (2.6). In Figures 2.7 and 2.8, we can see that for the SDP formulation given by (2.18) the choice of c does not appear to have much impact on the quality of the solution. While the choice of c may not have much impact on relaxation tightness, it still must be a large enough value for the problem formulation to be valid. In Table 2.2, we provide the minimum values of c required to satisfy the second-

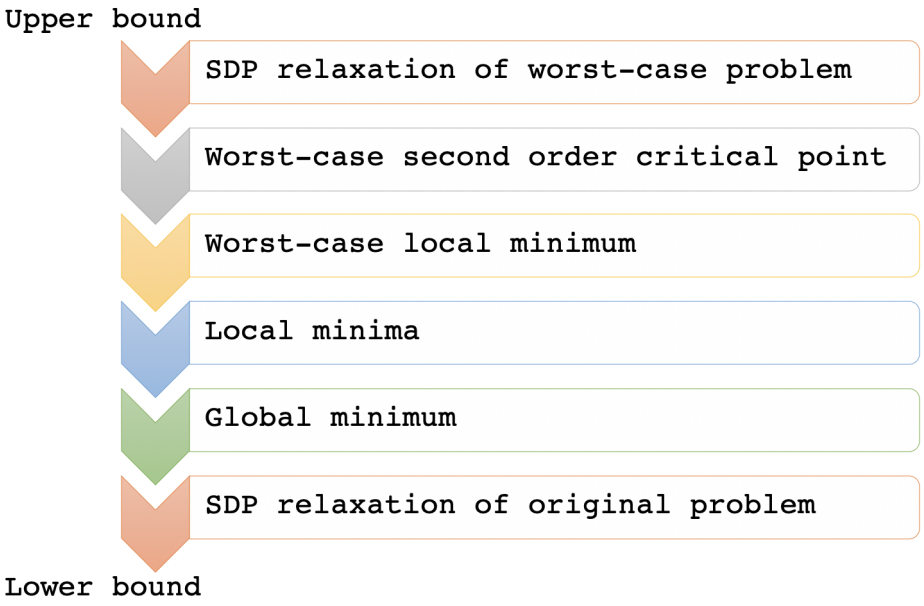

Figure 2.6: Hierarchy of worst-case and original problems and relaxations.

order necessary condition (2.15d) at local minima of various networks, all of which are known to have multiple local minima [4].

Next, we modify the objective function given by (2.18a) to have a penalty term, in order to improve the tightness of the relaxation. We introduce a penalty on the trace of matrix W in order to reduce the rank of the matrix, a technique that is commonly used [46]. From this trace penalty, we subtract out the elements that correspond to the squared real and reactive power generation, so as not to favor local minima with lower objective values. The new problem is given by (2.18) with the modified objective function defined as $\tilde{g}(W)$:

$$\tilde{g}(W) \triangleq 1/2 \cdot \text{trace}\{M_0 W_{22}\} + b_0^T W_{21} + a_0 - \epsilon \cdot (\text{trace}\{W\} - \text{trace}\{W_{22}^{pq}\}) \tag{2.24}$$

where $\epsilon > 0$ is the penalty parameter and W_{22}^{pq} refers to the block of $W_{22} := xx^T$ corresponding to the p_g and q_g variables.

In Figures 2.7 and 2.8, we can see that the relaxation given by (2.18) with modified objective (2.24) becomes tight when ϵ is sufficiently large. A range of values for the penalty parameter will result in a high quality relaxation. However, methods to efficiently tune the penalty parameter for these types of bi-objective problems are an area of open research. In Table 2.3, we provide the solution to the SDP relaxation of the worst-case local minimum problem (2.18) for a few test networks and choices of parameters c and ϵ and compare that value to the objective at the worst-case local minimum as discovered by the local search solver `fmincon`.

Table 2.2: Minimum values of c which satisfy the second-order condition (2.15d) at some local minima of networks from [45] when we consider the QCQP formulation with quadratic generation costs

Network	Type of Minimum	Objective value	Value of c
WB2	Global	877.78	0
WB2	Non-global	905.73	280569
WB5	Global	946.58	3
WB5	Non-global	1082.33	6617
LMBM3	Global	5694.54	0
LMBM3	Non-global	6833.68	1765
LMBM3	Non-global	9677.11	12542
case9mod	Global	3087.84	1031921
case9mod	Non-global	3398.03	2442355
case9mod	Non-global	4265.15	905274
case22loop	Global	4538.80	0
case22loop	Non-global	5929.14	177

Table 2.3: Solution to SDP relaxation of worst-case local minimum problem (2.18) with modified objective (2.24) for various networks from [45] when we consider the QCQP formulation with quadratic generation costs

Network	SDP (2.18) parameter c	SDP (2.18) parameter ϵ	SDP (2.18) Objective	Discovered worst-case min of QCQP (2.7)
WB2	280569	195	917.30	917.17
WB3	0	500	441.73	418.14
WB5	6617	195	1608.55	1082.33

2.6 Conclusions

This chapter formulates the problem of finding the worst-case local minimum for a canonical QCQP, with a focus on the application to OPF. Since the problem is nonconvex, an SDP relaxation is designed to find an upper bound on the objective value at the worst-case local minimum. We show that this SDP relaxation is exact in a particular case with many saddle points. Additionally, we find that the tightness of this upper bound depends on the choice of a parameter in the second-order necessary optimality condition. While the SDP relaxation is not tight in general, by modifying the parameter c and/or by introducing a penalty term, it can become tight, thereby recovering the true worst-case local minimum of OPF. Furthermore, we show that choices in the formulation, i.e. linear versus quadratic power generation costs, affect the choice of the parameter in the second-order necessary optimality condition. By comparing the objective value obtained from the SDP relaxation of the worst-case local minimum problem to the objective value of the SDP relaxation of

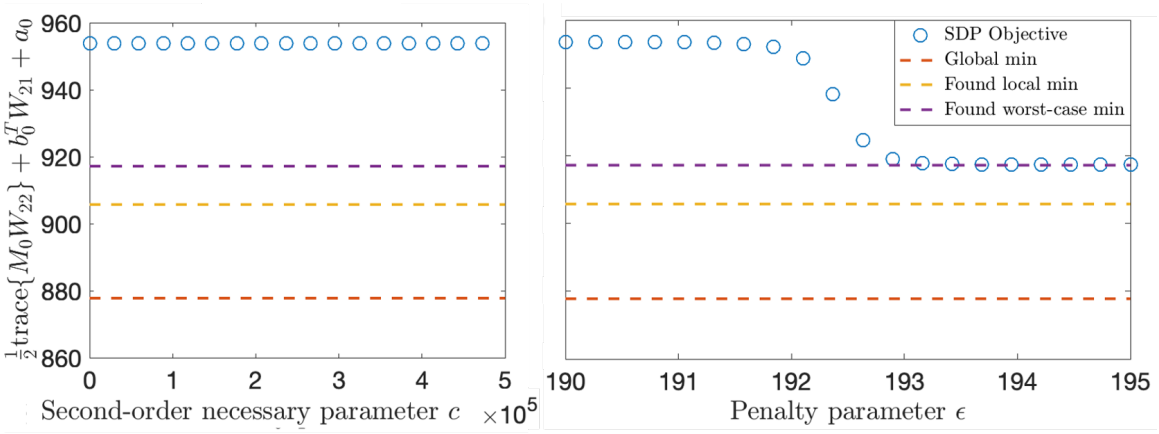

Figure 2.7: The objective value given by the strengthened SDP relaxation of the worst-case local minimum problem on the WB2 2-bus network from [45] is shown in blue circles. The left figure shows a sweep over various values of the second-order necessary parameter c for $\epsilon = 0$. The right figure shows a sweep over various values of the penalty parameter ϵ for $c = 280569$. The local minima for this network, as found by `fmincon`, are shown by dashed lines.

the original problem, we provide a metric on how much SDP can outperform local search. These two SDP relaxations for the upper and lower bounds allow us to evaluate the projected performance of local search methods when good initial points are not available. This method is an efficient tool to bound the worst-case performance of any type of local search method used to solve OPF or other types of nonconvex QCQPs.

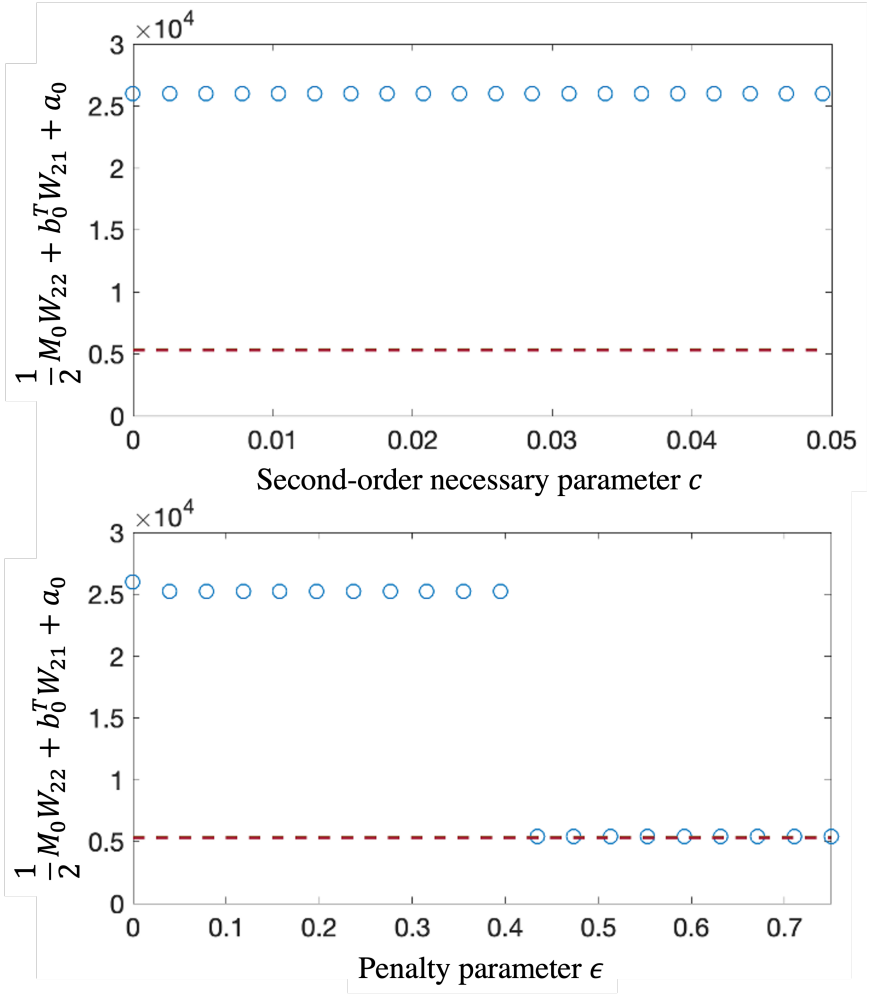

Figure 2.8: The objective value given by the SDP relaxation of the worst-case local minimum problem on the IEEE 9-bus network is shown in blue circles. The top figure shows a sweep over various values of the second-order necessary parameter c for $\epsilon = 0$. The bottom figure shows a sweep over various values of the penalty parameter ϵ for $c = 0.025$. The known local minimum for this network is shown by dashed lines.

Appendix

2.A Derivation of Canonical QCQP for AC-OPF Problem with Linear Power Generation Costs

In this section, we will show how to derive the matrices $M_0, \ldots, M_k$ and scalars $a_0, a_1, \ldots, a_p$ for the canonical QCQP of the AC-OPF problem (2.4) with linear power generation costs (2.2). We have $u \in \mathbb{R}^{2|\mathcal{V}|}$ as defined in (2.3). The real power flow over line (i, j) can be written as:

$$\begin{aligned} p_{ij} &= G_{ij}(\mathrm{Re}\{v_i\}^2 + \mathrm{Im}\{v_i\}^2 - \mathrm{Re}\{v_i\}\mathrm{Re}\{v_j\} - \mathrm{Im}\{v_i\}\mathrm{Im}\{v_j\}) \\ &\qquad + B_{ij}(\mathrm{Re}\{v_i\}\mathrm{Im}\{v_j\} - \mathrm{Re}\{v_j\}\mathrm{Im}\{v_i\}) && \text{(2.25a)} \\ &= G_{ij}(u_i^2 + u_{i+n}^2 - u_i u_j - u_{i+n}u_{j+n}) + B_{ij}(u_i u_{j+n} - u_j u_{i+n}) && \text{(2.25b)} \end{aligned}$$

The reactive power flow over line (i, j) can be written as:

$$\begin{aligned} q_{ij} &= -B_{ij}(\mathrm{Re}\{v_i\}^2 + \mathrm{Im}\{v_i\}^2 - \mathrm{Re}\{v_i\}\mathrm{Re}\{v_j\} - \mathrm{Im}\{v_i\}\mathrm{Im}\{v_j\}) \\ &\qquad + G_{ij}(\mathrm{Re}\{v_i\}\mathrm{Im}\{v_j\} - \mathrm{Re}\{v_j\}\mathrm{Im}\{v_i\}) && \text{(2.26a)} \\ &= -B_{ij}(u_i^2 + u_{i+n}^2 - u_i u_j - u_{i+n}u_{j+n}) + G_{ij}(u_i u_{j+n} - u_j u_{i+n}) && \text{(2.26b)} \end{aligned}$$

Then, the real and reactive power flow over any line can be written as:

$$p_{ij} = u^T A_{ij} u \tag{2.27a}$$

$$q_{ij} = u^T C_{ij} u \tag{2.27b}$$

where A_{ij} is a $2|\mathcal{V}| \times 2|\mathcal{V}|$ matrix of zeros except for the $[i, j, i+n, j+n]$ submatrix is:

$$A_{ij}[i, j, i+n, j+n] = \begin{pmatrix} G_{i,j} & -\frac{1}{2}G_{i,j} & 0 & \frac{1}{2}B_{i,j} \\ -\frac{1}{2}G_{i,j} & 0 & -\frac{1}{2}B_{i,j} & 0 \\ 0 & -\frac{1}{2}B_{i,j} & G_{i,j} & -\frac{1}{2}G_{i,j} \\ \frac{1}{2}B_{i,j} & 0 & -\frac{1}{2}G_{i,j} & 0 \end{pmatrix} \tag{2.28}$$

where C_{ij} is a $2|\mathcal{V}| \times 2|\mathcal{V}|$ matrix of zeros except for the $[i, j, i+n, j+n]$ submatrix is:

$$C_{ij}[i, j, i+n, j+n] = \begin{pmatrix} -B_{i,j} & \frac{1}{2}B_{i,j} & 0 & \frac{1}{2}G_{i,j} \\ \frac{1}{2}B_{i,j} & 0 & -\frac{1}{2}G_{i,j} & 0 \\ 0 & -\frac{1}{2}G_{i,j} & -B_{i,j} & \frac{1}{2}B_{i,j} \\ \frac{1}{2}G_{i,j} & 0 & \frac{1}{2}B_{i,j} & 0 \end{pmatrix} \tag{2.29}$$

The net real power flow out of any bus i is then:

$$p_i = \sum_{(i,j)\in\mathcal{E}} p_{ij} = \sum_{(i,j)\in\mathcal{E}} u^T A_{ij} u = u^T A_i u \tag{2.30}$$

where $A_i = \sum_{(ij)\in\mathcal{E}} A_{ij}$.

The net reactive power flow out of any bus i is then:

$$q_i = \sum_{(i,j)\in\mathcal{E}} q_{ij} = \sum_{(i,j)\in\mathcal{E}} u^T C_{ij} u = u^T C_i u \tag{2.31}$$

where $C_i = \sum_{(i,j)\in\mathcal{E}} C_{ij}$.

The constraint $\underline{V_i} \le |v_i| \le \overline{V_i}$ in (2.1) can be written in the following forms:

$$\underline{V_i}^2 \le |v_i|^2 \le \overline{V_i}^2 \tag{2.32a}$$
$$\underline{V_i}^2 \le \mathrm{Re}\{v_i\}^2 + \mathrm{Im}\{v_i\}^2 \le \overline{V_i}^2 \tag{2.32b}$$
$$\underline{V_i}^2 \le u_i^2 + u_{i+n}^2 \le \overline{V_i}^2 \tag{2.32c}$$

Then, the voltage magnitude bounds for bus i can be written:

$$\underline{V_i}^2 \le u^T D_i u \le \overline{V_i}^2 \tag{2.33}$$

where D_i is an $2|\mathcal{V}| \times 2|\mathcal{V}|$ matrix of zeros except for the $[i, i+n]$ submatrix is:

$$D_i[i, i+n] = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{2.34}$$

To redefine the AC-OPF problem (2.1) in terms of u, we have:

$$\min_{u\in\mathbb{R}^{2|\mathcal{V}|}} \quad \sum_{i\in\mathcal{G}} c_{i1}(u^T A_i u + P_i^d) + c_{i0} \tag{2.35a}$$
$$\text{s.t. } \underline{P_i^g} - P_i^d \le u^T A_i u \le \overline{P_i^g} - P_i^d, \qquad \forall i \in \mathcal{G} \tag{2.35b}$$
$$\underline{Q_i^g} - Q_i^d \le u^T C_i u \le \overline{Q_i^g} - Q_i^d, \qquad \forall i \in \mathcal{G} \tag{2.35c}$$
$$\underline{V_i}^2 \le u^T D_i u \le \overline{V_i}^2, \qquad \forall i \in \mathcal{V} \tag{2.35d}$$
$$0 = u^T A_i u + P_i^d, \qquad \forall i \in \mathcal{V} \setminus \mathcal{G} \tag{2.35e}$$
$$0 = u^T C_i u + Q_i^d, \qquad \forall i \in \mathcal{V} \setminus \mathcal{G} \tag{2.35f}$$

If we take $z \in \mathbb{R}^{4|\mathcal{G}|+2|\mathcal{V}|}$ to be the slack variables corresponding to the inequality constraints and $x^T = \begin{bmatrix} u^T & z^T \end{bmatrix}$, then we can convert inequality constraints of the form $u^T A u +$

$b^T u + c \leq 0$ to equality constraints of the form $u^T A u + b^T u + c + z^2 = 0$. Using this method, we arrive at (2.4) by finding the matrices $M_0, \ldots, M_k$ and scalars $a_0, a_1, \ldots, a_k$ in the method described below. Let e_i be the i-th canonical basis vector in $\mathbb{R}^p$. For the objective function, we have:

$$M_0 \triangleq 2 \begin{bmatrix} \sum_{i \in \mathcal{G}} c_{i1} A_i & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{(n+k)\times(n+k)} \tag{2.36a}$$

$$a_0 \triangleq \sum_{i \in \mathcal{G}} c_{i1} P_i^d + c_{i0} \tag{2.36b}$$

For the upper bound on real power generation, we have:

$$M_i \triangleq 2 \begin{bmatrix} A_i & 0 \\ 0 & e_i e_i^T \end{bmatrix} \in \mathbb{R}^{(n+k)\times(n+k)} \tag{2.37a}$$

$$a_i \triangleq P_i^d - \overline{P_i^g} \tag{2.37b}$$

Similar matrices can be found for the lower bound on real power generation, the upper and lower bounds on reactive power generation, and the upper and lower bounds on voltage magnitude.

For the real power equality constraints we have:

$$M_i \triangleq 2 \begin{bmatrix} A_i & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{(n+k)\times(n+k)} \tag{2.38a}$$

$$a_i \triangleq P_i^d \tag{2.38b}$$

Similarly for the reactive power equality constraints.

2.B Derivation of Canonical QCQP for AC-OPF Problem with Quadratic Power Generation Costs

In this section, we will show how to derive the matrices $M_0, \ldots, M_k$, vectors $b_0, \ldots, b_k$, and scalars $a_0, \ldots, a_k$ for the canonical QCQP of the AC-OPF problem (2.7) with quadratic power generation costs (2.5). We have $u \in \mathbb{R}^{2|\mathcal{V}|+2|\mathcal{G}|}$ as defined in (2.6). Just as in the preceding section, we can write real power flow p_{ij} over line (i, j) as (2.27a) with A_{ij} defined by (2.28) and reactive power flow q_{ij} over line (i, j) as (2.27b) with C_{ij} defined by (2.29). We also have Equations (2.30) and (2.31) for real and reactive power injection in matrix form as a function of u, and Equation (2.33) for voltage magnitude bounds as a function of u. Unlike in the preceding formulation, we also have the following relations due to the structure of u:

$$p_i^g = e_{2|\mathcal{V}|+i}^T u, \quad \forall i = 1, \ldots, |\mathcal{G}| \tag{2.39a}$$

$$q_i^g = e_{2|\mathcal{V}|+|\mathcal{G}|+i}^T u, \quad \forall i = 1, \ldots, |\mathcal{G}| \tag{2.39b}$$

From these relations, we also have:

$$(p_i^g)^2 = u^T e_{2|\mathcal{V}|+i} e_{2|\mathcal{V}|+i}^T u, \quad \forall i = 1, \ldots, |\mathcal{G}| \tag{2.40}$$

Define the following:

$$M_0 \triangleq 1/2 \cdot \sum_{i=1}^{|\mathcal{G}|} c_{i2} e_{2|\mathcal{V}|+i} e_{2|\mathcal{V}|+i}^T \tag{2.41a}$$

$$b_0 \triangleq \sum_{i=1}^{|\mathcal{G}|} c_{i1} e_{2|\mathcal{V}|+i} \tag{2.41b}$$

$$a_0 \triangleq \sum_{i=1}^{|\mathcal{G}|} c_{i0} \tag{2.41c}$$

$$b_i \triangleq e_{2|\mathcal{V}|+i} \tag{2.41d}$$

$$d_i \triangleq e_{2|\mathcal{V}|+|\mathcal{G}|+i} \tag{2.41e}$$

Then, can redefine the AC-OPF problem in terms of u:

$$\min_{x \in \mathbb{R}^{2|\mathcal{V}|+2|\mathcal{G}|}} \quad 1/2 \cdot x^T M_0 x + b_0^T x + a_0 \tag{2.42a}$$

$$\text{s.t. } \underline{P_i^g} \leq b_i^T x \leq \overline{P_i^g}, \qquad \forall i \in \mathcal{G} \tag{2.42b}$$

$$\underline{Q_i^g} \leq d_i^T x \leq \overline{Q_i^g}, \qquad \forall i \in \mathcal{G} \tag{2.42c}$$

$$\underline{V_i}^2 \leq x^T D_i x \leq \overline{V_i}^2, \qquad \forall i \in \mathcal{V} \tag{2.42d}$$

$$0 = x^T A_i x - b_i^T x + P_i^d, \qquad \forall i \in \mathcal{G} \tag{2.42e}$$

$$0 = x^T C_i x - d_i^T x + Q_i^d, \qquad \forall i \in \mathcal{G} \tag{2.42f}$$

$$0 = x^T A_i x + P_i^d, \qquad \forall i \in \mathcal{V} \setminus \mathcal{G} \tag{2.42g}$$

$$0 = x^T C_i x + Q_i^d, \qquad \forall i \in \mathcal{V} \setminus \mathcal{G} \tag{2.42h}$$

We also add line capacity constraints:

$$|p_{ij}| \leq \overline{P_{ij}}, \quad \forall (i,j) \in \mathcal{E} \tag{2.43a}$$

$$\Updownarrow$$

$$x^T A_{ij} x \leq \overline{P_{ij}}, \quad \forall (i,j) \in \mathcal{E} \tag{2.43b}$$

$$-x^T A_{ij} x \leq \overline{P_{ij}}, \quad \forall (i,j) \in \mathcal{E} \tag{2.43c}$$

It is clear that this formulation (2.42) can be written in the canonical form (2.7) by introducing the slack variable z corresponding to the inequality constraints as described in the preceding section.

2.3 Bibliography

[1] E. Glista and S. Sojoudi, "Convex model to evaluate worst-case performance of local search in the optimal power flow problem," *IEEE 59th Conference on Decision and Control*, 2020.

[2] J. Carpentier, "Contribution to the economic dispatch problem," *Bulletin de la Societe Francoise des Electriciens*, vol. 3, no. 8, pp. 431–447, 1962.

[3] M. B. Cain, R. P. O'Neill, and A. Castillo, "History of optimal power flow and formulations," *Federal Energy Regulatory Commision*, December 2012.

[4] W. Bukhsh, A. Grothey, K. McKinnon, and P. Trodden, "Local solutions of the optimal power flow problem," *IEEE Transactions on Power Systems*, November 2013.

[5] M. R. Narimani, D. K. Molzahn, D. Wu, and M. L. Crow, "Empirical investigation of non-convexities in optimal power flow problems," in *2018 Annual American Control Conference (ACC)*, 2018, pp. 3847–3854.

[6] D. Bienstock and A. Verma, "Strong NP-hardness of AC power flows feasibility," *Operations Research Letters*, vol. 47, no. 6, pp. 494–501, 2019.

[7] K. Lehmann, A. Grastien, and P. Van Hentenryck, "AC-feasibility on tree networks is NP-hard," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 798–801, 2016.

[8] O. Alsac and B. Stott, "Optimal load flow with steady-state security," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-93, no. 3, pp. 745–751, 1974.

[9] A. Monticelli, M. V. F. Pereira, and S. Granville, "Security-constrained optimal power flow with post-contingency corrective rescheduling," *IEEE Transactions on Power Systems*, vol. 2, no. 1, pp. 175–180, 1987.

[10] A. Castillo, C. Laird, C. A. Silva-Monroy, J.-P. Watson, and R. P. O'Neill, "The unit commitment problem with AC optimal power flow constraints," *IEEE Transactions on Power Systems*, vol. 31, no. 6, pp. 4853–4866, 2016.

[11] B. Knueven, J. Ostrowski, and J.-P. Watson, "On mixed-integer programming formulations for the unit commitment problem," *INFORMS Journal on Computing*, vol. 32, no. 4, pp. 857–876, 2020. [Online]. Available: https://doi.org/10.1287/ijoc.2019.0944

[12] F. Capitanescu, J. Martinez Ramos, P. Panciatici, D. Kirschen, A. Marano Marcolini, L. Platbrood, and L. Wehenkel, "State-of-the-art, challenges, and future trends in security constrained optimal power flow," *Electric Power Systems Research*, vol. 81, no. 8, pp. 1731 – 1741, 2011.

[13] A. Konar, N. D. Sidiropoulos, and O. Mehanna, "Parametric frugal sensing of power spectra for moving average models," *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1073–1085, 2015.

[14] M. Soltanalian and P. Stoica, "Designing unimodular codes via quadratic optimization," *IEEE Transactions on Signal Processing*, vol. 62, no. 5, pp. 1221–1234, mar 2014.

[15] U. Mustafa, A. Akansu, and M. Avellaneda, "Portfolio risk in multiple frequencies," *IEEE Signal Processing Magazine*, vol. 28, pp. 61 – 71, 10 2011.

[16] A. Castillo and R. P. O'Neill, "Survey of approaches to solving OPF," *Federal Energy Regulatory Commision*, March 2013.

[17] F. Zohrizadeh, C. Josz, M. Jin, R. Madani, J. Lavaei, and S. Sojoudi, "A survey on conic relaxations of optimal power flow problem," *European Journal of Operational Research*, vol. 287, no. 2, pp. 391–409, 2020.

[18] D. I. Sun, B. Ashley, B. Brewer, A. Hughes, and W. F. Tinney, "Optimal power flow by Newton approach," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-103, no. 10, pp. 2864–2880, 1984.

[19] G. L. Torres and V. H. Quintana, "An interior-point method for nonlinear optimal power flow using voltage rectangular coordinates," *IEEE Transactions on Power Systems*, vol. 13, no. 4, pp. 1211–1218, 1998.

[20] I. M. Nejdawi, K. A. Clements, and P. W. Davis, "An efficient interior point method for sequential quadratic programming based optimal power flow," *IEEE Transactions on Power Systems*, vol. 15, no. 4, pp. 1179–1183, 2000.

[21] R. A. Jabr, A. H. Coonick, and B. J. Cory, "A primal-dual interior point method for optimal power flow dispatching," *IEEE Transactions on Power Systems*, vol. 17, no. 3, pp. 654–662, 2002.

[22] C. Y. Chung, W. Yan, and F. Liu, "Decomposed predictor-corrector interior point method for dynamic optimal power flow," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1030–1039, 2011.

[23] J. Momoh and J. Zhu, "Improved interior point method for OPF problems," *IEEE transactions on power systems*, vol. 14, no. 3, pp. 1114–1120, August 1999.

[24] Q. Jiang and H.-D. Chiang, "Sequential feasible optimal power flow: Theoretical basis and numerical implementation," *European Transactions on Electrical Power*, vol. 20, no. 6, pp. 695–709, 2010.

[25] S. Park, E. Glista, J. Lavaei, and S. Sojoudi, "Homotopy method for finding the global solution of post-contingency optimal power flow," *American Control Conference*, 2020.

[26] X. Bai, H. Wei, K. Fujisawa, and Y. Wang, "Semidefinite programming for optimal power flow problems," *International Journal of Electrical Power & Energy Systems*, vol. 30, no. 6-7, pp. 383–392, July 2008.

[27] J. Lavaei and S. H. Low, "Zero duality gap in optimal power flow problem," *IEEE Transactions on Power Systems*, vol. 27, no. 1, February 2012.

[28] B. Zhang and D. Tse, "Geometry of feasible injection region of power networks," *Allerton Conference*, September 2011.

[29] S. Sojoudi and J. Lavaei, "Physics of power networks makes hard optimization problems easy to solve," *2012 IEEE Power and Energy Society General Meeting*, pp. 1–8, 2012.

[30] J. Lavaei, D. Tse, and B. Zhang, "Geometry of power flows and optimization in distribution networks," *IEEE Transactions on Power Systems*, vol. 29, no. 2, March 2014.

[31] S. Sojoudi and J. Lavaei, "Exactness of semidefinite relaxations for nonlinear optimization problems with underlying graph structure," *SIAM Journal on Optimization*, vol. 24, no. 4, pp. 1746–1778, 2014.

[32] B. Kocuk, S. S. Dey, and X. A. Sun, "Inexactness of SDP relaxation and valid inequalities for optimal power flow," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 642–651, 2016.

[33] B. Lesieutre, D. Molzahn, A. Borden, and C. DeMarco, "Examining the limits of the application of semidefinite programming to power flow problems," *Allerton Conference*, September 2011.

[34] D. K. Molzahn and I. A. Hiskens, "Convex relaxations of optimal power flow problems: An illustrative example," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 63, no. 5, pp. 650–660, 2016.

[35] R. Jabr, "Radial distribution load flow using conic programming," *IEEE Transactions on Power Systems*, vol. 21, no. 3, pp. 1458 – 1459, August 2006.

[36] B. Kocuk, S. S. Dey, and X. A. Sun, "Strong SOCP relaxations for the optimal power flow problem," *Operations Research*, vol. 64, no. 6, pp. 1177–1196, dec 2016.

[37] C. Coffrin, H. Hijazi, and P. Van Hentenryck, "Strengthening the SDP relaxation of AC power flows with convex envelopes, bound tightening, and valid inequalities," *IEEE Transactions on Power Systems*, vol. 32, no. 5, September 2017.

[38] R. Madani, S. Sojoudi, and J. Lavaei, "Convex relaxation for optimal power flow problem: Mesh networks," *IEEE Transactions on Power Systems*, vol. 30, no. 1, pp. 199–211, January 2015.

[39] A. Hauswirth, S. Bolognani, G. Hug, and F. Dörfler, "Generic existence of unique Lagrange multipliers in AC optimal power flow," *IEEE Control Systems Letters*, vol. 2, no. 4, pp. 791–796, 2018.

[40] D. P. Bertsekas, *Nonlinear programming*, 3rd ed. Athena scientific, 2016.

[41] B. Sturmfels, "Solving systems of polynomial equations," in *American Mathematical Society, CBMS Regional Conferences Series*, no. 97, 2002.

[42] P. A. Parrilo, "Sum of squares programs and polynomial inequalities," 2004.

[43] I. A. Fotiou, P. Rostalski, P. A. Parrilo, and M. Morari, "Parametric optimization and optimal control using algebraic geometry methods," *International Journal of Control*, vol. 79, no. 11, February 2006.

[44] K. C. Toh, M. J. Todd, and R. H. Tütüncü, "SDPT3 — a MATLAB software package for semidefinite programming, version 1.3," *Optimization Methods and Software*, vol. 11, no. 1-4, pp. 545–581, 1999.

[45] W. Bukhsh, May 2013. [Online]. Available: https://www.maths.ed.ac.uk/optenergy/LocalOpt/

[46] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum rank solutions to linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, p. 471–501, 2010.

Chapter 3

Homotopy Method for Post-contingency Optimal Power Flow

In this chapter, we present an efficient homotopy method for finding the global solution of post-contingency optimal power flow.[1]

3.1 Introduction

As presented in Chapter 2, optimal power flow (OPF) is a fundamental tool for power system network analysis [4]. The goal of OPF is to find a minimum cost production of the committed generating units while satisfying the technical constraints of the power system. To ensure security, additional care must be taken so that the system is able to operate within the technical limits even in the event of component failures, i.e. "contingencies." A system that is impaired by a contingency is "stressed" in the sense that a loss of a transmission line or a generator will generally make it more difficult for the power system to meet consumer demand. Thus, in practice, power operators are concerned with security-constrained OPF (SCOPF) instead of an idealistic OPF problem (such as the one found in Equation (2.1)).

SCOPF can be regarded as a large number (as high as 10,000) of OPF problems coupled to each other via physical constraints, where the first OPF corresponds to the operating point of the system under the normal conditions and the remaining ones are associated with a predetermined set of contingencies [5]. The SCOPF problem includes constraints that tie the normal-operation OPF problem to the contingencies, such as constraints that impose additional limits on line flows and bus voltages for the contingencies. Since the SCOPF problem is a larger version of the original OPF problem, it is also nonconvex and thus is hard to solve to global optimality.

[1] Chapter 3 includes materials from [1, 2] that were previously published. Some of this material also appears in [3] due to the collaborative nature of this work with SangWoo Park.

3.1.1 Existing Methods to Solve SCOPF Problems

SDP relaxations have been shown to yield high-quality solutions of SCOPF on small test networks [6, 7]. However, convex relaxations such as SDP and even local search methods are not scalable for some real-world SOCP problems that involve large contingency sets [5]. There are two primary methods to address the huge size of the SCOPF problem. One method is called "contingency filtering" or "screening," in which a smaller subset of binding contingencies is found to replace the full set of contingencies [8–10]. The goal of these methods is to find a contingency subset that is sufficient to recover the same SCOPF solution as the case where the full set of contingencies is considered. However, sometimes this method yields a set of binding contingencies that can still be quite large, producing a reduced SCOPF problem that is still too computationally burdensome to be solved in the desired time. In this case, it can be helpful to consider the second method: approximating or simplifying the formulation of contingencies in the SCOPF problem. There have been many proposed methods to simplify the model of post-contingency states in SCOPF, such as Benders decomposition, linearization of the power flow equations, Lagrangian relaxation, and network compression [11–14]. These contingency selection, approximation, and decomposition techniques can be combined to generate heuristic solutions to large-scale SCOPF problems, as in [15, 16]. Additionally, recent research has applied approaches from distributed control, stochastic programming, and machine learning to solve the SCOPF problem [17–20]. Note that for any of these simplification methods, there is no guarantee that the solution is locally optimal or even feasible for the post-contingency state [5].

Combining both contingency filtering and approximation methods to reduce the SCOPF problem, we can view the reduced SCOPF as a single OPF for the normal operating state (denoted as the "base case") subject to many surrogate contingency constraints expressed in terms of the base case variables, as is done in [21]. We will refer to this reduced SCOPF problem as the *base*-OPF problem. SDP and other methods may be used to solve such *base*-OPF problems, but they are mainly helpful to find the operating voltages for the base case, not those for each of the contingencies. While the *base*-OPF performs risk-aware decision making for the base case scenario by considering contingencies, finding the optimal (or near optimal) operating point for the base case considering the possibility of contingencies, it fails to find globally optimal operating points for each of the contingency scenarios. Thus, for each contingency, the system operator must solve an additional OPF-based optimization or feasibility problem in order to find the operating point of the post-contingency system, which leads to the post-contingency OPF problems.

3.1.2 Existing Methods to Solve Post-contingency OPF Problems

Currently, there is a rather limited literature that attempts to optimize the post-contingency scenarios. In the classic work [21], the optimal post-contingency actions were modeled as sub-problems and explicitly included in the SCOPF formulation. In order to overcome

the complexity of this two-level optimization problem, an algorithm based on Bender's decomposition was developed, for which convergence is not guaranteed for general nonconvex problems. More recently, the work in [22] proposed an approach to determine an optimal combination of preventive and corrective actions taking into the account the system dynamics, while [23] introduced a hybrid computational strategy to solve the pre-contingency and post-contingency OPF problems. In [24], the authors perform optimization over the post-contingency recourse variables using an interior-point solver. None of the previous works have ventured into finding the global optimum of each of the post-contingency OPF problems (which we refer to as the *contingency*-OPF problems), mainly because applying a computationally burdensome algorithm such as SDP to each of the contingency scenarios is unrealistic.

3.1.3 Contributions

In this chapter, we develop a fast homotopy method to solve each of the *contingency*-OPF problems to global optimality given the solution to the *base*-OPF problem.

Instead of solving for the solution to a *contingency*-OPF problem directly via a descent numerical algorithm or convex relaxation, we generate and solve (using local search) a series of intermediate optimization problems wherein we gradually remove a component of the power system. We show that the effectiveness of homotopy to find a global solution of the *contingency*-OPF problem is dependent on the choice of homotopy path, and we introduce new theory to characterize desirable homotopy paths.

The remainder of the chapter is organized as follows. In Section 3.2, we provide a literature review on homotopy methods and explain how they relate to our approach. In Section 3.3, we present the formulation of the two-stage Security-constrained Optimal Power Flow that can be decomposed into the *base*-OPF and *contingency*-OPF. Next, in Section 3.4, we introduce the homotopy method that connects *contingency*-OPF to *base*-OPF via parametrization. In Section 3.5, we develop theoretical results to characterize cases when homotopy will lead to a global solution of the deformed problem. Finally, in Section 3.6 we implement the homotopy method on actual test cases and verify its effectiveness.

3.1.4 Notations

The symbols $\mathbb{R}^N$ and $\mathbb{C}^N$ denote the spaces of N-dimensional real and complex vectors, respectively. The symbols $(\cdot)^T$ and $(\cdot)^*$ denote the transpose and conjugate transpose of a vector or matrix. $\mathrm{Re}\{\cdot\}$ and $\mathrm{Im}\{\cdot\}$ denote the real and imaginary part of a given scalar or matrix. The symbol $|\cdot|$ is the absolute value operator if the argument is a scalar, vector, or matrix; otherwise, it is the cardinality of a measurable set. The elementwise multiplication of two matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{m\times n}$ is denoted as $A \odot B$. Let $\mathbf{1}_n$ and $\mathbf{0}_n$ denote the n-dimensional vectors of ones and zeros, respectively. Furthermore, $\mathbf{1}_n^k$ denotes an n-dimensional vector of ones except for the k-th element that is zero. The imaginary unit is

denoted by $\mathbf{j} = \sqrt{-1}$. Given a function $f(x, \cdot)$, $\nabla_x f(x, \cdot)$ and $\nabla_x^2 f(x, \cdot)$ denote the Jacobian and Hessian of f with respect to x, respectively.

3.2 Background on Homotopy Methods

Homotopy and continuation methods have long been used in mathematics and engineering to solve systems of nonlinear algebraic equations [25]. Continuation methods in mathematics describe the continuous transformation of an easy problem into the given hard problem [26]. A parametrized homotopy map defines this transformation, which generates a family of problems with each problem initialized by the solution from the previous problem. Path tracing methods such as the predictor-corrector method are used to numerically trace the homotopy map [27]. The benefit of homotopy methods compared to other iterative methods is that homotopy methods may yield global rather than local convergence. While homotopy methods have been shown to be accurate and robust, they are computationally expensive and should be reserved for highly nonlinear problems [26]. These methods are most useful for problems where convergence to a global solution is heavily dependent on a good initial point, which can be hard to obtain.

The development of probability-one homotopy methods in the 1970s created a globally-convergent framework for solving nonlinear systems of equations [28]. For these probability-one methods, almost all choices of the parameter in the homotopy map yield no singular points in the Jacobian and thereby global convergence.

3.2.1 Homotopy for Power Systems

Homotopy methods have been applied in the field of power systems, primarily to solve the power flow (PF) problem for cases that do not converge. The continuation power flow (CPF) problem is used to find a set of solutions of the power flow problem, starting at some base load and ending at an operating point near the voltage stability limit [29]. The power flow Jacobian is singular at the voltage stability limit, which results in convergence issues for solving PF. However, the CPF formulation allows the problem to stay well-conditioned at all possible loading conditions. Homotopy methods are also used to solve the PF problem when the convergence of the problem is dependent on a good initial point, which may be hard to find. It has been shown that standard iterative methods for power flow, such as Newton-Raphson, may diverge due to a poor initial point [30]. Homotopy methods have been shown to improve convergence of the PF problem [31, 32] and used to compute all possible solutions to the PF problem [33].

3.2.2 Homotopy for Optimization

More recently, probability-one homotopy methods have been applied to solving optimization problems. The applications include optimal control [34, 35] and statistical learning [36].

Typically, the homotopy methods in optimization focus on parametrizing the KKT conditions [26, 37] or the objective function [38, 39]. Our method is similar to the homotopy optimization method described in [38], wherein a series of local minimization problems are solved, rather than tracing a path of zeros to the KKT conditions. However, we will focus on a more generalized theoretical analysis of homotopy, allowing for a homotopy map on the set of constraints.

While convergence to a global minimum with probability one is guaranteed for a convex optimization problem [38], this is generally not true for nonconvex problems. In order to understand when homotopy can be effective in finding a global solution for nonconvex optimization, we explore a minimization problem of the form: $\min_x f(x)$ where $f : \mathbb{R}^n \to \mathbb{R}$ is a nonconvex function of $x \in \mathbb{R}^n$. This problem is named (P^o). Note that the function $f(\cdot)$ can incorporate exact/inexact penalty functions to enforce constraints on x, implying that this formulation is general for both unconstrained and constrained optimization [40]. We refer to (P^o) as the "base case" problem. A deformed version of the base case, which is also a nonconvex minimization problem, is denoted by (P^f) and defined as $\min_x \tilde{f}(x)$. For our application, (P^o) corresponds to the *base*-OPF problem and (P^f) corresponds to the *contingency*-OPF problem (the definition of these two problems are provided in the next section). We consider two possible methods for solving the deformed problem that are based on local search algorithms:

1. *One-shot method:* Use the solution of P^o as the initial point for any descent numerical algorithm to solve P^f.

2. *Homotopy method:* Generate a (discretized) homotopy map from P^o to P^f. Use the solution of P^o as the initial point, but update it at each step of the homotopy by solving an intermediate problem using local search that is initialized at the solution of the previous step. A linear (un-discretized) homotopy map can be defined as:

$$P(\lambda) \triangleq \min_x \left\{ \lambda \tilde{f}(x) + (1-\lambda) f(x) \right\}, \quad 0 \leq \lambda \leq 1 \tag{3.1}$$

   with the property that $P(0) = P^o$ and $P(1) = P^f$.

Depending on $f(x)$ and $\tilde{f}(x)$, homotopy may or may not lead to better results than solving the deformed problem in one shot. In Figure 3.1, we see an example where homotopy is effective to find the global minimum of a deformed problem and another example where homotopy leads to a non-global local minimum whereas solving the problem in one shot leads to the global minimum. Knowing when homotopy will be effective is highly dependent on understanding how the shape of the function changes from the base case to the deformed problem. In the current literature, there is a lack of theoretical results to characterize the performance of homotopy in finding a global optimum. While [38] presents algorithms that make use of homotopy to solve nonconvex, unconstrained minimization problems, these algorithms are similar to other stochastic search methods in that they do not guarantee convergence to the global minimum.

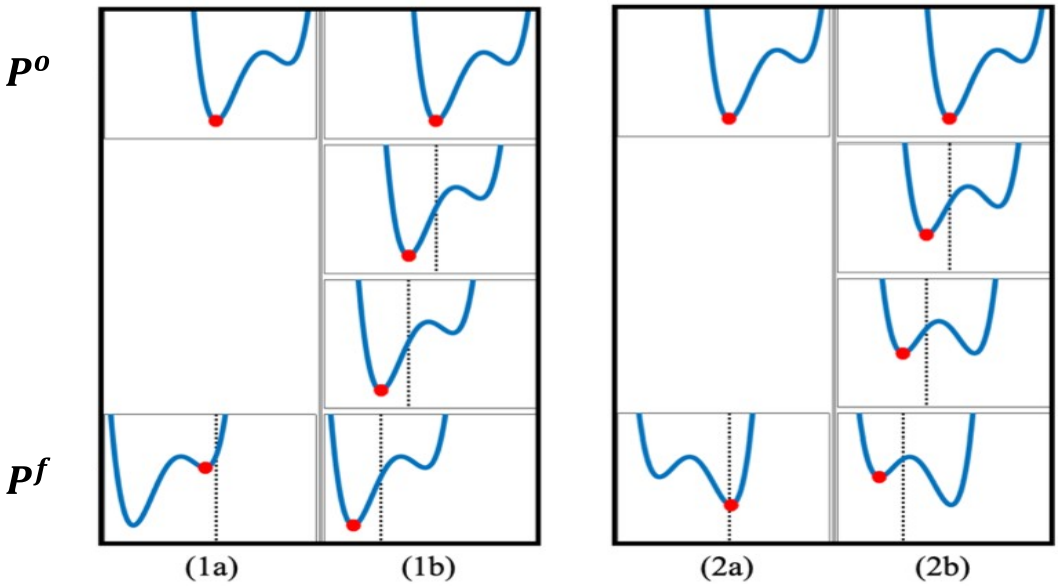

Figure 3.1: Evaluating the performance of homotopy on one-dimensional unconstrained minimization problems. The figure compares two different problems (1) and (2), with two different methods (a) and (b). The dotted lines show how the solution from the previous iteration is used in local search algorithms to solve the next problem. The red dots show the solution at each iteration using the position of the dotted lines as the initial point. For the one-shot method (a), the result of P^o is used as the initial point for P^f. For the homotopy method (b), the base problem P^o is gradually transformed to P^f over three iterations, updating the initial point as the solution to the previous problem.

3.3 Formulation of *base*-OPF and *contingency*-OPF Problems

In this section, we present the mathematical formulations for the *base*-OPF with security constraints and the *contingency*-OPF. The *base*-OPF resembles the conventional SCOPF that finds a base case operational point which is robust against potential contingencies (as described in Section 3.1). The *contingency*-OPF focuses on a single contingency and attempts to find an adjusted operating point that minimizes constraint violations.

To begin, let the power network be defined by a graph $\mathcal{N}(\mathcal{V}, \mathcal{E})$ with the set of generators $\mathcal{G}$, where $\mathcal{V}$ and $\mathcal{E}$ are the vertex set and the edge set of this graph, respectively. The "classic" optimal power flow problem without contingency considerations is a static optimization

problem formulated as:

$$\begin{aligned}
\min_{v\in\mathbb{C}^n} \quad & f(v)+\psi(v) && \text{(3.2a)}\\
\text{s.t.} \quad & p_i^g - \sum_{(i,j)\in\mathcal{E}} p_{ij} = P_i^d, \quad \forall i\in\mathcal{V} && \text{(3.2b)}\\
& q_i^g - \sum_{(i,j)\in\mathcal{E}} q_{ij} = Q_i^d, \quad \forall i\in\mathcal{V} && \text{(3.2c)}\\
& p_{ij} = \mathrm{Re}\{v_i(v_i-v_j)^* Y_{ij}^*\}, \quad \forall (i,j)\in\mathcal{E} && \text{(3.2d)}\\
& q_{ij} = \mathrm{Im}\{v_i(v_i-v_j)^* Y_{ij}^*\}, \quad \forall (i,j)\in\mathcal{E} && \text{(3.2e)}
\end{aligned}$$

where $f(\cdot)$ represents the operating cost (usually a quadratic function of the active power generations) and $\psi(\cdot)$ represents the exact penalty or inexact penalty function that forces the variables to stay within the feasible set defined by:

$$\Psi \triangleq \left\{ v \;\middle|\; \begin{array}{ll} \underline{P_i^g}\le p_i^g \le \overline{P_i^g}, & \forall i\in\mathcal{G} \\ \underline{Q_i^g}\le q_i^g \le \overline{Q_i^g}, & \forall i\in\mathcal{G} \\ \underline{V_i}\le |v_i| \le \overline{V_i}, & \forall i\in\mathcal{V} \\ |p_{ij}+\mathbf{j}q_{ij}|\le \overline{S_{ij}}, & \forall (i,j)\in\mathcal{E} \end{array} \right\} \tag{3.3}$$

In this problem, the decision variable v represents the vector of complex voltages of the power system, and v_i is the voltage at the i-th bus. Furthermore, $p_i^g, q_i^g, p_{ij}, q_{ij}, P_i^d$ and Q_i^d are the active/reactive power generation at the i-th bus, active/reactive power flow from bus i to j, and active/reactive power demand at bus i, respectively. $Y_{ij} = G_{ij} + \mathbf{j}B_{ij}$ is the line admittance, whose real and imaginary parts are the line conductance and susceptance, respectively. The constraints model technical limits, such as the power flow equations, bounds on voltage magnitudes, and bounds on power generations and flows. Nonlinearities are introduced to the constraints with the AC power flow equations, and these nonlinearities with the voltage magnitude lower bounds result in the nonconvexity of the problem. In a standardized optimization form, the "classic" OPF problem can be expressed in a compact form as follows:

$$\begin{aligned}
\min_{v\in\mathbb{C}^n} \quad & f(v)+\psi(v) \\
\text{subject to:} \quad & h(v) = 0
\end{aligned} \tag{3.4}$$

where we note that $h(\cdot)$ is a vector.

3.3.1 SCOPF as the *base*-OPF

Now, suppose that there is a set of possible contingencies, namely $\mathcal{K}$, where each contingency corresponds to a line or generator outage. Each contingency $k\in\mathcal{K}$ introduces a new set of voltage variables v^k, and therefore, for a network with $|\mathcal{V}|$ buses and $|\mathcal{K}|$ contingencies, the SCOPF problem will involve optimizing over $|\mathcal{V}|(|\mathcal{K}|+1)$ scalar complex voltage variables.

The contingencies also add operational constraints of their own. In addition, there are physical limitations on how the post-contingency network can adapt from the base case, and these limits are added as constraints that are functions of the base case voltages.

However, since this extremely high-dimensional problem is cumbersome to solve, in practice the contingency constraints are approximated via methods such as LODF and PTDF [41]. In essence, this approximates the contingency voltage v^k as a function of the base case voltage v. Therefore, post-contingency operating constraints for contingency k are approximated by a composite function of the form $h_k(v) \triangleq c_k(a_k(v))$, where $a_k(v)$ represents the control actions that are taken in the event of a contingency.

Finally, another important consideration is how SCOPF performs when the problem is infeasible. In other words, the SCOPF modeling should be flexible enough to return a best possible solution when all of the physical constraints cannot be met simultaneously. Therefore, we model some operational limits using soft constraints with extra variables that capture the amount of violation. The objective function that is minimized is the sum of active power generation costs in the base case as well as a weighted sum of constraint violation penalties in the base case and contingencies. The standard optimization form is presented below:

$$\big[\textit{base}\text{-OPF}\big] \qquad \begin{aligned} \min_{v,\sigma,\sigma_k} \quad & f(v) + \psi(v) + \phi(\sigma) + \sum_{k=1}^{|\mathcal{K}|} \phi_k(\sigma_k) \\ \text{s.t.} \quad & h(v) = \sigma \\ & h_k(v) = \sigma_k, \quad \forall\ k = 1, \ldots, |\mathcal{K}| \end{aligned} \tag{3.5}$$

where $\phi(\cdot)$ and $\phi_k(\cdot)$ represent the penalty functions for the violations. With no loss of generality, we focus on the case when $\phi(\sigma^p, \sigma^q) = \sum_i \{c_i^p(\sigma_i^p)^2 + c_i^q(\sigma_i^q)^2\}$, where c_i^p and c_i^q are cost coefficients. We denote this SCOPF problem as the *base*-OPF, distinguishing it from the *contingency*-OPF presented next.

3.3.2 Formulating the *contingency*-OPF

Recall that the *base*-OPF solves for the base case operating point by taking into account the possible failures in the network. In the process, it approximates the relationship between the contingency operation point v^k and the base case operating point v. However, it does not actually solve for the optimal v^k's. Therefore, for each contingency we propose to solve a *contingency*-OPF formulated below to find the best operating point for the specific contingency scenario, given the base case solution. This problem resembles the classic OPF problem except that there are additional coupling constraints that tie the problem to the original base case. For instance, the voltage magnitude at a bus must be equal to its base case value unless the reactive capacity of the generators at that bus is exhausted. From an operational standpoint, the relation between the *base*-OPF problem and the *contingency*-OPF problems is shown in Figure 3.2.

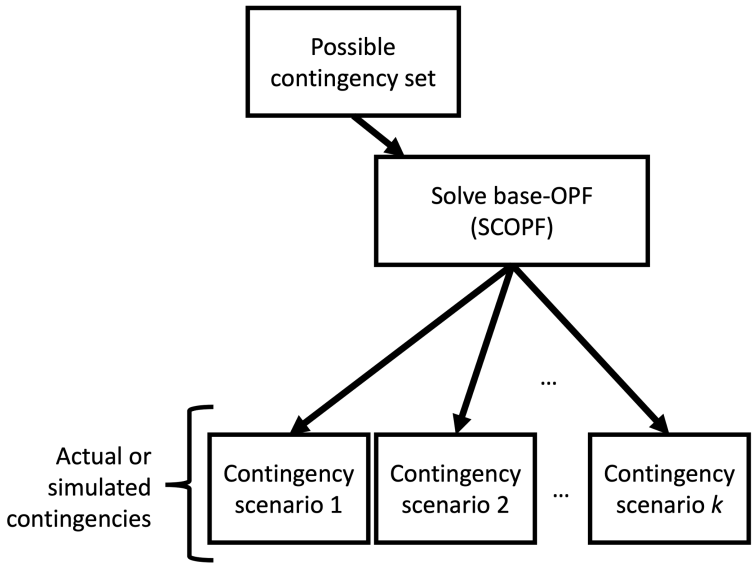

Figure 3.2: Relationship between the SCOPF (referred to as the *base*-OPF) and *contingency*-OPF problems. Each contingency scenario corresponds to solving the *contingency*-OPF given in (3.7).

We model a contingency, such as a line or generator outage, by changing the system parameters from their base values. For example, a line outage physically means that power cannot flow over that connection, which can be modeled by setting the impedance of the line to infinity (or equivalently its admittance to zero). In the event of a line outage, the power is re-routed through other paths and therefore the amount of loss in the system changes. However, the difference in loss is small enough such that there is no need for additional participation from other generators, unlike in the scenario of a generator outage. Therefore, we fix the real power generation to be equal to the base case values and solve for the remaining variables so that the violations for the bus balance equations are small and distributed across the network as much as possible (note that the proposed method can handle generator participation, which is explained in Section 3.4.2). This is because a concentrated violation in a few buses can result in serious issues for the power network, whereas small power mismatches can be dealt with by real-time feedback controllers. Taking these into consideration, the *contingency*-OPF under study is given as:

$$\min_{v,\sigma^p,\sigma^q} \quad \phi(\sigma^p,\sigma^q)+\psi(v) \tag{3.6a}$$

$$\text{subject to:} \quad P_i^g - \sum_{(i,j)\in\mathcal{E}} p_{ij} = P_i^d + \sigma_i^p \qquad \forall i\in\mathcal{V} \tag{3.6b}$$

$$q_i^g - \sum_{(i,j)\in\mathcal{E}} q_{ij} = Q_i^d + \sigma_i^q \qquad \forall i\in\mathcal{V} \tag{3.6c}$$

$$p_{ij} = \mathrm{Re}\{v_i(v_i - v_j)^* \widetilde{Y}_{ij}^*\} \qquad \forall (i,j) \in \mathcal{E} \tag{3.6d}$$

$$q_{ij} = \mathrm{Im}\{v_i(v_i - v_j)^* \widetilde{Y}_{ij}^*\} \qquad \forall (i,j) \in \mathcal{E} \tag{3.6e}$$

$$|v_i| = |v_i|^{base} \qquad \forall i \in \mathcal{V} \setminus \mathcal{V}^q \tag{3.6f}$$

where $\psi(\cdot)$ represents a exact/inexact penalty function that forces the variables to stay within the feasible set defined by Ψ in (3.3). The set $\mathcal{V}^q$ is the set of buses that hit their upper or lower reactive power generation bounds in the base case, and $|v_i|^{base}$ $\forall i \in \mathcal{V}$ is the voltage magnitude of bus i in the base case. The notation $\widetilde{Y}_{ij}$ reflects the potential change in the admittance matrix from the base case value. Note that real power generation is now a fixed parameter obtained from a solution of the *base*-OPF and therefore has been denoted by capital P^g.

For generator outage contingencies, there is an additional aspect to consider. A generator outage corresponds to setting the real power generation at that generator to zero. However, in order to compensate for the lost generation, the system operator needs to increase the power generation at other generators that participate in the outage response. The above framework is general enough to incorporate this difference: simply set $P^g = P^{g,f}$ where $P^{g,f}$ is the new setpoint for the real power generation.

Denoting $x = [v, \sigma^p, \sigma^q]$ as the combined variable, *contingency*-OPF in a standard optimization form would be:

$$[\textit{contingency}\text{-OPF}] \qquad \begin{aligned} \min_{x} \quad & f(x) \\ \text{subject to:} \quad & h(x) = 0 \end{aligned} \tag{3.7}$$

Note that $f(\cdot)$ is the not the same as the objective function used in (3.4) or (3.5) but a comprehensive objective function that includes all the penalty functions. Similarly, $h(\cdot)$ is the not the same as the constraint functions used in (3.4) or (3.5).

If the optimal objective value of the *contingency*-OPF (3.7) is zero, it means that the system is capable of maintaining zero violations by adjusting the parameters from the base case. However, the primary focus of the proposed method is on hard instances with a nonzero optimal cost, meaning that some of the constraints must be violated to accommodate the outage. In these cases, since taking corrective actions to deal with nodal power violations is expensive, it is essential to find a global solution.

3.4 Homotopy for Solving *contingency*-OPF

In the following subsections, we present a homotopy method that parametrizes the *contingency*-OPF to model a gradual line or a generator outage.

3.4.1 Homotopy Method for a Line Outage

In order to solve the *contingency*-OPF problem, we propose a homotopy method that gradually changes certain parameters of the problem from the *base*-OPF, rather than abruptly

changing the structure of the network. For instance, a transmission line outage can be modeled by physically removing the line from the network, by limiting the apparent power flow over the line, or by assigning a low conductance and susceptance value to the line such that effectively no power flows through it. We use the third method to construct a homotopy map for a line outage. The benefit of this method is that it yields better numerical conditioning and is a straightforward method for constructing a homotopy map. For this homotopy method, we solve a series of *contingency*-OPF problems, each with a slightly lower admittance value than the previous problem which uses the solution of the previous problem as an initial point.

Let $\ell \in \mathcal{E}$ be a line that connects bus i and j. Now, consider a contingency scenario in which the line ℓ is out. The active and reactive power over line ℓ can be expressed by the following power flow equations:

$$p_{ij} = \mathrm{Re}\{v_i(v_i - v_j)^* Y_{ij}^*\} \tag{3.8}$$

$$q_{ij} = \mathrm{Im}\{v_i(v_i - v_j)^* Y_{ij}^*\} \tag{3.9}$$

To formalize the line outage contingency, we introduce an aggregate homotopy parameter $\lambda = [\lambda_1, \lambda_2]$ corresponding to the conductance and susceptance where $\lambda_1, \lambda_2 \in \mathbb{R}^{|\mathcal{E}|}$. To be more precise, we parameterize the admittance in *contingency*-OPF as:

$$Y_{ij}(\lambda) = G_{ij}^0 \lambda_{1,\ell} + \mathbf{j} B_{ij}^0 \lambda_{2,\ell}, \qquad \forall (i,j) \in \mathcal{E} \tag{3.10}$$

where G_{ij}^0 and B_{ij}^0 represent the initial admittance of line ℓ. Notice that $\lambda^o = [\mathbf{1}_{|\mathcal{E}|}, \mathbf{1}_{|\mathcal{E}|}]$ corresponds to the original network before the line outage, and $\lambda^f = [\mathbf{1}^\ell_{|\mathcal{E}|}, \mathbf{1}^\ell_{|\mathcal{E}|}]$ corresponds to the post-contingency network after the line outage. Then, the series of homotopy problems, $H(\lambda)$, parametrized by λ can be written in the standard form as:

$$\begin{bmatrix} \textit{homotopy}\text{-OPF} \\ H(\lambda) \end{bmatrix} \qquad \begin{aligned} \min_{x} \quad & f(x, \lambda) \\ \text{subject to:} \quad & h(x, \lambda) = 0 \end{aligned} \tag{3.11}$$

By varying λ from λ^o to λ^f, the homotopy map allows us to create fictitious power networks that constitute a series of intermediate OPF problems. An example flowchart demonstrating the homotopy method for a line outage contingency is given in Figure 3.3.

3.4.2 Homotopy Method for a Generator Outage

A generator outage can also be modeled in a similar way by gradually removing it and adjusting the participation of other generators to compensate for the loss in power. Our proposed homotopy map gradually decreases the real power generation at the generators that are out and gradually increases the real power generation at the generators participating in the contingency response. For the simplicity of presentation, consider contingencies associated with a single generator (generator k) outage. This is common practice in power

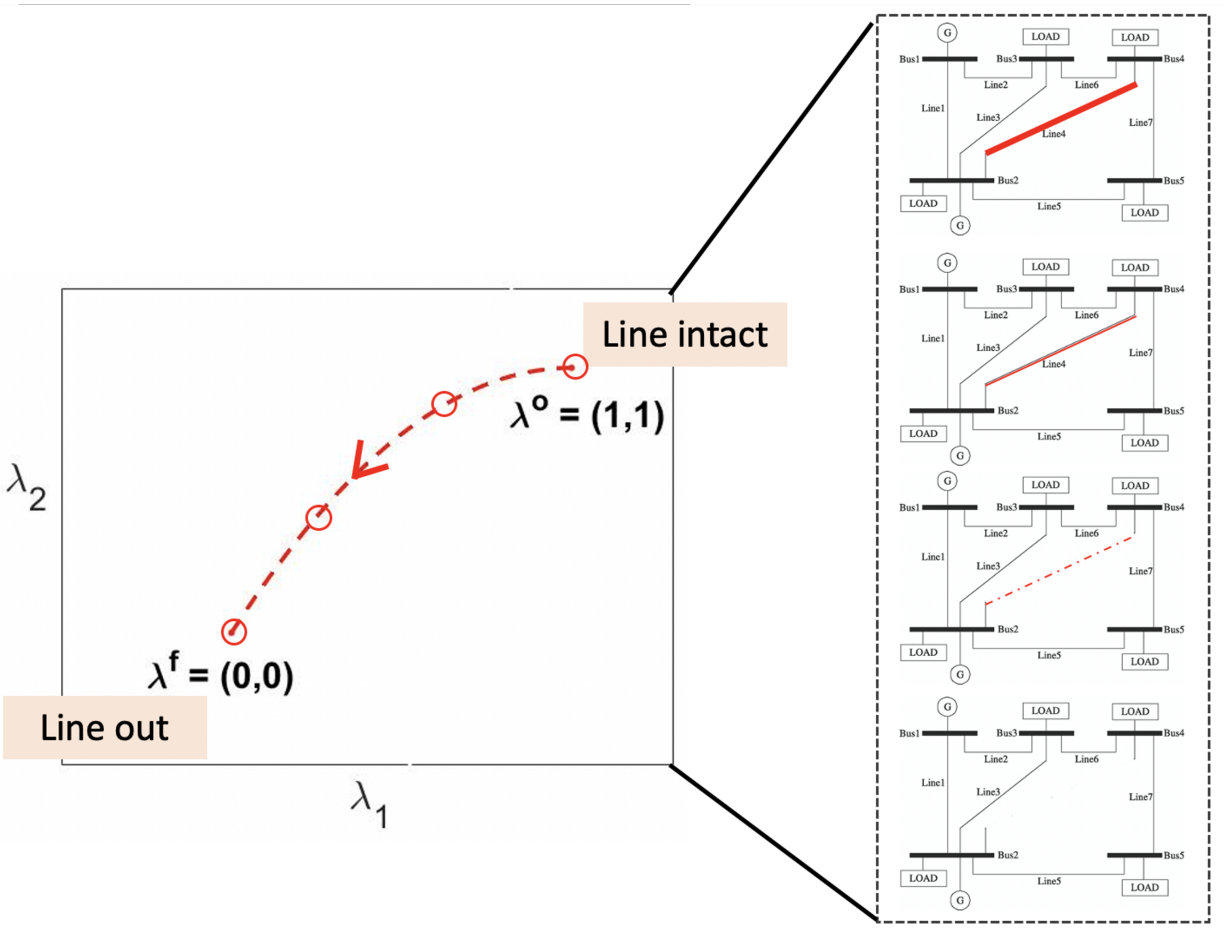

Figure 3.3: An example of a homotopy implementation on a contingency scenario corresponding to a single line (connecting buses 2 and 4) outage. The left plot shows the homotopy path (red dotted curve) and the discretized points (red circles) in the parameter space. The right figure shows how solving the *homotopy*-OPF along the discretized homotopy path affects the system parameters.

systems and is referred to as the $N-1$ criterion. Yet, note that the proposed method can easily be extended to multiple generator outages and is incorporated in Algorithm 2.

Let $P^{g,o} \in \mathbb{R}^{|\mathcal{V}|}$ be the real power generated at all generators in the base case. Using the participation factors of generators that are still active in the contingency, we can compute $P^{g,f} \in \mathbb{R}^{|\mathcal{V}|}$, the real power generated at all generators after the contingency. Since generator k is down in this contingency scenario, $P_k^{g,f} = 0$. One possible method to choose the participation factors that determine $P^{g,f}$ is provided in the Appendix (see Algorithm 3). Similar to the case for line outage contingencies, we introduce an aggregate homotopy parameter $\lambda = [\gamma, \beta]$ with $\gamma, \beta \in \mathbb{R}^{|\mathcal{V}|}$ to create the following homotopy map:

$$P^g(\gamma) = P^{g,o} \odot \gamma + P^{g,f} \odot (\mathbf{1}_{|\mathcal{V}|} - \gamma) \tag{3.12a}$$

$$Q^d(\beta) = Q^{d,o} \odot \beta + Q^{d,f} \odot (\mathbf{1}_{|\mathcal{V}|} - \beta) \tag{3.12b}$$

Focusing on the first equation where we parametrize the real power generation, notice that $\lambda^o = [\mathbf{1}_{|\mathcal{V}|}, \mathbf{1}_{|\mathcal{V}|}]$ corresponds to the original network before the generator outage, and $\lambda^f = [\mathbf{0}_{|\mathcal{V}|}, \mathbf{0}_{|\mathcal{V}|}]$ corresponds to the post-contingency network after the generator outage. By varying λ from λ^o to λ^f, the homotopy map allows us to trace a gradual generator outage. Equation (3.12b) parametrizes the reactive power demand, and we will set the value $Q^{d,f} \simeq Q^{d,o}$. The justification for the extra parametrization of reactive power demand is described in [2] which explains that a parametrization with high enough dimension results in a good homotopy path with high probability. The series of homotopy problems have the same form as those for the line outage, given by Equation (3.11).

3.4.3 Connecting the *contingency*-OPF with the *base*-OPF

Starting with a solution to the *base*-OPF (3.5), we aim to iteratively solve a series of *homotopy*-OPF problems (3.11) to eventually arrive at the *contingency*-OPF problem (3.7). In order to proceed, we assume that the *base*-OPF has a unique global solution that is available (known). The availability of a global solution is a reasonable assumption because a good initial point is usually provided for the *base*-OPF, and also because more time is allocated to solving it compared to a large set of *contingency*-OPF problems for different outages, allowing the use of various convex relaxation techniques.

If the violation cost for $H(\lambda^o)$ is non-zero, the global solution will be unique with overwhelming probability. Furthermore, even if the violation cost for $H(\lambda^o)$ is zero, it will immediately become non-zero during the next homotopy iteration if removing that line introduces inflexibilities that the network cannot accommodate. In fact, these near-infeasible problems where a contingency will make the system "stressed" are the cases where homotopy can be useful and is the focus of this work. Note that the global minimum of the *base*-OPF is also a global minimum of $H(\lambda^o)$ because at $\lambda = \lambda^o$, the parameters of the *homotopy*-OPF correspond to the pre-contingency network, for which the violations are zero.

3.4.4 Implementation of *homotopy*-OPF

Now, we describe how the solution to the *base*-OPF can be used to find the solution to the *contingency*-OPF via a homotopy method. First, a series of *homotopy*-OPF problems is formulated as a parametrization of the *contingency*-OPF problem as described above. To define a physically implementable homotopy map, we must discretize the homotopy parameter λ. Then, the first *homotopy*-OPF problem is initialized as the solution to the *base*-OPF problem. The series of *homotopy*-OPF problems is solved via local search methods, and the initial point is updated at each iteration of homotopy to be the solution of the previous *homotopy*-OPF problem. We define a stepsize that is sufficiently small so that the convergence properties of the continuous homotopy map hold (see [26, 27, 42] for theoretical subtleties). Note that the discretization of the homotopy path can also be represented by the set $\Lambda := \{\Lambda^1, \ldots, \Lambda^T\}$, where $\Lambda^i = \lambda(i)$ for $i = 0, \ldots, T$, $\Lambda^1 = \lambda(0) = \lambda^o$ and $\Lambda^T = \lambda(T) = \lambda^f$. Please refer to Algorithms 1 and 2 for complete details of the method.

By finding a solution to the problem at each step of the homotopy, we are tracing a path of zeros to the KKT equations, assuming constraint qualification holds (see [2] for a complete discussion of the assumptions in our method). The path-tracing becomes difficult if there are bifurcations or turning points along the path. In this chapter, we use a sufficiently small step-size which allows a "jump" over a bifurcation, as discussed in [27, 38].

To develop intuition on when a homotopy method may or may not lead to the global solution, we consider the basin of attraction of the global solution to the *contingency*-OPF problem. The "basin of attraction" of a local solution is the set of initial points that lead to the solution using an iterative search method, and is dependent on both the problem geometry and the choice of algorithm.

At each iteration of homotopy, the problem is initialized as the solution to the previous problem with the hope that the previous point will be in the basin of attraction of the new solution. The homotopy method will find the global solution of the final problem if at some point along the homotopy path, the solution to the intermediate problem enters and stays within the basin of attraction of the global solution to the final problem. Because of this, homotopy is only useful for problems where the initial point, i.e. the solution to *base*-OPF, is not in the basin of attraction of the global minimum of the final problem, *contingency*-OPF. If the initial point is within the basin of attraction of the global solution to the final problem, then the intermediate problems are unnecessary.

Proposition 1. *If the global solution along the homotopy path is unique, then a sufficiently small step-size $\Delta\lambda$ will ensure that the solution to each intermediate problem is a global solution.*

Under the uniqueness assumption mentioned above, the solution to some intermediate problem will enter and remain within the basin of attraction of the global solution to the final problem, and we will obtain the global minimum of the final problem. In the next section, we find some conditions under which the global solution along the path is unique.

Algorithm 1 Homotopy-OPF for Line Outages

Given:

1. Power network $\mathcal{N}(\mathcal{V}, \mathcal{E})$ and generators $\mathcal{G}$
2. Contingency set $\mathcal{K}$ with line outages $L_k \subset \mathcal{E}$ for each $k \in \mathcal{K}$
3. Discretized homotopy path Λ

Initialize: Solve *base*-OPF problem given by Equation (3.5) to find a globally optimal solution $(|v|_*, \theta_*, p^g_*, q^g_*, \{\sigma_{k*}\})$

Formulate the *contingency*-OPF problem in Equation (3.7):

1. Fix real power generation to base case solution: $P^g := p^g_*$
2. Find $\mathcal{V}^q$ based on q^g_*

for $k \in \mathcal{K}$ **do**

Set up *homotopy*-OPF family $H(\Lambda)$ for given line outages L_k.

Initialize $(|\tilde{v}|, \tilde{\theta}, \tilde{q}^g, \tilde{\sigma}^p, \tilde{\sigma}^q) \leftarrow (|v|_*, \theta_*, p^g_*, q^g_*, \{\sigma_{k*}\})$

for $i \in \{0, ..., T\}$ **do**

Solve $H(\Lambda^i)$ using initial point $(|\tilde{v}|, \tilde{\theta}, \tilde{q}^g, \tilde{\sigma}^p, \tilde{\sigma}^q)$, and obtain new solution $(|v|, \theta, q^g, \sigma^p, \sigma^q)$.

Update $(|\tilde{v}|, \tilde{\theta}, \tilde{q}^g, \tilde{\sigma}^p, \tilde{\sigma}^q) \leftarrow (|v|, \theta, q^g, \sigma^p, \sigma^q)$

end for

Return $(|v|, \theta, q^g, \sigma^p, \sigma^q)$ and violation cost $\phi(\sigma^p, \sigma^q)$.

end for

3.5 Analysis of Homotopy Paths

While probability-one homotopy methods almost surely guarantee algorithm convergence, they do not necessarily result in convergence to a global minimum [38]. In Section 3.2, we offered two examples of nonconvex optimization: one in which the homotopy method resulted in the global minimum and another in which the homotopy method resulted in a non-global local minimum (see Figure 3.1). In this section, we describe a theoretical framework that describes when homotopy can be used to obtain a global minimum. We apply this framework to analyze the performance of *homotopy*-OPF in finding the global solution of the *contingency*-OPF. The results developed in this section have implications for homotopy methods in a broad range of optimization problems. See [2] for a complete discussion of the convergence and complexity of the proposed *homotopy*-OPF method.

Remark 1. *To simplify the presentation, we make the assumption that homotopy-OPF has a unique global solution at the initial point of the path. The "uniqueness" of the global solution (in this assumption and Theorem 4 to be presented next) can be replaced by the "connectivity" of the set of all global solutions (this allows having infinitely many possible solutions for contingency-OPF with zero violation cost).*

Algorithm 2 Homotopy-OPF for Generator Outages

Given:

1. Power network $\mathcal{N}(\mathcal{V}, \mathcal{E})$ and generators $\mathcal{G}$
2. Contingency set $\mathcal{K}$ with generator outages $R_k \subset \mathcal{G}$ for each $k \in \mathcal{K}$
3. Discretized homotopy path Λ

Initialize: Solve *base*-OPF problem given by Equation (3.5) to find a globally optimal solution $(|v|_*, \theta_*, p^g_*, q^g_*, \{\sigma_{k*}\})$

for $k \in \mathcal{K}$ **do**

Formulate the *contingency*-OPF problem in Equation (3.7):

1. Define P^g_r as the fixed real power generation at $r \in \mathcal{G}$
2. Define ΔP^g_k as the total lost real power generation at k: $\Delta P^g_k \triangleq \sum_{r \in R_k} p^g_{*,r}$
3. Find $\mathcal{V}^q$.
4. Remove real power generation for generators in R_k: $P^g_r \leftarrow 0 \quad \forall r \in R_k$
5. Compute participation factors α^g_r for $r \in \mathcal{G} \setminus R_k$ (see Algorithm 3 in the Appendix)
6. Add real power generation for participating generators:

for $r \in \mathcal{G} \setminus R_k$ **do**

if $\alpha^g_r > 0$ **then**

$P^g_r \leftarrow \max\{\alpha^g_r \Delta P^g_k, \overline{P^g_r} - p^g_{*,r}\}$

end if

end for

Set up *homotopy*-OPF family $H(\Lambda)$ for given generator outages R_k.

Let $P^{g,o} := p^g_*$ and $P^{g,f} := P^g$

Initialize $(|\tilde{v}|, \tilde{\theta}, \tilde{q}^g, \tilde{\sigma}^p, \tilde{\sigma}^q) \leftarrow (|v|_*, \theta_*, p^g_*, q^g_*, \{\sigma_{k*}\})$

for $i \in \{0, ..., T\}$ **do**

Solve $H(\Lambda^i)$ using initial point $(|\tilde{v}|, \tilde{\theta}, \tilde{q}^g, \tilde{\sigma}^p, \tilde{\sigma}^q)$ and obtain new solution $(|v|, \theta, q^g, \sigma^p, \sigma^q)$

Update $(|\tilde{v}|, \tilde{\theta}, \tilde{q}^g, \tilde{\sigma}^p, \tilde{\sigma}^q) \leftarrow (|v|, \theta, q^g, \sigma^p, \sigma^q)$

end for

Return $(|v|, \theta, q^g, \sigma^p, \sigma^q)$ and violation cost $\phi(\sigma^p, \sigma^q)$

end for

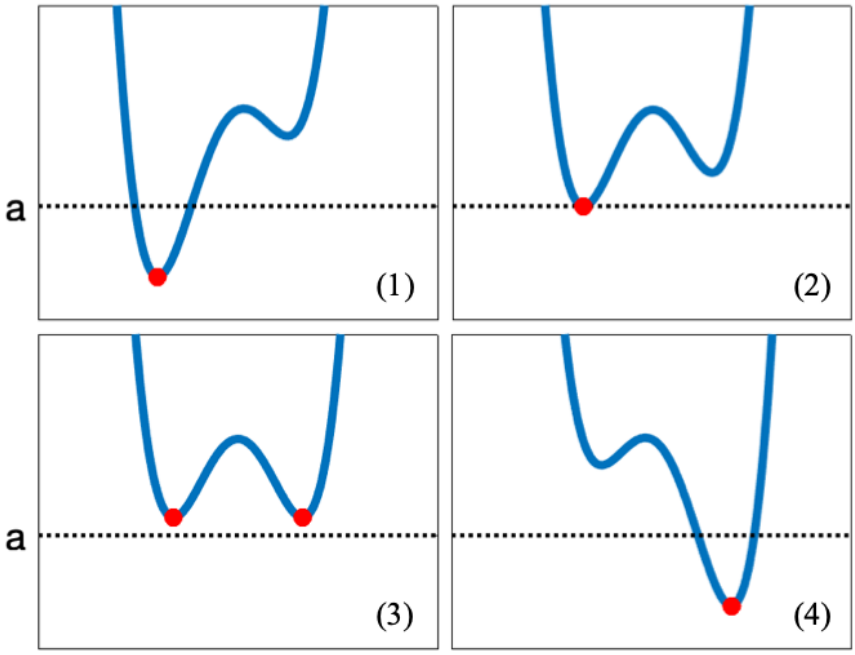

Figure 3.4: Red dots denote the global min of the functions. Homotopy may not be effective for cases where the global minimum of the base case becomes a non-global local minimum of the deformed problem. For any problem where a global minimum for the initial problem is transformed into a non-global local solution for the final problem, by continuity, there must exist a point along the deformation path where the problem has two global minima. In this example, continuously changing from the initial curve (1) to the final curve (4) requires passing through curves (3) with two global minima. Point a is defined in Definition 1.

3.5.1 Characterization of Desirable Homotopy Path

The path we take to change the homotopy parameter λ defines how the constraints and objective function of $H(\lambda)$ will change, and this in turn affects the series of global solutions obtained throughout the homotopy process. Therefore, choosing a good homotopy path is directly correlated with the success of the method. Note that even though Algorithms 1 and 2 work on a discretized homotopy path, their analyses require working on the continuous path. In Figure 3.4, we have presented an example in which homotopy fails to find the global solution of the final problem. The major cause of this breakdown is the emergence of two global solutions in the (continuous) homotopy path, which is followed by a change in the relative positions of the global solution and next best local solution. In order to better characterize this, consider the KKT conditions for the *homotopy*-OPF problem defined in

(3.11):

$$\begin{aligned} \nabla f(x,\lambda) + \nu^T \nabla h(x,\lambda) &= 0 \\ h(x,\lambda) &= 0 \end{aligned} \tag{3.13}$$

where ν is the vector of dual variables. We assume that constraint qualifications hold for the problem $H(\lambda)$ defined in (3.11) for all $\lambda \in \Lambda$, which implies the KKT conditions are satisfied for every local minimum. For a given λ, let $\mathcal{X}(\lambda)$ be the set of all x that satisfy the KKT conditions in Equation (3.13). Note that the goal is not to solve the KKT conditions directly but is to merely use them as a necessary condition for all local solutions. Before proceeding to a main theorem of this work, below we make one basic assumption on the KKT conditions and define a concept called the "dividing midpoint zone" (DMZ).

Assumption 1. *The cardinality of set $\mathcal{X}(\lambda)$ as a function of λ is finite and constant for all $\lambda \in \Lambda$.*

The first part of Assumption 1 is essential to guarantee that a local solution does not suddenly appear or disappear along the homotopy path, in which case we cannot trace it back to the local solutions of the original problem to track it. Using techniques in algebraic geometry, one can study the satisfaction of the first part of Assumption 1 [43]. The second part of Assumption 1 is to make sure that we can apply the implicit function theorem. In [44], it is shown that this is generically true under mild regularity conditions. Note that if this assumption is violated, it is very hard to make any conclusions about the performance of homotopy, since we may have the emergence of many local minima with small perturbations in λ.

Definition 1. *At λ^o, we order all the elements in $\mathcal{X}(\lambda^o)$ in a way such that $f(x^{(1)},\lambda^o) < f(x^{(2)},\lambda^o) \leq \cdots \leq f(x^{(|\mathcal{X}|)},\lambda^o)$. Furthermore, let a be the midpoint objective value of the first and second best KKT points. In other words,*

$$a = \frac{f(x^{(1)},\lambda^o) + f(x^{(2)},\lambda^o)}{2} \tag{3.14}$$

Define $\mathcal{S}$ to be the set of all λ for which there exists a KKT point with the objective value equal to a:

$$\mathcal{S} \triangleq \{\lambda \in \Lambda \mid \quad f(x,\lambda) = a \text{ for some } x \in \mathcal{X}(\lambda)\} \tag{3.15}$$

Here, we define a to be the "dividing midpoint" between $f(x^{(1)},\lambda^o)$ and $f(x^{(2)},\lambda^o)$. In practice, a wide range of values that are slightly above or below the point a, within the DMZ, would lead to the same implications. The optimal choice of a depends on the knowledge of how the shape of the curve changes with respect to λ. We are now ready to state the first theorem.

Theorem 4. *Let $\rho(\lambda) = 0$ be a homotopy path of λ with two end-points λ^o and λ^f. In other words, the set of λ's satisfying $\rho(\lambda) = 0$ can be parametrized by $t \in [0, T]$ such that $\lambda(0) = \lambda^o$ and $\lambda(T) = \lambda^f$. If $\rho(\lambda) = 0$ does not intersect with the set $\mathcal{S}$, then the homotopy problem (3.11) has a unique global minimum for all values of λ along the path $\rho(\lambda) = 0$.*

Proof. The proof is provided in Appendix 3.B. □

According to Theorem 4, the success of homotopy in finding the global optimum depends on the geometry of the set $\mathcal{S}$. To illustrate this, suppose that the set $\mathcal{S}$ is described by the blue area in Figure 3.5. We wish to design a homotopy method that starts from $\lambda^o = (1, 1)$ and ends at $\lambda^f = (0, 0)$. However, in some cases, this may not be possible without crossing the set $\mathcal{S}$. This could be because set $\mathcal{S}$ fully encompasses the final λ^f and blocks any path from entering, such as in case (a) of Figure 3.5.

Directly analyzing the geometry of the set $\mathcal{S}$ is challenging. Therefore, we introduce a method to certify whether a path is a successful homotopy path or not. The following theorem offers a dual certificate.

Theorem 5. *Let $\rho(\lambda) = 0$ define the homotopy path used to solve the homotopy-OPF problem (3.11). Consider the following feasibility problem and denote it by (P) :*

$$
\begin{aligned}
(P) \quad p(x^*, \lambda^*, \mu^*) = \min_{x,\lambda,\mu} \quad & 0 && \text{(3.16a)}\\
s.t. \quad & \nabla f(x, \lambda) + \mu^T \nabla h(x, \lambda) = 0 && \text{(3.16b)}\\
& h(x, \lambda) = 0 && \text{(3.16c)}\\
& f(x, \lambda) = a && \text{(3.16d)}\\
& \rho(\lambda) = 0 && \text{(3.16e)}
\end{aligned}
$$

Let the corresponding dual problem be denoted by (D), written as:

$$\max_{\omega_1,\omega_2,\omega_3,\omega_4} \; d(\omega_1, \omega_2, \omega_3, \omega_4) \tag{3.17}$$

where $\omega_1, \omega_2, \omega_3, \omega_4$ are the dual variables for the constraints (3.16b), (3.16c), (3.16d) and (3.16e). If there exists a quadruplet $(\omega_1, \omega_2, \omega_3, \omega_4)$ such that $d(\omega_1, \omega_2, \omega_3, \omega_4) > 0$, then the homotopy-OPF problem (3.11) attains a unique global minimum along the path $\rho(\lambda) = 0$.

Proof. The proof is provided in Appendix 3.C. □

Note that the dual problem is convex and finding it is easy for certain problems, for example in the case where *homotopy*-OPF is cast as a nonconvex quadratically-constrained quadratic program. In essence, finding a dual feasible point $(\omega_1, \omega_2, \omega_3, \omega_4)$ for which $d(\omega_1, \omega_2, \omega_3, \omega_4) > 0$ provides a certificate that guarantees that the homotopy path $\rho(\lambda)$ will never intersect with set $\mathcal{S}$. Then, by Theorem 4, we can conclude that the homotopy method will have a unique global minimum along its path and therefore Algorithms 1 and 2 are able to solve *contingency*-OPF to global optimality using iterative local search due to

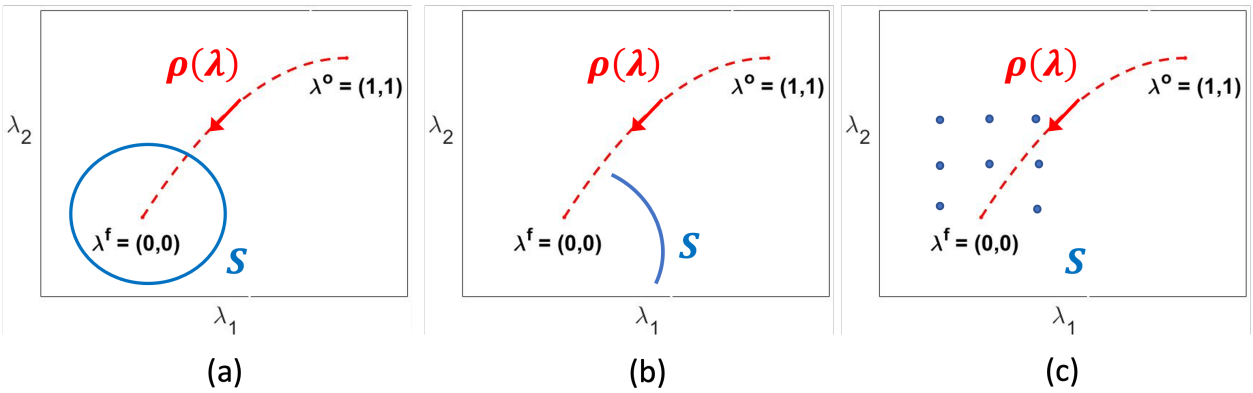

Figure 3.5: Examples of sets $\mathcal{S}$ (blue) and homotopy paths $\rho(\lambda)$ (red). In Figure (a), we see an example where it is impossible to design a homotopy path that reaches the origin without intersecting $\mathcal{S}$. In Figure (b), we see an example where homotopy will work with a carefully designed path that avoids set $\mathcal{S}$. In Figure (c), we see an example where any homotopy path with probability 1 will avoid set $\mathcal{S}$.

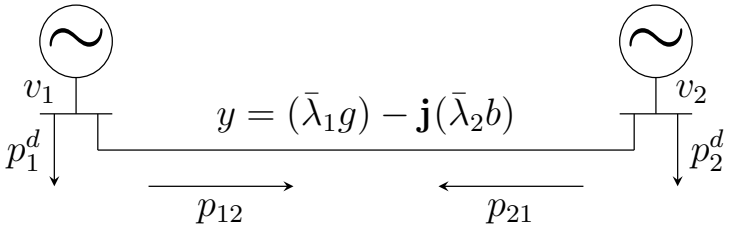

Figure 3.6: Diagram of the two-bus network

Proposition 1. Note that the results of Theorem 4 are still valid if one adds valid inequalities to (P) to increase the likelihood of the existence of a desirable dual feasible point via reducing the duality gap [45].

3.5.2 Geometry of the Homotopy Path: Two-bus Example

Consider a simple two-bus example as shown in Figure 3.6. Each bus has a corresponding voltage magnitude and voltage angle associated with it. The voltage magnitude of bus i is denoted by $|v_i|$ and the voltage angle of bus i is denoted by θ_i. The line connecting the two buses have admittance $y = (\bar{\lambda}_1 g) - \mathbf{j}(\bar{\lambda}_2 b)$. The active power injection and demand at bus i are denoted by p_i^{inj} and $P_i^d > 0$, respectively. Furthermore, there is a lower bound Q^{min} on reactive power injection at both buses. Assume that (1) $|v_1| = |v_2| = 1$, (2) $-\Delta \leq \theta_1 - \theta_2 \leq \Delta$ and (3) $0 < Q^{min} < q(\Delta)$, where $\Delta = \tan^{-1}(\lambda_2 b / \lambda_1 g)$ and $q(\cdot)$ denotes the reactive power

injection as a function of the solely the angle difference, which is due to the fact that voltage magnitudes are fixed. Note that the second constraint on angle difference is reasonable for the secure operation of power systems and is also used in [46] in order to restrict the two-bus active power injection region to be within the Pareto front of the original feasible region. In mathematical terms, suppose that the corresponding OPF problem takes the following form:

$$\begin{aligned} \min_{p_1^{inj}, p_2^{inj}} \quad & (p_1^{inj} + P_1^d)^2 + c(p_2^{inj} + P_2^d)^2 \\ \text{subject to:} \quad & h(p_1^{inj}, p_2^{inj}) = 0 \end{aligned} \tag{3.18}$$

The feasible set of the two-bus injection region belongs to the Pareto front of an ellipse, which is partially removed due to the reactive power constraints (the details can be found in [46]). The following lemma characterizes the set of homotopy parameters for which there are at least two global solutions.

Lemma 2. *Denote* $\alpha = cos^{-1}(\frac{-Q^{min}+b\lambda_2}{|y|})$*, and define two polynomial functions of* $\lambda = (\lambda_1, \lambda_2)$ *as follows:*

$$w_1(\lambda_1, \lambda_2) \triangleq \frac{2\lambda_2 b}{|y|}\big(\lambda_2 b \cdot \sin\alpha + \alpha \cdot \lambda_1 g\big) \tag{3.19a}$$

$$w_2(\lambda_1, \lambda_2) \triangleq 2\lambda_1 g - \frac{2\lambda_1 g}{|y|}\big(- \lambda_1 g \cdot \sin\alpha + \alpha \cdot \lambda_2 b\big) \tag{3.19b}$$

Define also the set $\widetilde{S}$ *as:*

$$\widetilde{\mathcal{S}} \triangleq \{\lambda \in \mathbb{R}^2 \mid (1-c)\cdot w_1(\lambda_1, \lambda_2)\cdot w_2(\lambda_1, \lambda_2) + 2P_1^d \cdot w_1(\lambda_1, \lambda_2) - 2cP_2^d \cdot w_1(\lambda_1, \lambda_2) = 0\} \tag{3.20}$$

If $(\bar{\lambda}_1, \bar{\lambda}_2) \in \widetilde{\mathcal{S}}$*, then the two-bus OPF problem has two global solutions.*

Proof. The proof is provided in Appendix 3.D. □

We can view this set $\widetilde{\mathcal{S}}$ as an equivalent if not a subset of $\mathcal{S}$. This is a set of measure zero in general and as long as the homotopy path does not intersect with this set, our homotopy algorithms will work.

The set $\widetilde{S}$ for the two-bus network is depicted in Figure 3.7. In this case, we can see that there is no homotopy path $\rho(\lambda)$ moving from $\lambda^o = (1, 1)$ to $\lambda^f = (0, 0)$ that does not cross this set and therefore cannot guarantee that the homotopy method will always converge to the global solution.

3.6 Simulations

In this section, we illustrate the success of the homotopy method in finding the global solution of the *contingency*-OPF. In doing so, we present simulations of different line and generator outage scenarios on various networks. We also evaluate the performance using different homotopy paths and discretizations, and verify our earlier theoretical results.

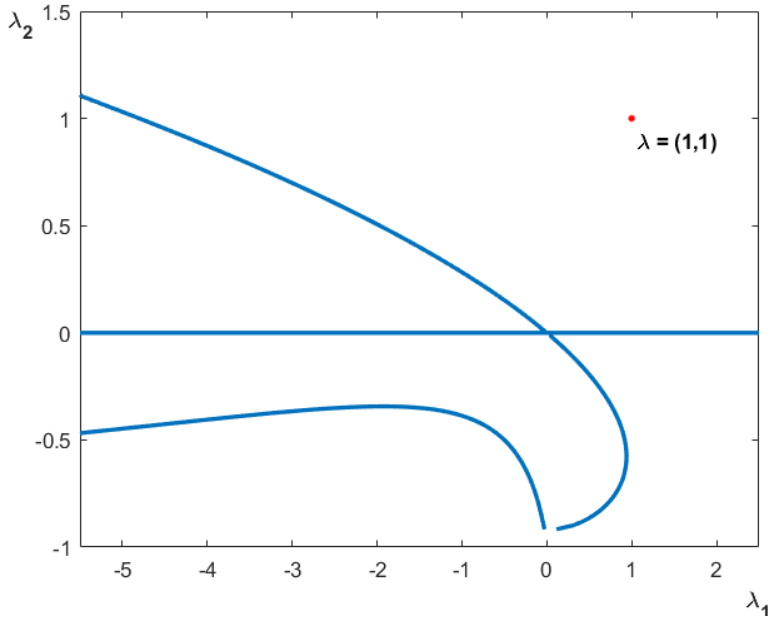

Figure 3.7: An example of set $\widetilde{\mathcal{S}}$ for the two-bus network

3.6.1 Simulation Setup

These simulations are all run in MATLAB on a standard laptop (2.6 GHz 6-Core Intel Core i7 with 16 GB 2400MHz RAM). The *contingency*-OPF problems with and without homotopy all solve in less than two minutes and typically solve on the order of seconds. With the given machine configuration, we were able to solve six *homotopy*-OPF problems in parallel in less than three minutes. The typical time frame for solving OPF problems in practice is 5 to 15 minutes. The applicability of our methodology will depend on how much parallel-computing resources are available to the user.

In these simulations, we consider $N-1$ contingencies wherein there is one line or generator out as well as $N-2$ and $N-3$ contingencies wherein there are multiple outages. Although $N-1$ contingencies occur more frequently in practice, $N-2$ and $N-3$ contingencies are catastrophic events that are worth considering as they are harder to correct. Extreme weather events, attacks, or component aging could cause these $N-k$ (where $k \geq 2$) contingency scenarios to occur [47]. Adding uncertain renewable energy sources such as wind energy to power networks increases the probability of correlated faults and thus the possibility of $N-2$ and $N-3$ contingencies [48]. Additionally, these multi-contingency scenarios can capture cascading failures that occur in a short window where corrective action is not possible between contingencies [48].

In order to implement the *contingency*-OPF within the MATPOWER format [49], we introduce virtual generators that model the violations of real and reactive power balance

equations (σ^p and σ^q). Virtual generators are modeled so that they only generate or consume (virtual) power when there is a nonzero violation in the respective power balance equation. Therefore, by penalizing the virtual generation in the modified objective function, we fully implement the *contingency*-OPF as formulated in Section 3.3.2. The benefit of this formulation is that there always exists a feasible solution to *contingency*-OPF. By adding power generation flexibility with virtual generators, we aim to find a feasible point (equivalent to a zero objective value) or an infeasible point for the network but with the minimum violations (such solutions could still be implemented via corrective actions taken by real-time feedback controllers). To solve each of the homotopy simulations, we use the MATPOWER Interior Point Solver (MIPS) [50].

For both line and generator outages, we solve the corresponding *contingency*-OPF problems via both homotopy and the one-shot method. The one-shot method uses the solution for the *base*-OPF as the initial point for directly solving the *contingency*-OPF. We compare various homotopy discretization schemes to the one-shot method. Note that the one-shot method is equivalent to solving the *contingency*-OPF problem via interior point methods and thus represents the current state-of-the-art.

3.6.2 Simulated Line Outages

For the line outages, we consider three different homotopy paths. If we take line ℓ connecting buses i and j to be out, then the three homotopy paths are given by:

- Scheme 1: Uniformly decrease $(\lambda_{1,\ell}, \lambda_{2,\ell})$ from $(1, 1) \to (0, 0)$
- Scheme 2: Decrease $\lambda_{1,\ell}$ from $1 \to 0$, then $\lambda_{2,\ell}$ from $1 \to 0$
- Scheme 3: Decrease $\lambda_{2,\ell}$ from $1 \to 0$, then $\lambda_{1,\ell}$ from $1 \to 0$

These schemes can be applied to multiple line outages by simultaneously modifying $\lambda_{1,\ell}$ and $\lambda_{2,\ell}$ for each line $\ell \in \mathcal{E}$ that is out. For line outage scenarios on the 3375-bus and 3120-bus Polish networks [49], Figures 3.8 and 3.9 show the evolution of the violation cost over these homotopy schemes (with a 10-iteration discretization) compared to the violation cost of the one-shot method.

Next, we consider changing the discretization of homotopy scheme 1 in a line outage scenario. Figure 3.10 shows line outage scenarios on the 3375-bus and 3120-bus Polish networks [49] using homotopy scheme 1 with a varying number of iterations. For these line outages, we implement Scheme 1 described above that decreases $\lambda_{1,\ell}$ and $\lambda_{2,\ell}$ simultaneously from 1 to 0 at each of the outed lines $\ell \in \mathcal{E}$. We consider three possible discretizations of this homotopy path, using 3, 5, and 10 iterations of homotopy and compare them to the one-shot method.

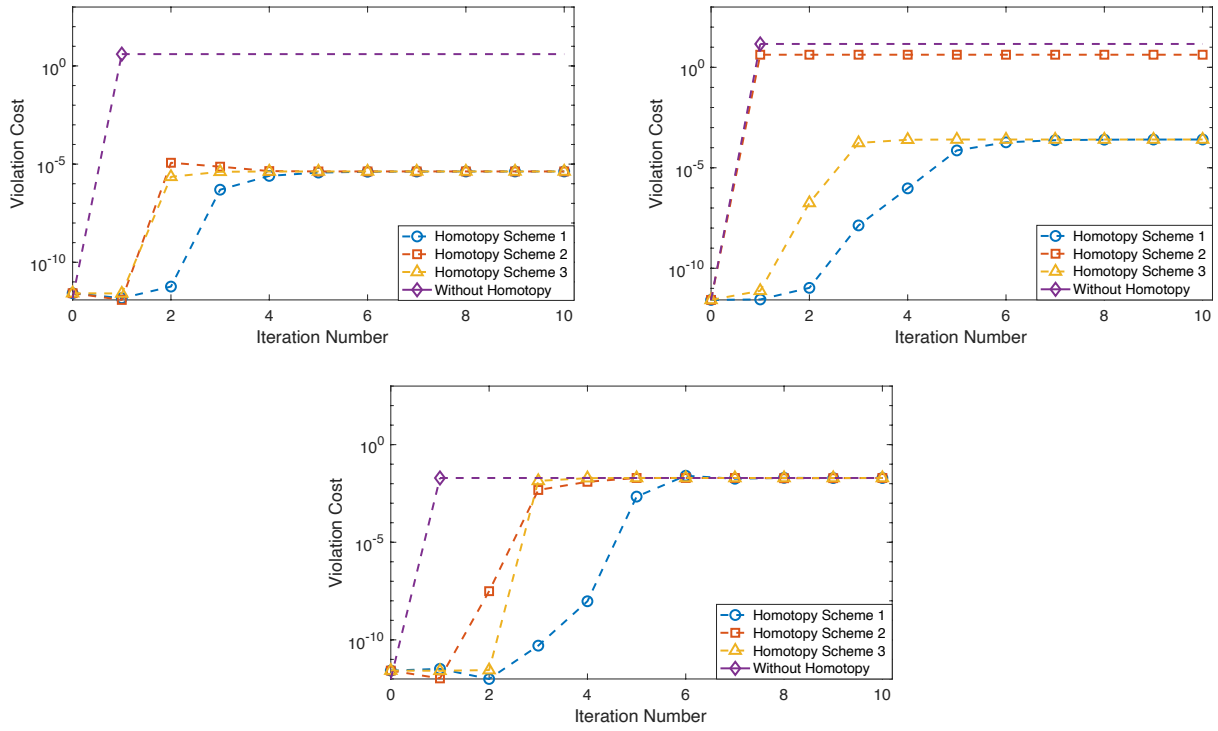

Figure 3.8: Performance of proposed homotopy method on the 3375-bus Polish network (`case3375wp` with real and reactive power demand scaled up by 10%) with single line outages. Homotopy schemes 1 through 3 are tested with 10 iterations. In the top left figure (line out ID: 3596, with about 370 MW flowing from bus 553 to bus 554 in the base case), we have a scenario where all three homotopy schemes outperform the one-shot method. In the top right figure (line out ID: 3551, with about 180 MW flowing from bus 451 to bus 450 in the base case), we have a scenario where only homotopy schemes 1 and 3 significantly outperform the one-shot method. In the bottom figure (line out ID: 268, with about 198 MW flowing from bus 26 to bus 77 in the base case), we have a scenario where the one-shot method performs the same as all three homotopy methods. This was the most common scenario for our experiments on the 3375-bus network, representing about 95% of 4161 tested single line outages. While homotopy in general yielded the same solution as the one-shot method, in scenarios where it outperformed the one-shot method, the results were often better by a factor of 10^5, as seen in the top two figures. Note that all solutions shown here are convergent, compared to the scenarios for the 3120-bus Polish network shown in Figure 3.9 which have some non-convergent solutions.

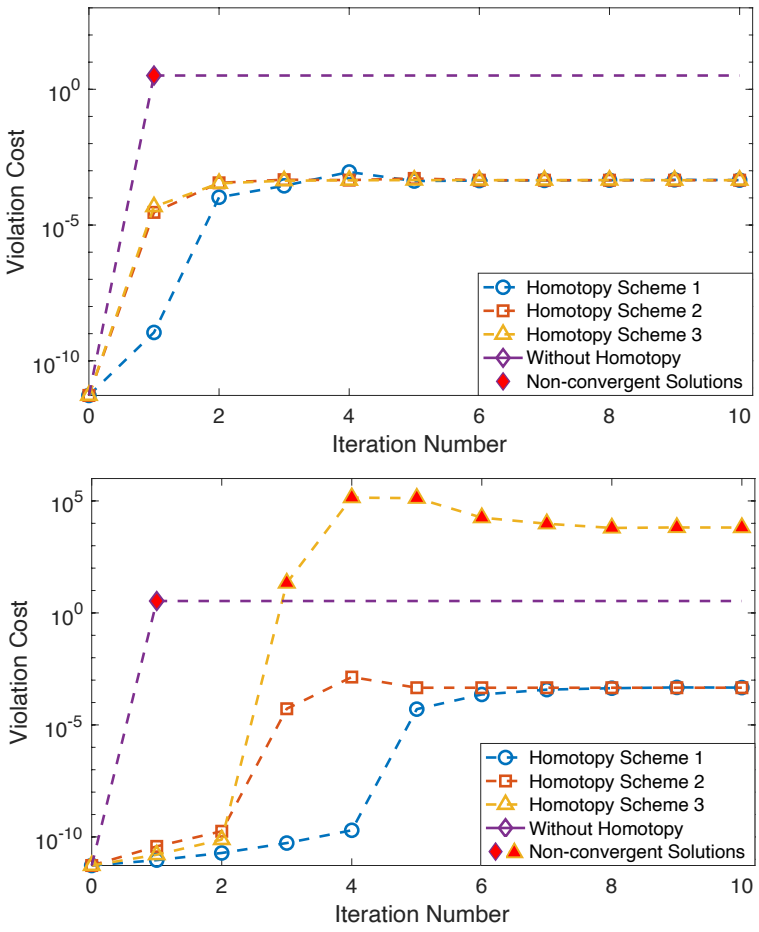

Figure 3.9: Performance of proposed homotopy method on the 3120-bus Polish network (`case3120sp` with real and reactive power demand scaled up by 10%) with multiple line outages. Homotopy schemes 1 through 3 are tested with 10 iterations. By introducing multiple line outages, we make the *contingency*-OPF problem more difficult to solve, which makes it a good candidate for the proposed homotopy method. In both of these scenarios, homotopy schemes 1 and 2 find a convergent solution while the one-shot method does not. In the top figure, the IDs of the outed lines are 438 (with about 67 MW flowing from bus 710 to bus 797 in the base case), 439 (with about 67 MW flowing from bus 797 to bus 578 in the base case), and 3150 (with about 59 MW flowing from bus 578 to bus 577 in the base case). In the bottom figure, the IDs of the outed lines are 2056 (with about 3 MW flowing from bus 1625 to bus 1145 in the base case) and 3082 (with about 51 MW flowing from bus 336 to bus 337 in the base case).

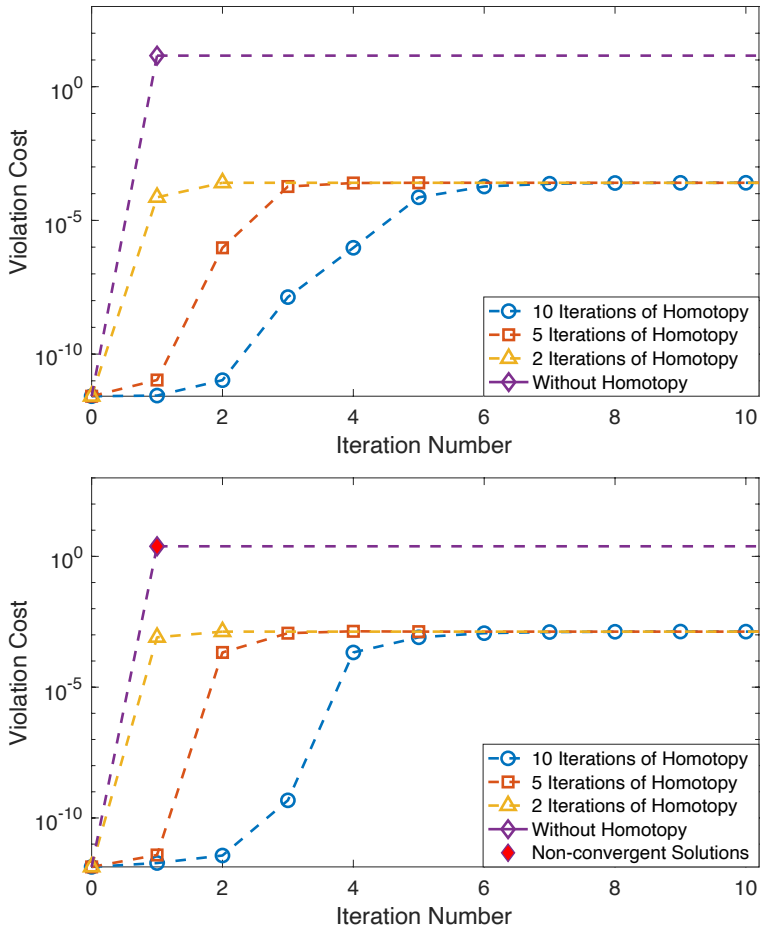

Figure 3.10: Performance of proposed homotopy scheme 1 tested with a varying number of iterations. The top figure shows the 3375-bus Polish network (`case3375wp` with real and reactive power demand scaled up by 10%) with a single line outage (line out ID: 3551, with about 180 MW flowing from bus 451 to bus 450 in the base case). In this scenario, we see that the 2, 5, and 10-iteration homotopy methods converge to a solution that is much better than that obtained by the one-shot method. The bottom figure shows the 3012-bus Polish network (`case3012wp` with real and reactive power demand scaled up by 8%) with a single line outage (line out ID: 1604, with about 68 MW flowing from bus 1590 to bus 1360 in the base case). In this scenario, we see that the 2, 5, and 10-iteration homotopy methods result in a convergent solution while the one-shot method does not. For both scenarios, by introducing even a 2-iteration homotopy scheme, we outperform the one-shot method.

3.6.3 Simulated Generator Outages

For generator outages, we implement a homotopy path that decreases λ from $[\mathbf{1}_{|\mathcal{V}|}, \mathbf{1}_{|\mathcal{V}|}]$ to $[\mathbf{0}_{|\mathcal{V}|}, \mathbf{0}_{|\mathcal{V}|}]$ uniformly throughout the iterations. For this homotopy path, we also consider

varying the discretization of the path. Figure 3.11 shows generator outage scenarios on the 89-bus and 1354-bus PEGASE networks [51, 52]. Figure 3.12 shows the homotopy results on the 3375-bus Polish network with a single generator outage. From these figures, we can see that the final violation cost obtained using the given homotopy paths can vary significantly depending on the number of iterations (or equivalently the step-size $\Delta\lambda$) of *homotopy*-OPF.

3.6.4 Comparison of Homotopy and One-shot Methods

In some of the examples from Figures 3.8, 3.10, and 3.12, we can see that solving the *contingency*-OPF problems with our homotopy method results in a lower violation cost than solving the same problems via the one-shot method. We also considered how far the bus voltages in the *contingency*-OPF problem were from the base case voltages when we solved the problem with homotopy versus one-shot methods, as shown in Figure 3.13 and 3.14. To quantify the severity of the contingency, we also show the voltage variations when using a simple powerflow solver after a contingency. The results show that with homotopy we can obtain a solution that is relatively close to that of the base case, while the solution obtained without homotopy can be unnecessarily far away from that of the base case.

In other cases from Figures 3.9, 3.10, and 3.11, solving the *contingency*-OPF problems via the one-shot method results in non-convergence while the homotopy method can find a convergent solution.

In order to formally compare the performance of homotopy versus the one-shot method, we say that homotopy "outperforms" the one-shot method if either of the following are true:

1. If the homotopy scheme converges and the one-shot method does not converge.

2. If the homotopy scheme converges to a value that is better than that of the one-shot method by at least 0.01% of the optimal *base*-OPF cost.

For the 1354-bus PEGASE network, we tested 1, 2, and 3 line and generator outages, testing 100 simulations of each type of outage. The homotopy paths for these line and generator outages are the same as those described for the simulations in Figures 3.10, 3.11, and 3.12. The percent of simulations where homotopy outperformed the one-shot method is given in Table 3.1 for the network with base-level demand and with demand scaled up by 10%. It can be observed that for the line outage contingencies, the homotopy methods appear to be more useful when the demand is higher. This is likely because the increased demand makes the problem harder, and thus homotopy is more useful. However, the inverse appears true for the generator outage scenarios, i.e. the homotopy methods appear to be more useful when demand is at the base-level. This could be because the removal of a generator could lead to many possibilities for operating the post-contingency network in a lower demand scenario, which may introduce bad local minima.

Although the percent of simulations where homotopy outperforms the one-shot method is less than 20% for the considered cases, it is important to note that in these cases the homotopy method can lead to a significant reduction in the violation cost during a contingency

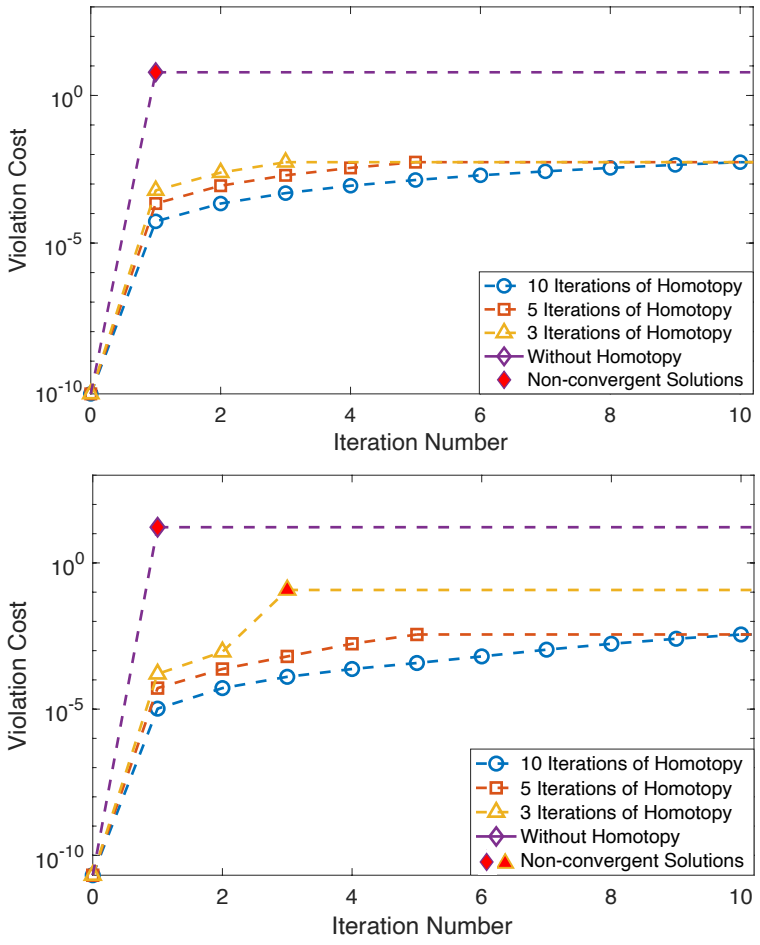

Figure 3.11: Performance of proposed homotopy method for generator outages. The top figure shows a 2 generator outage (generator out IDs: 4 and 7 that respectively generate about 667 MW at bus 37 and about 532 MW at bus 56 in the base case) in the 89-bus PEGASE network (`case89pegase`). The bottom figure shows a 1 generator outage (generator out ID: 30 that generates 1.2 GW at bus 152 in the base case) in the 1354-bus PEGASE network (`case1354pegase`). In these scenarios, the homotopy method can be used to find a convergent solution when the one-shot method fails to find one.

scenario or to a convergent solution when the one-shot method fails to converge. Since a small penalty price is applied to minor violations and an extremely severe penalty is applied to high violations, having even a single contingency scenario with a high violation is problematic for the entire SCOPF problem. Therefore, even if this methodology can improve the current industry solution for 1 out of 10 contingency scenarios, it is extremely beneficial for the security of the entire system. For the cases where the proposed homotopy method does not outperform the one-shot method, the homotopy method typically is at least as good as

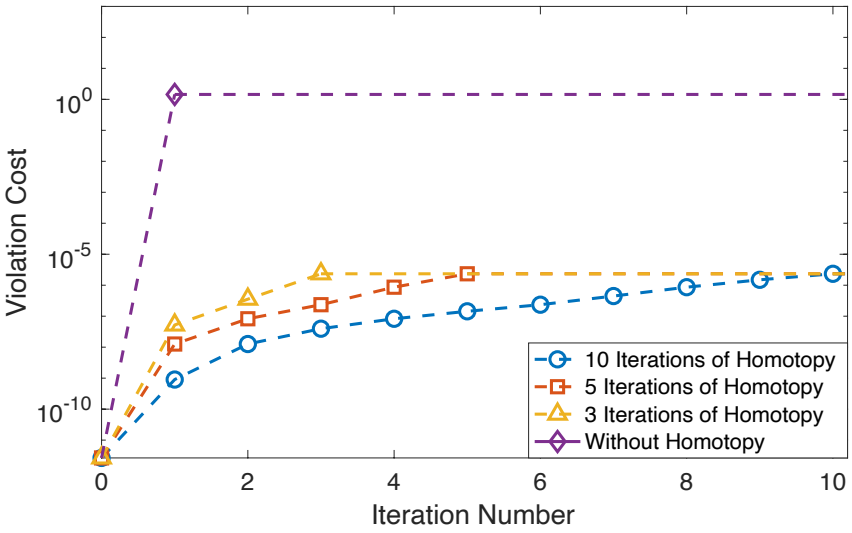

Figure 3.12: Performance of proposed homotopy method on the 3375-bus Polish network (`case3375wp` with real and reactive power demand scaled up by 10%) with a single generator outage. The figure (generator out ID: 100 that generates 370 MW at bus 400 in the base case) shows a scenario where all homotopy discretization schemes result in a violation cost much lower than that obtained by the one-shot method.

Table 3.1: Percent of simulations where 5-iteration homotopy scheme outperformed one-shot method for 1354-bus PEGASE network. For each type of contingency and demand level, we have tested 100 random outage scenarios.

Type of contingency	Base-level power demand	10% greater power demand
1 line outage	10%	12%
2 line outage	7%	12%
3 line outage	12%	15%
1 generator outage	9%	7%
2 generator outage	10%	9%
3 generator outage	17%	12%

the one-shot method.

3.7 Conclusions

This chapter studies the *contingency*-OPF problem, which is used to find an optimal operating point in the case of a line or generator outage. Unlike the *base*-OPF problem that is a

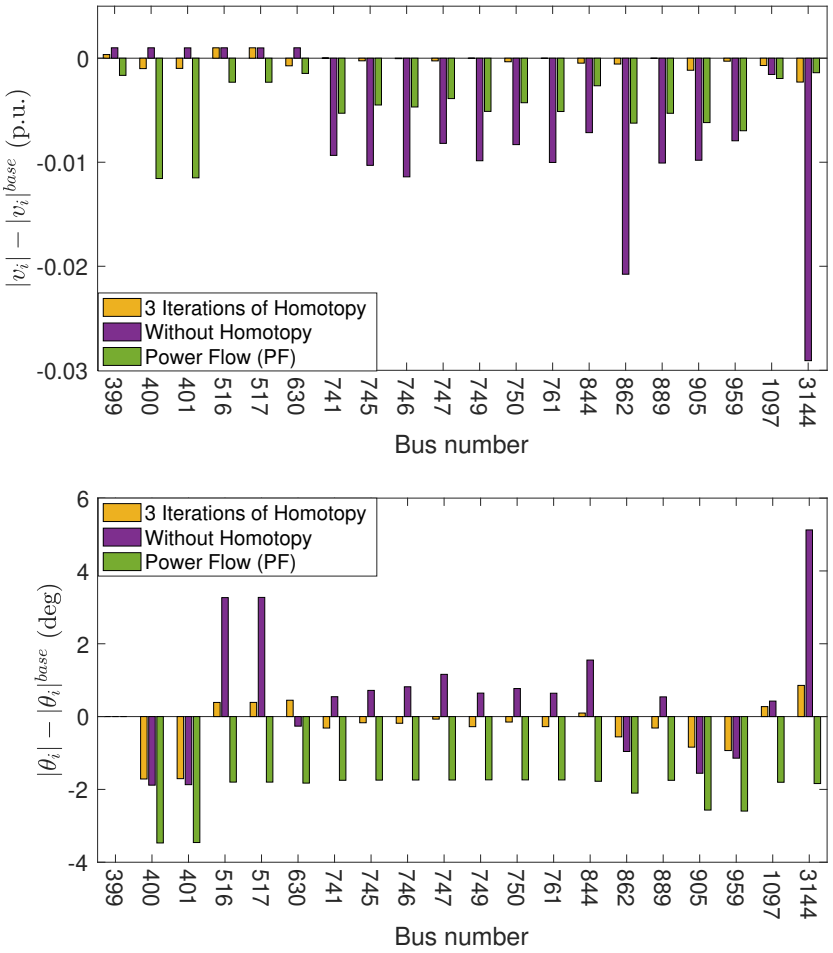

Figure 3.13: Solution analysis for the 3375-bus Polish network (`case3375wp` with real and reactive power demand scaled up by 10%) with a single generator outage (generator out ID: 100 that generates 370 MW at bus 400 in the base case). In these figures, we compare the voltage variation between the solutions of the *contingency*-OPF problem, solved with the 3-iteration homotopy and one-shot methods, to the solution of the post-contingency power flow (PF) problem, solved with Newton-Raphson. For all these post-contingency voltage solutions, we subtract out the base case voltage to determine the variation of the solution from the base case. The twenty *generator* buses with the largest variation from the base case voltage magnitude in the PF solution are shown here since they provide a measure on the severity of the contingency. In this scenario, all three homotopy methods converge to the same solution, as shown in Figure (3.12), so we only compare the solution from the 3-iteration method to that obtained by the one-shot method. For this scenario, the homotopy method gradually deforms the base case to yield a solution that is much closer to the base case while the one-shot method yields a solution that is far from the base case.

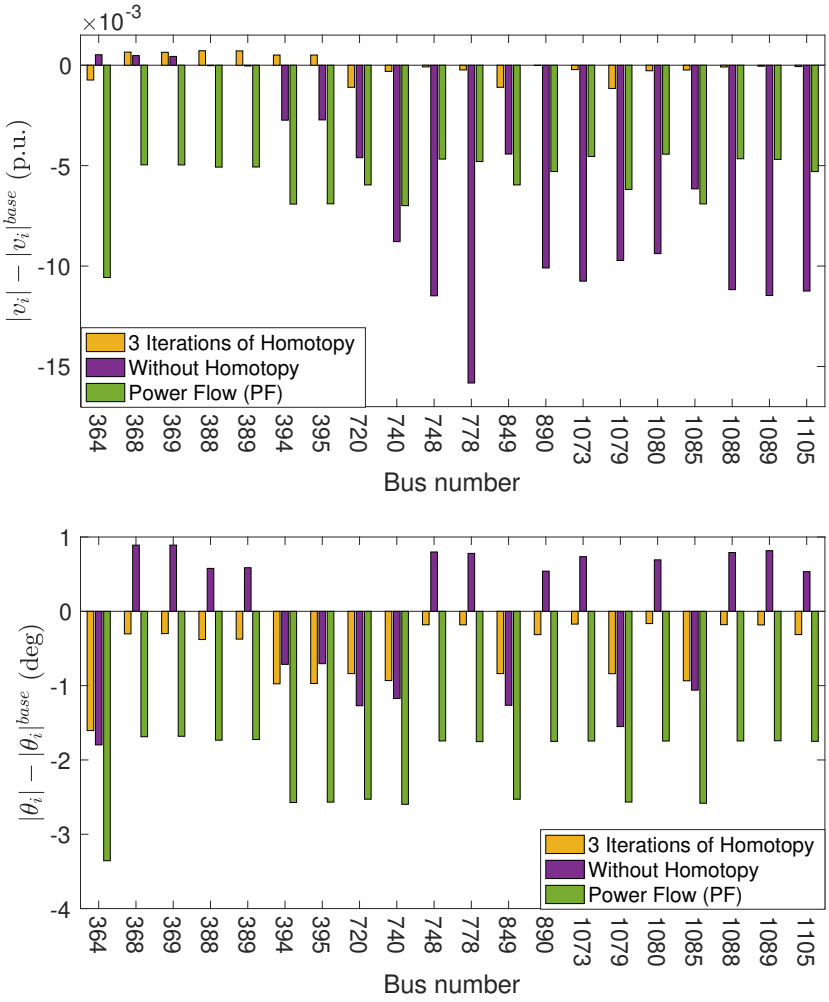

Figure 3.14: Solution analysis for the 3375-bus Polish network (`case3375wp` with real and reactive power demand scaled up by 10%) with a single generator outage (generator out ID: 100 that generates 370 MW at bus 400 in the base case). In these figures, we compare the voltage variation between the solutions of the *contingency*-OPF problem, solved with the 3-iteration homotopy and one-shot methods, to the solution of the post-contingency power flow (PF) problem, solved with Newton-Raphson. For all these post-contingency voltage solutions, we subtract out the base case voltage to determine the variation of the solution from the base case. The twenty *load* buses with the largest variation from the base case voltage magnitude in the PF solution are shown here since they provide a measure on the severity of the contingency. See Figures (3.12) and (3.13) for more details on this scenario.

single optimization problem, there are many *contingency*-OPF problems that should all be solved in a short period of time. Recognizing that the *contingency*-OPF problem is a challenging variant of the classical OPF problem, we introduce a new homotopy method to find the best solution of the *contingency*-OPF problem. This method involves solving a series of intermediate *homotopy*-OPF problems using simple local search methods, and we study conditions that guarantee convergence to a global solution of the *contingency*-OPF. We perform simulations on real-world networks and show that the proposed homotopy method can result in a lower objective value than existing methods. In the majority of considered cases, the proposed homotopy method resulted in the same solution as that obtained by current state-of-the-art methods. However, in other critical cases, homotopy significantly outperformed state-of-the-art interior point methods as measured by both violation cost reduction and solver convergence. Our work improves the security of the grid against low-probability events with catastrophic effects.

Appendix

3.A Computation of Participation Factors for Generator Outage

During the outage of one or more generators, a collection of other generators will increase their power generation in order to respond to the outage and meet power demand. The "participation factor" of a generator determines the portion of the generation response that is assigned to that generator. There are a variety of ways to compute participation factors, including scaling the participation factors based on the remaining power capacity. In Algorithm 3, we present one method for computing participation factors which is based on the topology of the network, i.e. it redirects generation from the outed generators to generators that supply the same set of buses as the outed generators in the *base*-OPF. This method is based on the work [53]. In our simulations of generator outages, we use this method for computing participation factors with Algorithm 2.

3.B Proof of Theorem 4

Let the *homotopy*-OPF problem at λ^o, $H(\lambda^o)$, have a set of KKT points $\mathcal{X}(\lambda^o) = \{x^{(1)}, \ldots, x^{(|\mathcal{X}|)}\}$ that are ordered in a way such that $f(x^{(1)}, \lambda^o) < f(x^{(2)}, \lambda^o) \leq ... \leq f(x^{(|\mathcal{X}|)}, \lambda^o)$. The first strict inequality between the global minimum and the next best local minimum implies that there is a unique global minimum $x^{(1)}$, which is the assumption made in Remark 1. Therefore, by definition, $\lambda^o \notin \mathcal{S}$. We will prove the theorem by proving its contrapositive. Suppose that there exists a $\tau \in [0, T]$ for which the *homotopy*-OPF problem $H(\lambda(\tau))$ has two global solutions, $x_\tau^{(1)}$ and $x_\tau^{(2)}$. To show that the contrapositive is true, we have to show that the path described by $\rho(\lambda) = 0$ intersects with the set $\mathcal{S}$. There are two scenarios that can happen:

(i) When $f(x_\tau^{(1)}, \lambda(\tau)) = f(x_\tau^{(2)}, \lambda(\tau)) \geq a$

(ii) When $f(x_\tau^{(1)}, \lambda(\tau)) = f(x_\tau^{(2)}, \lambda(\tau)) < a$

Note that a is defined in Definition 1. For scenario (i), since $f(x^{(1)}, \lambda^o) < a$, $\exists t \in [0, \tau]$ such that $f(x_t^{(1)}, \lambda(t)) = a$ where $\rho(\lambda(t)) = 0$. This is due to Assumption 1 and the fact

Algorithm 3 Calculation of Participation Factors for Power Redistribution, Contingency k

Given:
(i) Power network $\mathcal{N}(\mathcal{V}, \mathcal{E})$ and generators $\mathcal{G}$
(ii) Solution to *base*-OPF problem $(|v|_*, \theta_*, p^g_*, q^g_*, \{\sigma_{k*}\})$
(iii) Generators out in contingency k: $R_k \subset \mathcal{G}$
Let $(|v|, \theta, p^g, q^g, \{\sigma_k\}) \leftarrow (|v|_*, \theta_*, p^g_*, q^g_*, \{\sigma_{k*}\})$
Compute real power flow for all $(i,j) \in \mathcal{E}$ in the base case:
$p_{ij} = G_{ij}|v_i|^2 - G_{ij}|v_i||v_j|\cos(\theta_{ij}) + B_{ij}|v_i||v_j|\sin(\theta_{ij})$
Generate a directed graph $\mathcal{D}(\mathcal{V}, \mathcal{A})$ based on direction of power flow: $(i,j) \in \mathcal{A}$ if $p_{ij} \geq 0$
Use shortest path algorithm to compute the domain of each generator
Group the buses supplied by the same set of generators into commons $\mathcal{C}$ (see [53])
Use algorithm in [53] to determine the contribution C_{rj} of each generator r to common j
Remove contribution of generators that are out:
$C_{rj} \leftarrow 0, \quad \forall r \in R_k, \ \forall j \in \mathcal{C}$
Distribute lost generation over generations that supply the same common:
for $j \in \mathcal{C}$ **do**
  Define $C_j \triangleq \sum_r C_{rj}$
  if $C_j \neq 0$ **then**
    for $r \in \mathcal{G}$ **do**
      $C_{rj} \leftarrow C_{rj}/C_j$
    end for
  end if
end for
Initialize participation factors: $\alpha^g_r = 0$ for all $r \in \mathcal{G}$
Define participation factors based on contribution to common:
for $r \in R_k$ **do**
  for $j \in \mathcal{C}$ **do**
    $\alpha^g_t \leftarrow \alpha^g_t + C_{tj}$ for all generators t in common j
  end for
end for
Normalize the participation factors α^g so that $\sum_{r\in\mathcal{G}} \alpha^g_r = 1$

that the KKT points change continuously with respect to the parameter λ. Similarly, for scenario (ii), since $f(x^{(2)}, \lambda^o) > a$, $\exists t \in [0, \tau]$ such that $f(x^{(2)}_t, \lambda(t)) = a$ where $\rho(\lambda(t)) = 0$. In both scenarios, the path described by $\rho(\lambda) = 0$ intersects with the set $\mathcal{S}$, which proves the contrapositive and completes the proof.

3.C Proof of Theorem 5

Due to Theorem 4, it is sufficient to show the equivalence between the following two statements:

(i) The path described by $\rho(\lambda) = 0$ does not intersect with the set $\mathcal{S}$.

(ii) There exists a dual variable quadruplet $(\omega_1, \omega_2, \omega_3, \omega_4)$ such that $d(\omega_1, \omega_2, \omega_3, \omega_4) > 0$.

By definition of the set $\mathcal{S}$, statement (i) is equivalent to saying that the following set of equations do not have a solution:

$$\begin{aligned} \nabla f(x,\lambda) + \mu^T \nabla h(x,\lambda) &= 0 \\ h(x,\lambda) &= 0 \\ f(x,\lambda) &= a \\ \rho(\lambda) &= 0 \end{aligned} \tag{3.21}$$

In other words, the following feasibility problem is infeasible:

$$(P) \quad \min_{x,\lambda,\mu} \quad 0 \tag{3.22a}$$

$$\text{s.t. } \nabla f(x,\lambda) + \mu^T \nabla h(x,\lambda) = 0 \tag{3.22b}$$

$$h(x,\lambda) = 0 \tag{3.22c}$$

$$f(x,\lambda) = a \tag{3.22d}$$

$$\rho(\lambda) = 0 \tag{3.22e}$$

By duality theory, if the dual problem (D) is unbounded, then the primal problem (P) must be infeasible. However, since the primal objective value is zero and the dual problem should provide a lower bound to the primal, finding a dual certificate that gives a positive dual objective value is sufficient in proving that the primal problem is infeasible. This completes the proof.

3.D Proof of Lemma 2

Let us start with the equation for the reactive power injections. Let θ_1 and θ_2 denote the voltage phasor angles at bus 1 and 2, respectively. Then after denoting $\theta = \theta_1 - \theta_2$, we have the following:

$$q_1^{inj} = \lambda_2 b - \lambda_1 g \cdot \sin\theta - \lambda_2 b \cdot \cos\theta \tag{3.23a}$$

$$q_2^{inj} = \lambda_2 b + \lambda_1 g \cdot \sin\theta - \lambda_2 b \cdot \cos\theta \tag{3.23b}$$

A lower bound of $Q^{\min}$ on q_1^{inj} results in the following calculations:

$$Q^{\min} \leq \lambda_2 b - \lambda_1 g \cdot \sin\theta - \lambda_2 b \cdot \cos\theta$$

$$\begin{aligned}
&\iff -Q^{\min} + \lambda_2 b \geq \lambda_1 g \cdot \sin\theta + \lambda_2 b \cdot \cos\theta \\
&\qquad = \sqrt{(\lambda_1 g)^2 + (\lambda_2 b)^2} \cdot \cos(\theta - \Delta) \qquad \text{where } \Delta = \tan^{-1}\left(\frac{\lambda_1 g}{\lambda_2 b}\right) \\
&\iff \cos(\theta - \Delta) \leq \frac{-Q^{\min} + b \cdot \lambda_2}{\sqrt{(\lambda_1 g)^2 + (\lambda_2 b)^2}} \\
&\qquad \iff \theta \geq \cos^{-1}\left(\frac{-Q^{\min} + b \cdot \lambda_2}{\sqrt{(\lambda_1 g)^2 + (\lambda_2 b)^2}}\right) + \Delta \\
&\qquad\qquad \text{or} \quad \theta \leq -\cos^{-1}\left(\frac{-Q^{\min} + b \cdot \lambda_2}{\sqrt{(\lambda_1 g)^2 + (\lambda_2 b)^2}}\right) + \Delta
\end{aligned} \tag{3.24}$$

From the lower bound on q_2^{inj}, we can perform a similar derivation and arrive at the following inequality:

$$\begin{aligned}
&\theta \geq \cos^{-1}\left(\frac{-Q^{\min} + b \cdot \lambda_2}{\sqrt{(\lambda_1 g)^2 + (\lambda_2 b)^2}}\right) - \Delta \\
&\qquad \text{or} \quad \theta \leq -\cos^{-1}\left(\frac{-Q^{\min} + b \cdot \lambda_2}{\sqrt{(\lambda_1 g)^2 + (\lambda_2 b)^2}}\right) - \Delta
\end{aligned} \tag{3.25}$$

Therefore, combining inequalities (3.24) and (3.25), we derive the following inequality:

$$\begin{aligned}
&\theta \geq \cos^{-1}\left(\frac{-Q^{\min} + b \cdot \lambda_2}{\sqrt{(\lambda_1 g)^2 + (\lambda_2 b)^2}}\right) + \Delta \\
&\qquad \text{or } \theta \leq -\cos^{-1}\left(\frac{-Q^{\min} + b \cdot \lambda_2}{\sqrt{(\lambda_1 g)^2 + (\lambda_2 b)^2}}\right) - \Delta.
\end{aligned} \tag{3.26}$$

Furthermore, we assume that

$$-\tan^{-1}\left(\frac{\lambda_2 b}{\lambda_1 g}\right) \leq \theta \leq \tan^{-1}\left(\frac{\lambda_2 b}{\lambda_1 g}\right) \tag{3.27}$$

which is equivalent to the following using basic trigonometry:

$$-\left(\frac{\pi}{2} - \Delta\right) \leq \theta \leq \left(\frac{\pi}{2} - \Delta\right) \tag{3.28}$$

Combining the two inequalities (3.26) and (3.28), and using the definition of α, we get the final constraint on θ:

$$\alpha + \Delta \leq \theta \leq \left(\frac{\pi}{2} - \Delta\right) \quad \text{or} \quad -\left(\frac{\pi}{2} - \Delta\right) \leq \theta \leq -\alpha - \Delta. \tag{3.29}$$

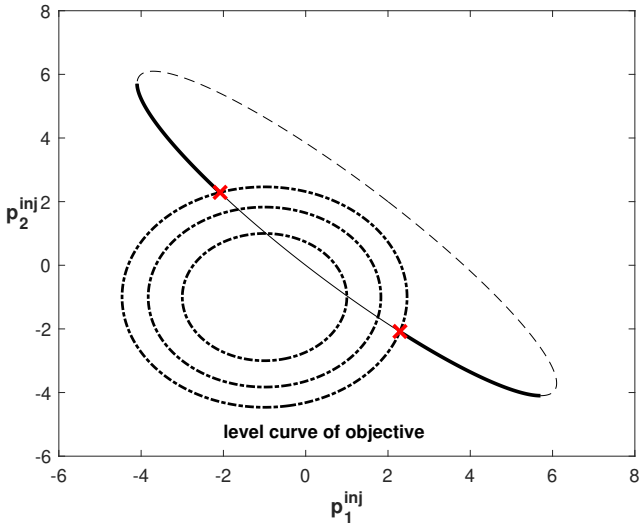

Figure 3.15: An example of two-bus network for which there are two global solutions to the OPF.

This feasible region of θ is reflected in the feasible region of the active power injections, as shown in the bolded part of the ellipse in Figure 3.15. As illustrated in the figure, the two red points are active power injections, corresponding to $\theta = \alpha + \Delta$ and $\theta = -\alpha - \Delta$. Let the first red point, (P_1^{inj}, P_2^{inj}), be generated by $\theta = \alpha + \Delta$. Then, the following is true for P_1^{inj}:

$$\begin{aligned} P_1^{inj} &= \lambda_1 g + \lambda_2 b \cdot \sin\theta - \lambda_1 g \cdot \cos\theta \\ &= \lambda_1 g + \lambda_2 b \cdot \sin(\alpha + \Delta) - \lambda_1 g \cdot \cos(\alpha + \Delta) \\ &= \lambda_1 g + \lambda_2 b \cdot (\sin\alpha \cdot \cos\Delta + \alpha \sin\Delta) - \lambda_1 g \cdot (\alpha \cos\Delta - \sin\alpha \cdot \sin\Delta) \\ &= \lambda_1 g + \frac{\lambda_2 b}{|y|}(\lambda_2 b \cdot \sin\alpha + \alpha \cdot \lambda_1 g) - \frac{\lambda_1 g}{|y|}(\alpha \cdot \lambda_2 b - \lambda_1 g \cdot \sin\alpha) \end{aligned} \tag{3.30}$$

Similarly, if we let the second red point $(\bar{P}_1^{inj}, \bar{P}_2^{inj})$, be generated by $\theta = -\alpha - \Delta$, we have

$$\bar{P}_1^{inj} = \lambda_1 g - \frac{\lambda_2 b}{|y|}(\lambda_2 b \cdot \sin\alpha + \alpha \cdot \lambda_1 g) - \frac{\lambda_1 g}{|y|}(\alpha \cdot \lambda_2 b - \lambda_1 g \cdot \sin\alpha) \tag{3.31}$$

Also, note that due to symmetry, $P_2^{inj} = \bar{P}_1^{inj}$ and $\bar{P}_2^{inj} = P_1^{inj}$. Let's define the following two functions:

$$w_1(\lambda_1, \lambda_2) \triangleq P_1^{inj} - \bar{P}_1^{inj} = \frac{2\lambda_2 b}{|y|}\big(\lambda_2 b \cdot \sin\alpha + \alpha \cdot \lambda_1 g\big) \tag{3.32a}$$

$$w_2(\lambda_1, \lambda_2) \triangleq P_1^{inj} + \bar{P}_1^{inj} = 2\lambda_1 g - \frac{2\lambda_1 g}{|y|}\big(- \lambda_1 g \cdot \sin\alpha + \alpha \cdot \lambda_2 b\big) \tag{3.32b}$$

If the two points (P_1^{inj}, P_2^{inj}) and $(\bar{P}_1^{inj}, \bar{P}_2^{inj})$ are both globally optimal, their objective values must be equal. In other words,

$$\begin{aligned} & (P_1^{inj} + P_1^d)^2 + c(P_2^{inj} + P_2^d)^2 = (\bar{P}_1^{inj} + P_1^d)^2 + c(\bar{P}_2^{inj} + P_2^d)^2 \\ \iff \quad & (1-c)\{(P_1^{inj})^2 - (\bar{P}_1^{inj})^2\} + 2P_1^d(P_1^{inj} - \bar{P}_1^{inj}) - 2cP_2^d(P_1^{inj} - \bar{P}_1^{inj}) = 0 \\ \iff \quad & (1-c) \cdot w_1(\lambda_1, \lambda_2) \cdot w_2(\lambda_1, \lambda_2) + 2P_1^d \cdot w_1(\lambda_1, \lambda_2) - 2cP_2^d \cdot w_1(\lambda_1, \lambda_2) = 0 \end{aligned} \tag{3.33}$$

This completes the proof.

3.5 Bibliography

[1] S. Park, E. Glista, J. Lavaei, and S. Sojoudi, "Homotopy method for finding the global solution of post-contingency optimal power flow," *American Control Conference*, 2020.

[2] ——, "An efficient homotopy method for solving the post-contingency optimal power flow to global optimality," *IEEE Access*, vol. 10, pp. 124 960–124 978, 2022.

[3] S. Park, "Computational methods for the design and operations of electric power systems: Towards resiliency and security," Ph.D. dissertation, UC Berkeley, 2022.

[4] J. Carpentier, "Contribution to the economic dispatch problem," *Bulletin de la Societe Francoise des Electriciens*, vol. 3, no. 8, pp. 431–447, 1962.

[5] F. Capitanescu, J. Martinez Ramos, P. Panciatici, D. Kirschen, A. Marano Marcolini, L. Platbrood, and L. Wehenkel, "State-of-the-art, challenges, and future trends in security constrained optimal power flow," *Electric Power Systems Research*, vol. 81, no. 8, pp. 1731 – 1741, 2011.

[6] R. Madani, M. Ashraphijuo, and J. Lavaei, "Promises of conic relaxation for contingency-constrained optimal power flow problem," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 1297–1307, 2016.

[7] X. Bai and H. Wei, "Semi-definite programming-based method for security-constrained unit commitment with operational and optimal power flow constraints," *IET Generation, Transmission & Distribution*, vol. 3, no. 2, pp. 182–197, 2009.

[8] F. Capitanescu, M. Glavic, D. Ernst, and L. Wehenkel, "Contingency filtering techniques for preventive security-constrained optimal power flow," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 1690–1697, 2007.

[9] F. Bouffard, F. Galiana, and J. Arroyo, "Umbrella contingencies in security-constrained optimal power flow," in *Proc. of the 15th Power Systems Computation Conference (PSCC)*, 2005.

[10] R. Madani, J. Lavaei, and R. Baldick, "Constraint screening for security analysis of power networks," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 1828–1838, 2017.

[11] Y. Li and J. D. McCalley, "Decomposed SCOPF for improving efficiency," *IEEE Transactions on Power Systems*, vol. 24, no. 1, pp. 494–495, 2009.

[12] A. Marano-Marcolini, F. Capitanescu, J. L. Martinez-Ramos, and L. Wehenkel, "Exploiting the use of DC SCOPF approximation to improve iterative AC SCOPF algorithms," *IEEE Transactions on Power Systems*, vol. 27, no. 3, pp. 1459–1466, 2012.

[13] Q. Wang, J. D. McCalley, T. Zheng, and E. Litvinov, "Solving corrective risk-based security-constrained optimal power flow with Lagrangian relaxation and Benders decomposition," *International Journal of Electrical Power & Energy Systems*, vol. 75, pp. 255–264, 2016.

[14] K. Karoui, H. Crisciu, A. Szekut, and M. Stubbe, "Large scale security constrained optimal power flow," in *Proc. of the 16th Power Systems Computation Conference (PSCC)*, 2008.

[15] M. Bazrafshan, K. Baker, and J. Mohammadi, "Computationally efficient solutions for large-scale security-constrained optimal power flow," 2020.

[16] I. Avramidis, F. Capitanescu, S. Karagiannopoulos, and E. Vrettos, "A novel approximation of security-constrained optimal power flow with incorporation of generator frequency and voltage control response," *IEEE Transactions on Power Systems*, 2020.

[17] J. Mohammadi, G. Hug, and S. Kar, "Agent-based distributed security constrained optimal power flow," *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 1118–1130, 2018.

[18] W. Zhang, Y. Xu, Z. Dong, and K. P. Wong, "Robust security constrained-optimal power flow using multiple microgrids for corrective control of power systems under uncertainty," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 1704–1713, 2017.

[19] E. Karangelos and L. Wehenkel, "An iterative AC-SCOPF approach managing the contingency and corrective control failure uncertainties with a probabilistic guarantee," *IEEE Transactions on Power Systems*, vol. 34, no. 5, pp. 3780–3790, 2019.

[20] A. Velloso and P. Van Hentenryck, "Combining deep learning and optimization for preventive security-constrained DC optimal power flow," *IEEE Transactions on Power Systems*, 2021.

[21] A. Monticelli, M. V. F. Pereira, and S. Granville, "Security-constrained optimal power flow with post-contingency corrective rescheduling," *IEEE Transactions on Power Systems*, vol. 2, no. 1, pp. 175–180, 1987.

[22] F. Capitanescu, T. V. Cutsem, and L. Wehenkel, "Coupling optimization and dynamic simulation for preventive-corrective control of voltage instability," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 796–805, 2009.

[23] Y. Xu, Z. Y. Dong, R. Zhang, K. P. Wong, and M. Lai, "Solving preventive-corrective SCOPF by a hybrid computational strategy," *IEEE Transactions on Power Systems*, vol. 29, no. 3, pp. 1345–1355, 2014.

[24] A. Gholami, K. Sun, S. Zhang, and X. A. Sun, "An ADMM-based distributed optimization method for solving security-constrained ac optimal power flow," 2022.

[25] L. T. Watson, "Numerical linear algebra aspects of globally convergent homotopy methods," *SIAM Review*, vol. 28, no. 4, pp. 529–545, December 1986.

[26] L. T. Watson and R. T. Haftka, "Modern homotopy methods in optimization," *Computer Methods in Applied Mechanics and Engineering*, vol. 74, no. 3, pp. 289–305, September 1989.

[27] E. L. Allgower and K. Georg, "Introduction to numerical continuation methods," in *SIAM Classics in Applied Mathematics*, 2003.

[28] S.-N. Chow, J. Mallet-Paret, and J. A. Yorke, "Finding zeroes of maps: Homotopy methods that are constructive with probability one," *Mathematics of Computation*, vol. 32, no. 143, pp. 887–899, July 1978.

[29] V. Ajjarapu and C. Christy, "The continuation power flow: A tool for steady state voltage stability analysis," *IEEE Transactions on Power Systems*, vol. 7, no. 1, pp. 416–423, February 1992.

[30] S. Yu, H. D. Nguyen, and K. S. Turitsyn, "Simple certificate of solvability of power flow equations for distribution systems," *2015 IEEE Power & Energy Society General Meeting*, July 2015.

[31] H.-D. Chiang, T.-Q. Zhao, J.-J. Deng, and K. Koyanagi, "Homotopy-enhanced power flow methods for general distribution networks with distributed generators," *IEEE Transactions on Power Systems*, vol. 29, no. 1, pp. 93–100, January 2014.

[32] A. Pandey, M. Jereminov, M. Wagner, G. Hug, and L. Pileggi, "Robust convergence of power flow using TX stepping method with equivalent circuit formulation," November 2017.

[33] D. Mehta, H. D. Nguyen, and K. Turitsyn, "Numerical polynomial homotopy continuation method to locate all the power flow solutions," *IET Generation, Transmission & Distribution*, vol. 10, no. 12, pp. 2972 – 2980, August 2016.

[34] D. Zigic, L. T. Watson, E. G. Collins, and D. S. Bernstein, "Homotopy approaches to the H_2 reduced order model problem," 1991.

[35] H. Feng and J. Lavaei, "Damping with varying regularization in optimal decentralized control," 2019, available online at https://lavaei.ieor.berkeley.edu/ODC_hom_2019_2.pdf.

[36] P. Garrigues and L. E. Ghaoui, "An homotopy algorithm for the Lasso with online observations," in *Advances in Neural Information Processing Systems*, 2009, pp. 489–496.

[37] A. Poore and A.-H. Q., "The expanded Lagrangian system for constrained optimization problems," *SIAM Journal on Control and Optimization*, vol. 26, no. 2, p. 417–427, 1988.

[38] D. M. Dunlavy and D. P. O'Leary, "Homotopy optimization methods for global optimization," *Sandia National Laboratories*, December 2005.

[39] H. Mobahi and J. W. Fisher III, "A theoretical analysis of optimization by Gaussian continuation," *AAAI Conference on Artificial Intelligence, North America*, February 2015.

[40] D. P. Bertsekas, *Nonlinear programming*. Athena scientific, 2016.

[41] V. H. Hinojosa and F. Gonzalez-Longatt, "Preventive security-constrained DCOPF formulation using power transmission distribution factors and line outage distribution factors," *Energies*, vol. 11, no. 6, 2018.

[42] L. T. Watson, "Theory of globally convergent probability-one homotopies for nonlinear programming," *SIAM Journal on Optimization*, vol. 11, no. 3, p. 761–780, 2000.

[43] B. Sturmfels, *Solving systems of polynomial equations*, 2002.

[44] J. E. Spingarn, "Multifunctions and integrands." Springer, 1984, ch. Multifunctions associated with parametrized classes of constrained optimization problems, pp. 206–215.

[45] P. A. Parrilo, "Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization," 2000, PhD thesis, http://www.mit.edu/~parrilo/pubs/files/thesis.pdf.

[46] J. Lavaei, D. Tse, and B. Zhang, "Geometry of power flows and optimization in distribution networks," *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 572–583, 2014.

[47] L. A. Clarfeld, P. D. H. Hines, E. M. Hernandez, and M. J. Eppstein, "Risk of cascading blackouts given correlated component outages," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 3, pp. 1133–1144, 2020.

[48] Vaiman, Bell, Chen, Chowdhury, Dobson, Hines, Papic, Miller, and Zhang, "Risk assessment of cascading outages: Methodologies and challenges," *IEEE Transactions on Power Systems*, vol. 27, no. 2, pp. 631–641, 2012.

[49] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, "MATPOWER: Steady-state operations, planning and analysis tools for power systems research and education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, February 2011.

[50] H. Wang, C. E. Murillo-Sánchez, R. D. Zimmerman, and R. J. Thomas, "On computational issues of market-based optimal power flow," *IEEE Transactions on Power Systems*, vol. 22, no. 3, pp. 1185–1193, August 2007.

[51] C. Josz, S. Fliscounakis, J. Maeght, and P. Panciatic, "AC power flow data in MATPOWER and QCQP format: iTesla, RTE snapshots, and PEGASE," available online at http://arxiv.org/abs/1603.01533.

[52] S. Fliscounakis, P. Panciatici, F. Capitanescu, and L. Wehenkel, "Contingency ranking with respect to overloads in very large power systems taking into account uncertainty, preventive and corrective actions," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4909 – 4917, November 2013.

[53] D. Kirschen, R. Allan, and G. Strbac, "Contributions of individual generators to loads and flows," *IEEE Transactions on Power Systems*, vol. 12, no. 1, pp. 52–60, Feburary 1997.

Chapter 4

Optimal Measurement Choice in Robust Robust Power System State Estimation

In this chapter, we present a framework for Optimal Measurement Choice in Robust Power Grid State Estimation.[1]

4.1 Introduction

Power system state estimation (SE) is a critical problem for the reliability of the electric grid [2, 3]. SE uses data from sensors throughout a transmission or distribution network to monitor the state of the network [4]. The estimated state is in turn used to make decisions about real-time power dispatch, implement voltage control, and take action in the case of a contingency, such as a line or generator outage, as shown in Figure 4.1 [2]. During the Northeast power blackout of 2003, which affected over 50 million people in the U.S. and Canada, the propagation of cascading failures could have been mitigated had the operators been able to recover the true state of the network [5]. Because sensor measurements may be subject to both random noise and intentional cyberattacks, it is important to consider a robust version of the SE problem [6]. Furthermore, as cyberattacks increase in frequency, robust SE will become more important in the design of algorithms for the future smart grid [7–9].

A special case of graph-structured quadratic sensing, SE is formulated as the minimization of a loss function representing the difference between the actual set of measurements and the measurements that would be observed for the estimated state [10]. The state of a power network is defined by a complex voltage at each bus in the network. Due to the nonlinearity of alternating current (AC) power flow, the classical SE problem is nonlinear, making the problem NP-hard. In practice, nonlinear SE is often formulated as a weighted least-squares

[1] Chapter 4 includes materials from [1] that were previously published.

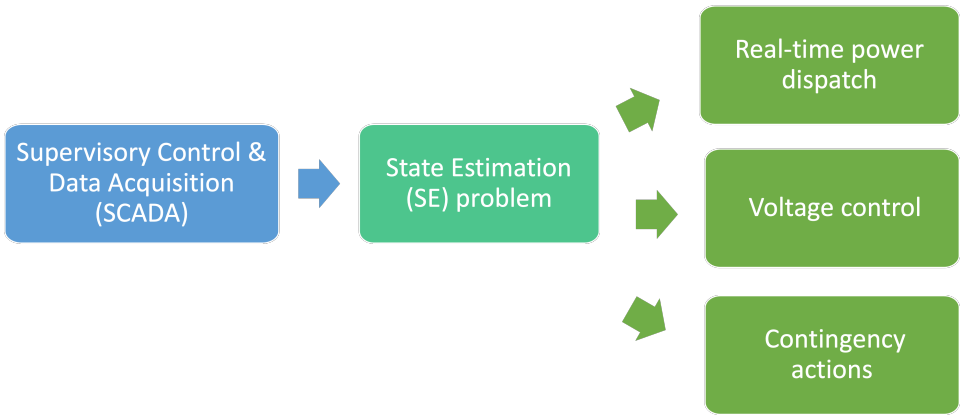

Figure 4.1: Role of state estimation in power systems operation.

problem and is solved with local search algorithms such as Newton's method or approximated with linearizations [11–14]. However, local search methods may yield spurious local minima with no physical meaning since SE does not satisfy the restricted isometry property (RIP) from quadratic sensing that can be used to certify a lack of spurious local minima [15–17]. Spurious local minima have no physical meaning in relation to the true state of the power system and are hard to distinguish from the global minimum that corresponds to the true state. Because of this, there is growing interest in methods that can yield global solutions to the SE problem such as stochastic and convex methods [18–25]. The paper [10] proposes a two-step power system SE method which allows for the recovery of the true state of the system in the case without noise or bad data. Because this method involves solving a linear SE problem, it is convex and can be solved to global optimality efficiently with existing local search methods. Additionally, [10] introduces a sufficient condition to verify the robustness of SE that explicitly depends on the support of the bad data, and [26] extends this work to propose a method which certifies that a network is locally robust to bad data without any dependence on the bad data support.

With the increasing number of phasor measurement units (PMUs) in power systems, there is growing interest in SE methods that make use of PMU data [27–30]. Because sensors such as PMUs are typically expensive, it is not always economical to install sensors at every possible location in a power network [31]. Thus, there is interest in optimizing the placement of sensors in a power network, with various objectives with regards to SE such as ensuring the system is observable [32–40]. Many of these methods formulate the measurement placement problem as a mixed-integer program (MIP), which is a natural formulation for making the binary choice whether or not to place a sensor at a particular location. While some papers attempt to optimize measurement placement in order to ensure SE robustness, these methods focus on ensuring overall redundancy rather than directly considering how bad data at some buses affect the rest of the measurements in the connected network [41]. Our work builds

on the existing literature by considering how bad data propagates in a power system SE problem in order to optimize measurement choice.

4.1.1 Contributions

This work proposes a novel framework to optimize the placement of sensors in a network in order to satisfy a machine learning condition for SE robustness. To do this, we leveraged a linearized SE framework and the concept of local partitioning in order to define a MILP that formalizes the measurement choice problem. The MILP is a coupled optimization problem over all lines in the network that has the goal of optimizing the local mutual incoherence metric for each line in the network. The coupling occurs through the choice of measurement set, which affects the mutual incoherence for all lines. The proposed framework provides measurement placement choices that yield greater system observability in case of cyberattack.

4.1.2 Notations

The symbol $\mathbb{R}$ denotes the set of real numbers, and $\mathbb{R}^N$ denotes the space of N-dimensional real vectors. The symbol $(\cdot)^T$ denotes the transpose of a vector or matrix. The symbol $|\cdot|$ is the absolute value operator if the argument is a scalar, vector, or matrix; otherwise, it is the cardinality of a measurable set. The imaginary unit is denoted by $\mathbf{j} = \sqrt{-1}$. The elementwise multiplication of two matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{m\times n}$ is denoted as $A \odot B$. The symbol $\dagger$ denotes the left pseudoinverse of a matrix given as $A^\dagger \triangleq (A^T A)^{-1} A^T$. The notation $||A||_\infty$ corresponds to the matrix infinity norm, e.g. the maximum absolute column sum of matrix A. The expression $\mathbf{1}_n$ is a vector of ones of dimension n, and the expression $\mathbf{1}\{\zeta\}$ is the indicator function which is 1 if ζ is true and 0 otherwise. The notation $A[\mathcal{B}, \mathcal{C}]$ or $A_{\mathcal{B},\mathcal{C}}$ represents a submatrix of matrix A formed by taking the rows and columns corresponding respectively to the sets $\mathcal{B}$ and $\mathcal{C}$. The notation $\mathcal{A} \setminus \mathcal{B}$ denotes the subtraction of set $\mathcal{B}$ from set $\mathcal{A}$, and $\mathcal{A} \cup \mathcal{B}$ denotes the union of sets $\mathcal{A}$ and $\mathcal{B}$. The notation $[n]$ denotes the index set $\{1, \ldots, n\}$.

4.2 Background

4.2.1 Power System State Estimation (SE)

Let a power network be defined as the graph $\mathcal{N}(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of buses and $\mathcal{E}$ is the set of lines. The goal of SE is to recover the true state of the network, given as the complex voltage $v_i \triangleq |v_i| e^{\mathbf{j}\theta_i}$ at each bus $i \in \mathcal{V}$. We are given some set of measurements $\mathcal{M}$, which can include measurements of the real or reactive power flows p_{ij}, q_{ij} on line $(i, j) \in \mathcal{E}$, the real or reactive power injected p_i, q_i at bus $i \in \mathcal{V}$, or the voltage magnitude $|v_i|$ at bus $i \in \mathcal{V}$. We can also extend this method to include phase angle measurements θ_i for $i \in \mathcal{V}$ from phasor measurement units (PMUs). We will use the SE method from [10], which introduces a linear

basis using the unknown state variables:

$$x_i^{\text{mg}} \triangleq |v_i|^2 \qquad \forall i \in \mathcal{V} \tag{4.1a}$$
$$x_{ij}^{\text{re}} \triangleq |v_i||v_j|\cos(\theta_{ij}) \qquad \forall (i,j) \in \mathcal{E} \tag{4.1b}$$
$$x_{ij}^{\text{im}} \triangleq |v_i||v_j|\sin(\theta_{ij}) \qquad \forall (i,j) \in \mathcal{E} \tag{4.1c}$$

where $\theta_{ij} \triangleq \theta_i - \theta_j$ for all $(i,j) \in \mathcal{E}$. We will take the set

$$\mathcal{X} \triangleq \left\{ \{x_i^{\text{mg}}\}_{\forall i \in \mathcal{V}}, \{x_{ij}^{\text{re}}\}_{\forall (i,j) \in \mathcal{E}}, \{x_{ij}^{\text{im}}\}_{\forall (i,j) \in \mathcal{E}} \right\} \tag{4.2}$$

to be the set of all states for the network, which is fixed given the network topology.

Given this linear basis, the equations which relate the measurements to the unknown state are:

$$|v_i|^2 = x_i^{\text{mg}} \qquad \forall i \in \mathcal{V} \tag{4.3a}$$
$$p_{ij} = G_{ij} x_i^{\text{mg}} - G_{ij} x_{ij}^{\text{re}} - B_{ij} x_{ij}^{\text{im}} \qquad \forall (i,j) \in \mathcal{E} \tag{4.3b}$$
$$p_{ji} = G_{ij} x_j^{\text{mg}} - G_{ij} x_{ij}^{\text{re}} + B_{ij} x_{ij}^{\text{im}} \qquad \forall (i,j) \in \mathcal{E} \tag{4.3c}$$
$$q_{ij} = -B_{ij}^* x_i^{\text{mg}} + B_{ij} x_{ij}^{\text{re}} - G_{ij} x_{ij}^{\text{im}} \qquad \forall (i,j) \in \mathcal{E} \tag{4.3d}$$
$$q_{ji} = -B_{ij}^* x_j^{\text{mg}} + B_{ij} x_{ij}^{\text{re}} + G_{ij} x_{ij}^{\text{im}} \qquad \forall (i,j) \in \mathcal{E} \tag{4.3e}$$
$$p_i = \sum_{j:(i,j)\in\mathcal{E}} p_{ij} + \sum_{j:(j,i)\in\mathcal{E}} p_{ij} \qquad \forall i \in \mathcal{V} \tag{4.3f}$$
$$q_i = \sum_{j:(i,j)\in\mathcal{E}} q_{ij} + \sum_{j:(j,i)\in\mathcal{E}} q_{ij} \qquad \forall i \in \mathcal{V} \tag{4.3g}$$

where G_{ij} is the conductance, B_{ij} is the susceptance, and $B_{ij}^* \triangleq B_{ij} + \frac{1}{2} B_{ij}^{\text{sh}}$ with shunt susceptance B_{ij}^{sh} for line $(i,j) \in \mathcal{E}$. Using the set of equations (4.3), we can define a state equation of the form $\mathbf{m} = A\mathbf{x}$ where $A \in \mathbb{R}^{m \times n}$ is the sensing matrix that relates the unknown state $\mathbf{x} \in \mathbb{R}^n$ to the vector of measurements $\mathbf{m} \in \mathbb{R}^m$. We have that $n \triangleq |\mathcal{X}| = |\mathcal{V}| + 2|\mathcal{E}|$ and $m \triangleq |\mathcal{M}|$. Note that A is sparse due to the sparse nature of power networks.

When $m > n$, the equation $\mathbf{m} = A\mathbf{x}$ represents an over-determined power flow problem. We will assume that we always have $m \geq n$. In a realistic scenario, the measurements $\mathbf{m}$ are corrupted with random noise and potentially other bad data, and therefore we cannot just solve this over-determined power flow to determine the true state. We can model the noisy and/or corrupted measurements $\mathbf{y} \in \mathbb{R}^m$ as:

$$\mathbf{y} = A\mathbf{x} + \mathbf{w} + \mathbf{b} \tag{4.4}$$

where $\mathbf{w} \in \mathbb{R}^m$ represents random noise and $\mathbf{b} \in \mathbb{R}^m$ represents the bad data vector. Typical assumptions on these vectors are that $\mathbf{w}$ follows a Gaussian distribution and that $\mathbf{b}$ is a sparse vector [42–45]. Note that the local recovery method in [26] is one of the most general methods as it does not make assumptions on the sparsity of $\mathbf{b}$.

The SE methods of [10] and [26] use a two-step process:

1. Solve SE problem defined by (4.4) to get an estimate $\hat{\mathbf{x}}$.
2. Recover an estimate of the complex voltages using the relations:
   - *Voltage magnitudes*: $|\hat{v}_i| = \sqrt{\hat{x}_i^{\text{mg}}}$ for all $i \in \mathcal{V}$
   - *Voltage phase angle differences*: $\hat{\theta}_{ij} = \arctan\left(\hat{x}_{ij}^{\text{im}}/\hat{x}_{ij}^{\text{re}}\right)$ for all $(i,j) \in \mathcal{E}$
   - *Voltage phase angles recovered by solving least-squares (LS) problem*:

$$\hat{\theta} = \arg\min_{\theta \in \mathbb{R}^{|\mathcal{V}|}} \sum_{(i,j)\in\mathcal{E}} (\theta_i - \theta_j - \hat{\theta}_{ij})^2 \tag{4.5}$$

If step 1 is able to recover the true state, then step 2 will recover the true complex voltage vector [10]. In the case of corrupted and/or noisy data, it will be impossible to recover the true state in step 1, but it is stated in [10] that the propagation of error is not too great in step 2. Thus, the focus of this work for robust SE is on step 1, which we will call ℓ-SE (linearized SE) from this point forward.

Consider the case when we have sparse corruption $\mathbf{b}$ but no random noise $\mathbf{w} = \mathbf{0}$. The ℓ-SE problem would be to solve the L_1 minimization problem given in [10]:

$$\hat{x} = \min_{\mathbf{x},\mathbf{b}} ||\mathbf{b}||_1 \quad \text{s.t. } A\mathbf{x} + \mathbf{b} = \mathbf{y} \tag{4.6}$$

In the case when both random Gaussian noise and sparse corruption are present, one version of ℓ-SE problem would be to solve the LASSO problem given in [10]:

$$\hat{x} = \min_{\mathbf{x},\mathbf{b}} \frac{1}{2|\mathcal{M}|}||\mathbf{y} - A\mathbf{x} - \mathbf{b}||_2^2 + \lambda||\mathbf{b}||_1 \tag{4.7}$$

for some regularization parameter $\lambda > 0$ that promotes the sparsity of $\mathbf{b}$. As an alternative, the paper [26] proposes minimizing a Huber loss which is more robust to outliers.

4.2.2 Mutual Incoherence

Mutual coherence is a measure of the cross-correlation of the columns of a matrix $A \in \mathbb{R}^{m\times n}$, which is a powerful notion in the area of compressed sensing [46–48]. The authors of [10] propose a new metric, which they call "mutual incoherence," a measure of the alignment of two particular submatrices of the sensing matrix A, one related to the clean data and one related to the corrupted data. As it is proposed in [10], this metric relies on the knowledge of the support of the bad data vector $\mathbf{b}$, denoted as $\mathcal{B} \subset \mathcal{M}$. The mutual incoherence metric $\rho(\mathcal{B})$ is then defined as:

$$\rho(\mathcal{B}) \triangleq \left|\left| A_{\mathcal{B}^c}^{T\dagger} A_{\mathcal{B}}^T \right|\right|_\infty \tag{4.8}$$

where $\mathcal{B}^c \triangleq \mathcal{M} \setminus \mathcal{B}$, $A_{\mathcal{B}}$ is the submatrix of A with rows corresponding to $\mathcal{B}$, and $A_{\mathcal{B}^c}$ is the submatrix of A with rows corresponding to $\mathcal{B}^c$. The mutual incoherence metric (4.8)

is a measure of how correlated the measurements are in $\mathcal{B}$ to the variables determined by measurements in $\mathcal{B}^c$. If $\rho(\mathcal{B}) < 1$, then ℓ-SE recovers $\hat{\mathbf{x}}$ with small error from the true state as shown in [10].

We need to make a few assumptions about the matrix A in order to use the mutual incoherence metric to certify the robustness of the ℓ-SE problem.

Assumption 2 (Preconditioning of sensing matrix)**.** *Each row of A is normalized so that $||\boldsymbol{a}_i||_2 = 1, \ \forall i \in [m]$, where $\boldsymbol{a}_i$ is the i^{th} row of A.*

Assumption 3 (Lower eigenvalue condition)**.**

$$\min \left\{ \lambda_{\min} \left(A_{\mathcal{B}^c}^T A_{\mathcal{B}^c} \right), \lambda_{\min} \left(\begin{bmatrix} A \\ I_{\mathcal{B}} \end{bmatrix} \begin{bmatrix} A^T & I_{\mathcal{B}}^T \end{bmatrix} \right) \right\} > 0 \tag{4.9}$$

where $I_{\mathcal{B}}$ corresponds to a submatrix formed by the $\mathcal{B}$ rows of the identity matrix $I \in \mathbb{R}^{m \times n}$ and $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue of a matrix.

This second assumption implies that the true vector must be identifiable if the bad data support $\mathcal{B}$ were known. The authors of [10] show that under these assumptions on A, if $\rho(\mathcal{B}) < 1$, then problem (4.7) with a given choice of regularization parameter λ recovers an estimated state with a small error from the true state as well as a large degree of bad data detection with high probability. However, because this method relies on knowledge of the support of the bad data vector, its application is limited.

The paper [26] builds on [10] and proposes a way to avoid using the bad data support, by developing a method for certification which can be ensured locally for each line in the network $(i, j) \in \mathcal{E}$ without considering the actual attack set. This method partitions the graph into attack, boundary, and safe regions for a given line $(i, j) \in \mathcal{E}$ and then looks at the mutual incoherence metric defined on subsections of the partitioned boundary measurements, which are fixed for a given line $(i, j) \in \mathcal{E}$ and measurement set $\mathcal{M}$. During an actual attack, if measurements at a node i are attacked and if every line $(i, j) \in \mathcal{E}$ attached to node i satisfies the mutual incoherence condition, then the attack will not propagate through the network.

In the next section, we present a modified version of the graph partitioning that was first introduced in [26]. While [26] partitions based on k^{th}-connected neighbors in the network, this method partitions through variable coupling in the sensing matrix and thus takes into account the choice of measurements to determine the variable partition. Unlike that in [26], our method results in the minimum number of boundary variables and maximum number of safe variables and measurements. This version is effectively the same as that in [26], i.e. it does not change the mutual incoherence metric or results of [26], but it streamlines the partitioning process and results in a more intuitive partition for the application.

4.3 Graph Partitioning for Local Certification

For a given line of attack $i \to j$, we aim to partition the set of state variables $\mathcal{X}$ into the sets of attacked variables $\mathcal{X}_{\text{a}}^{ij}$, boundary variables $\mathcal{X}_{\text{b}}^{ij}$, and safe variables $\mathcal{X}_{\text{s}}^{ij}$, where we use

the superscript ij to indicate that the partition is specific to the chosen attack line $i \to j$. It is desirable to partition the measurement sets into the following subsets:

$\mathcal{M}_{\mathrm{a}}^{ij}$	attacked measurements	$\to$ depend only on $\mathcal{X}_{\mathrm{a}}^{ij}$
$\mathcal{M}_{\mathrm{db}}^{ij}$	dependent boundary measurements	$\to$ depend on both $\mathcal{X}_{\mathrm{a}}^{ij}$ and $\mathcal{X}_{\mathrm{b}}^{ij}$
$\mathcal{M}_{\mathrm{ib}}^{ij}$	independent boundary measurements	$\to$ depend only on $\mathcal{X}_{\mathrm{b}}^{ij}$
$\mathcal{M}_{\mathrm{s}}^{ij}$	safe measurements	$\to$ all remaining measurements, can depend on both $\mathcal{X}_{\mathrm{s}}^{ij}$ and $\mathcal{X}_{\mathrm{b}}^{ij}$

We note that the "independent" and "dependent" boundary measurements are defined as dependent in relation to the attacked variables $\mathcal{X}_{\mathrm{a}}^{ij}$. The algorithm to formulate the variable and measurement partitions is given in Algorithm 4.

Algorithm 4 Sensing matrix partition for local attack $i \to j$

Inputs: $\mathcal{N}(\mathcal{V}, \mathcal{E})$, $\mathcal{M}$, $\mathcal{X}$, (i, j)
Compute sensing matrix A from $\mathcal{N}(\mathcal{V}, \mathcal{E})$
Set $\mathcal{X}_{\mathrm{a}}^{ij} \leftarrow \{x_i^{\mathrm{mg}}, x_{ij}^{\mathrm{re}}, x_{ij}^{\mathrm{im}}\}$
Set $\mathcal{M}_{\mathrm{a}}^{ij} \leftarrow \{\text{all-zero rows of } A[:, (\mathcal{X} \setminus \mathcal{X}_{\mathrm{a}}^{ij})]\}$
Set $\mathcal{M}_{\mathrm{db}}^{ij} \leftarrow \{\text{non-zero rows of } A[:, \mathcal{X}_{\mathrm{a}}^{ij}]\} \setminus \mathcal{M}_{\mathrm{a}}^{ij}$
Set $\mathcal{X}_{\mathrm{b}}^{ij} \leftarrow \{\text{non-zero columns of } A[\mathcal{M}_{\mathrm{db}}^{ij}, :]\} \setminus \mathcal{X}_{\mathrm{a}}$
Set $\mathcal{M}_{\mathrm{ib}}^{ij} \leftarrow \{\text{non-zero rows of } A[:, \mathcal{X}_{\mathrm{b}}^{ij}]\} \setminus \mathcal{M}_{\mathrm{db}}^{ij}$
Set $\mathcal{M}_{\mathrm{s}}^{ij} \leftarrow \mathcal{M} \setminus (\mathcal{M}_{\mathrm{a}}^{ij} \cup \mathcal{M}_{\mathrm{db}}^{ij} \cup \mathcal{M}_{\mathrm{ib}}^{ij})$
Set $\mathcal{X}_{\mathrm{s}}^{ij} \leftarrow \mathcal{X} \setminus (\mathcal{X}_{\mathrm{a}}^{ij} \cup \mathcal{X}_{\mathrm{b}}^{ij})$
Outputs: $\{\mathcal{X}_{\mathrm{a}}^{ij}, \mathcal{X}_{\mathrm{b}}^{ij}, \mathcal{X}_{\mathrm{s}}^{ij}\}$, $\{\mathcal{M}_{\mathrm{a}}^{ij}, \mathcal{M}_{\mathrm{db}}^{ij}, \mathcal{M}_{\mathrm{ib}}^{ij}, \mathcal{M}_{\mathrm{s}}^{ij}\}$

With this partition, we can rewrite the sensing matrix A as coupled through the boundary region:

$$A = \begin{bmatrix} A_{\mathcal{M}_{\mathrm{a}}^{ij}, \mathcal{X}_{\mathrm{a}}^{ij}} & 0 & 0 \\ A_{\mathcal{M}_{\mathrm{db}}^{ij}, \mathcal{X}_{\mathrm{a}}^{ij}} & A_{\mathcal{M}_{\mathrm{db}}^{ij}, \mathcal{X}_{\mathrm{b}}^{ij}} & 0 \\ 0 & A_{\mathcal{M}_{\mathrm{ib}}^{ij}, \mathcal{X}_{\mathrm{b}}^{ij}} & 0 \\ 0 & A_{\mathcal{M}_{\mathrm{s}}^{ij}, \mathcal{X}_{\mathrm{b}}^{ij}} & A_{\mathcal{M}_{\mathrm{s}}^{ij}, \mathcal{X}_{\mathrm{s}}^{ij}} \end{bmatrix} \tag{4.10}$$

If the matrix A satisfies some mutual incoherence condition for independent and dependent boundary measurement sets given by the partition in Algorithm 4, then line $i \to j$ is robust and bad data cannot propagate from i to j. In this case, if node i is part of the unknown attack set, then it will still be possible to recover a reasonable estimate of the state at node j with high probability. The required local mutual incoherence condition is given as:

$$\rho^{ij} \triangleq \left|\left| A^{T\dagger}_{\mathcal{M}_{\mathrm{ib}}^{ij}, \mathcal{X}_b^{ij}} A^{T}_{\mathcal{M}_{\mathrm{db}}^{ij}, \mathcal{X}_b^{ij}} \right|\right|_\infty < 1 \tag{4.11}$$

Proof. Same as in [26]. □

We can see that condition (4.11) depends on the measurement-variable partition. In this case, the mutual incoherence ρ^{ij} captures the alignment between measurements in the independent boundary set and the dependent boundary set. This condition ensures that attacked measurements do not propagate from the dependent boundary set to the independent boundary set.

Because condition (4.11) depends on the measurement set, it is apparent that we can optimize the choice of measurements $\mathcal{M}$ in order to decrease ρ^{ij} with the goal of finding measurements such that $\rho^{ij} < 1$. If we can find a measurement set $\mathcal{M}$ such that $\rho^{ij} < 1$ for all $i \to j$ and $j \to i$ for $(i,j) \in \mathcal{E}$, then we can say that the network is fully robust. If the network is fully robust, then we can find good estimates for local recovery of the safe and boundary region state variables via the method in [26]. In order to formalize the goal of placing sensors in a power network so that the network is robust, we will consider this mutual incoherence condition in an optimization framework, as presented in the next section.

4.4 Problem Formulation

In this section, we present the measurement choice MILP and its LP and SDP relaxations.

4.4.1 Formulation of the Measurement Choice MILP

The goal is to find a minimum choice of measurements over the network such that the mutual incoherence condition is satisfied for all boundary measurement sets $\{\mathcal{M}_{\text{db}}^{ij}, \mathcal{M}_{\text{ib}}^{ij}\}$ in both $i \to j$ and $j \to i$ directions for every line $(i,j) \in \mathcal{E}$. Note that in the formulations below we will use the notation $(i,j) \in \mathcal{E}$ to denote lines in both $i \to j$ and $j \to i$ directions. Let $\phi \in \{0,1\}^m$ be a binary vector which indicates the choice of measurements such that $\phi_i = 1$ if measurement $i \in [m]$ is chosen and $\phi_i = 0$ otherwise. Note that m is equal to the total possible number of measurements for a given power network.

When we consider the mutual incoherence condition across a line $i \to j$, we can define a partition of all possible measurements $\tilde{\mathcal{M}}$, which is invariant to the choice of measurements ϕ and depends only on the graph topology. Given this partition, let $\tilde{\mathcal{M}}_{\text{db}}^{ij}$ be the set of total possible dependent boundary measurements and $\tilde{\mathcal{M}}_{\text{ib}}^{ij}$ be the set of total possible independent boundary measurements, where the dependency is defined in relation to the attacked variables as in Section 4.3. Let $m_{\text{db}}^{ij} \triangleq |\tilde{\mathcal{M}}_{\text{db}}^{ij}|$, $m_{\text{ib}}^{ij} \triangleq |\tilde{\mathcal{M}}_{\text{ib}}^{ij}|$, and $n_{\text{b}}^{ij} \triangleq |\mathcal{X}_{\text{b}}^{ij}|$. We could formulate an optimization problem with condition (4.11) as a constraint. However, this problem may be infeasible if the constraints (4.11) cannot be satisfied for all lines $(i,j) \in \mathcal{E}$. Thus, it is more useful to consider the following mixed-integer nonlinear program (MINLP):

$$\min_{\substack{\beta \in \mathbb{R}, \phi \in \{0,1\}^m \\ X^{ij}, E^{ij}, J^{ij}, \forall (i,j) \in \mathcal{E}}} \quad \beta \tag{4.12a}$$

$$\text{subject to: } \underline{M} \leq \sum\nolimits_{i=1}^{m} \phi_i \leq \overline{M} \tag{4.12b}$$

$$\left(R^{ij} \odot E^{ij}\right) X^{ij} = S^{ij} \odot J^{ij} \qquad \forall (i,j) \in \mathcal{E} \tag{4.12c}$$

$$E_k^{ij} = \phi[\tilde{\mathcal{M}}_{\text{ib}}^{ij}(k)]\,\mathbf{1}_{n_{\text{b}}^{ij}} \qquad \forall k \in [m_{\text{ib}}^{ij}],\ \forall (i,j) \in \mathcal{E} \tag{4.12d}$$

$$J_k^{ij} = \phi[\tilde{\mathcal{M}}_{\text{db}}^{ij}(k)]\,\mathbf{1}_{n_{\text{b}}^{ij}} \qquad \forall k \in [m_{\text{db}}^{ij}],\ \forall (i,j) \in \mathcal{E} \tag{4.12e}$$

$$||X^{ij}||_\infty \leq \beta \qquad \forall (i,j) \in \mathcal{E} \tag{4.12f}$$

where $R^{ij} \triangleq A^T_{\tilde{\mathcal{M}}_{\text{ib}}^{ij},\mathcal{X}_{\text{b}}^{ij}} \in \mathbb{R}^{n_{\text{b}}^{ij} \times m_{\text{ib}}^{ij}}$ and $S^{ij} \triangleq A^T_{\tilde{\mathcal{M}}_{\text{db}}^{ij},\mathcal{X}_{\text{b}}^{ij}} \in \mathbb{R}^{n_{\text{b}}^{ij} \times m_{\text{db}}^{ij}}$ are subsets of the transposed sensing matrix A^T. We have introduced the variable $X^{ij} \in \mathbb{R}^{m_{\text{ib}}^{ij} \times m_{\text{db}}^{ij}}$ in order to represent the mutual incoherence as $||X^{ij}||_\infty$ for each line $(i,j) \in \mathcal{E}$. The matrix variables $E^{ij} \in \mathbb{R}^{n_{\text{b}}^{ij} \times m_{\text{ib}}^{ij}}$ and $J^{ij} \in \mathbb{R}^{n_{\text{b}}^{ij} \times m_{\text{db}}^{ij}}$ are used to choose columns of the sensing matrix corresponding respectively to independent and dependent boundary measurements. E_k^{ij} and J_k^{ij} represent the k^{th} columns of E^{ij} and J^{ij}, respectively. The notation $\phi[\tilde{\mathcal{M}}_{\text{ib}}^{ij}(k)]$ represents the element of ϕ corresponding to the k^{th} entry of $\tilde{\mathcal{M}}_{\text{ib}}^{ij}$ (similarly for $\phi[\tilde{\mathcal{M}}_{\text{db}}^{ij}(k)]$). The given parameters $\underline{M}$ and $\overline{M}$ are respectively the minimum and maximum numbers of measurements, where we select $\underline{M}$ such that $\underline{M} \geq n$.

Theorem 6. *If the objective of (4.12) is strictly less than 1, then a measurement set can be found such that the network is robust in terms of the mutual incoherence condition (4.11).*

Proof. Using equations (4.12d) and (4.12e), we have that $R^{ij} \odot E^{ij}$ is equivalent to $A^T_{\mathcal{M}_{\text{ib}}^{ij},\mathcal{X}_{\text{b}}^{ij}}$ and $S^{ij} \odot J^{ij}$ is equivalent to $A^T_{\mathcal{M}_{\text{db}}^{ij},\mathcal{X}_{\text{b}}^{ij}}$, thus $X^{ij} = A^{T\dagger}_{\mathcal{M}_{\text{ib}}^{ij},\mathcal{X}_{\text{b}}^{ij}} A^T_{\mathcal{M}_{\text{db}}^{ij},\mathcal{X}_{\text{b}}^{ij}}$ by constraint (4.12c). We have that $||X^{ij}||_\infty$ corresponds to ρ^{ij} as defined in Equation (4.11), and if $\beta < 1$ then (4.12f) enforces that ρ^{ij} is under 1 for every line $(i,j) \in \mathcal{E}$. □

Note that Problem (4.12) is nonconvex due to both the discrete nature of the binary variables ϕ and the nonlinearity of the $E^{ij}X^{ij}$ term in constraint (4.12c). If we examine the constraint (4.12c) for some line $(i,j) \in \mathcal{E}$, we see that it is equivalent to:

$$\sum\nolimits_{r=1}^{m_{\text{ib}}^{ij}} R_{kr}^{ij} X_{rl}^{ij} \phi[\tilde{\mathcal{M}}_{\text{ib}}^{ij}(r)] = S_{kl}^{ij} \phi[\tilde{\mathcal{M}}_{\text{db}}^{ij}(l)], \qquad \forall k \in [n_{\text{b}}^{ij}],\ \forall l \in [m_{\text{db}}^{ij}] \tag{4.13}$$

We can relax the nonconvexity due to the nonlinearity by introducing new variables:

$$Z_{rl}^{ij} \triangleq X_{rl}^{ij} \phi[\tilde{\mathcal{M}}_{\text{ib}}^{ij}(r)] \in \mathbb{R}, \qquad \forall r \in [m_{\text{ib}}^{ij}],\ \forall l \in [m_{\text{db}}^{ij}] \tag{4.14}$$

Then we can reformulate (4.13) with linear relations:

$$\sum_{r=1}^{m_{\text{ib}}^{ij}} R_{kr}^{ij} Z_{rl}^{ij} = S_{kl}^{ij} \phi[\tilde{\mathcal{M}}_{\text{db}}^{ij}(l)], \qquad \forall k \in [n_{\text{b}}^{ij}],\ \forall l \in [m_{\text{db}}^{ij}] \tag{4.15}$$

With this reformulation, all the nonlinearity is in the constraints (4.14). If we relax (4.14), we have:

$$Z_{rl}^{ij} \leq X_{rl}^{ij} \phi[\tilde{\mathcal{M}}_{\text{ib}}^{ij}(r)], \qquad \forall r \in [m_{\text{ib}}^{ij}],\ \forall l \in [m_{\text{db}}^{ij}] \tag{4.16}$$

We also note that $X_{rl}^{ij} = Z_{rl}^{ij}$. If $\phi[\tilde{\mathcal{M}}_{\text{ib}}^{ij}(r)] = 1$, this is obvious. If $\phi[\tilde{\mathcal{M}}_{\text{ib}}^{ij}(r)] = 0$, then the only constraint X_{rl}^{ij} appears in is (4.12f), and since we are minimizing the infinity norm of X^{ij}, we have that X_{rl}^{ij} will be equal to zero. Thus, we can substitute Z^{ij} into (4.12f) and (4.16) in place of X^{ij}. We can also reformulate constraint (4.16) using the big-M method by introducing some large constant $C > 0$ such that $Z_{rl}^{ij} \leq C$ for all $r \in [m_{\text{ib}}^{ij}]$, $l \in [m_{\text{db}}^{ij}]$, for all $(i,j) \in \mathcal{E}$ to yield the constraints:

$$Z_{rl}^{ij} \leq C\phi[\tilde{\mathcal{M}}_{\text{ib}}^{ij}(r)], \qquad \forall r \in [m_{\text{ib}}^{ij}],\ \forall l \in [m_{\text{db}}^{ij}] \tag{4.17}$$

To reformulate the constraint (4.12f) in order to yield a MILP, we introduce a new variable Y_{rl}^{ij} corresponding to $|Z_{rl}^{ij}|$ for all $r \in [m_{\text{ib}}^{ij}]$ and $l \in [m_{\text{db}}^{ij}]$ which can be related to Z_{rl}^{ij} by the following constraints:

$$Y_{rl}^{ij} \geq \max\{-Z_{rl}^{ij}, Z_{rl}^{ij}\}, \qquad \forall r \in [m_{\text{ib}}^{ij}],\ \forall l \in [m_{\text{db}}^{ij}] \tag{4.18}$$

We modify (4.17) to be upper bounds on Y^{ij}:

$$Y_{rl}^{ij} \leq C\phi[\tilde{\mathcal{M}}_{\text{ib}}^{ij}(r)], \qquad \forall r \in [m_{\text{ib}}^{ij}],\ \forall l \in [m_{\text{db}}^{ij}] \tag{4.19}$$

This formulation allows (4.12f) to be recast in terms of Y^{ij}:

$$\sum\nolimits_{l=1}^{m_{\text{db}}^{ij}} Y_{rl}^{ij} \leq \beta, \qquad \forall r \in [m_{\text{ib}}^{ij}] \tag{4.20}$$

In order for the power flow solution to be fully defined in the case without noise, i.e. $\mathbf{m} = A\mathbf{x}$, we need A to be full rank, as the authors suggest in [10]. Instead of enforcing the rank constraint in this optimization problem, we can enforce a weaker constraint which says that every variable must appear in at least one of the measurement equations. We can model this by taking $\phi^x \in \mathbb{R}^m$ to be the indicator variables corresponding to the set of measurements that depend on the variable $x \in \mathcal{X}$ defined as:

$$\phi_i^x \triangleq \begin{cases} \phi_i & \text{if measurement } i \text{ depends on variable } x \\ 0 & \text{otherwise} \end{cases} \tag{4.21}$$

where ϕ_i^x corresponds to the i^{th} element of ϕ^x for $x \in \mathcal{X}$.

The variables ϕ^x for all $x \in \mathcal{X}$ are defined based on the structure of the graph and therefore can easily be incorporated into the constraints. To enforce that every variable appears at least once in the chosen measurement equations, we use the constraints:

$$\sum\nolimits_{i=1}^{m} \phi_i^x \geq 1, \qquad \forall x \in \mathcal{X} \tag{4.22}$$

Combining these constraints, we have the MILP of interest:

$$\min_{\substack{\beta\in\mathbb{R},\phi\in\{0,1\}^m \\ Z^{ij},Y^{ij},\forall(i,j)\in\mathcal{E}}} \quad \beta \tag{4.23a}$$

$$\text{s.t. } \underline{M} \leq \sum\nolimits_{i=1}^{m} \phi_i \leq \overline{M} \tag{4.23b}$$

$$\sum\nolimits_{i=1}^{m} \phi_i^x \geq 1 \qquad \forall x \in \mathcal{X} \tag{4.23c}$$

$$\sum_{r=1}^{m_{\text{ib}}^{ij}} R_{kr}^{ij} Z_{rl}^{ij} = S_{kl}^{ij} \phi[\tilde{\mathcal{M}}_{\text{db}}^{ij}(l)] \qquad \forall k \in [n_{\text{b}}^{ij}], \forall l \in [m_{\text{db}}^{ij}], \forall (i,j) \in \mathcal{E} \tag{4.23d}$$

$$Y_{rl}^{ij} \geq \max\{-Z_{rl}^{ij}, Z_{rl}^{ij}\} \qquad \forall r \in [m_{\text{ib}}^{ij}], \forall l \in [m_{\text{db}}^{ij}], \forall (i,j) \in \mathcal{E} \tag{4.23e}$$

$$Y_{rl}^{ij} \leq C\phi[\tilde{\mathcal{M}}_{\text{ib}}^{ij}(r)] \qquad \forall r \in [m_{\text{ib}}^{ij}], \forall l \in [m_{\text{db}}^{ij}], \forall (i,j) \in \mathcal{E} \tag{4.23f}$$

$$\sum\nolimits_{l=1}^{m_{\text{db}}^{ij}} Y_{rl}^{ij} \leq \beta \qquad \forall r \in [m_{\text{ib}}^{ij}], \forall (i,j) \in \mathcal{E} \tag{4.23g}$$

For Problem (4.23), it may be impossible to recover a set of measurements that yields $\beta < 1$ (if the mutual incoherence condition (4.11) cannot be satisfied for all lines $(i,j) \in \mathcal{E}$). Thus, it will be more helpful to minimize the number of violations of the mutual incoherence condition, i.e. where $||Y^{ij}||_\infty \geq 1$. We can do this by solving the related MIP:

$$\min_{\substack{\phi\in\{0,1\}^m \\ Z^{ij},Y^{ij},\beta^{ij},\forall(i,j)\in\mathcal{E}}} \quad \sum_{(i,j)\in\mathcal{E}} \mathbf{1}\{\beta^{ij} \geq 1\} \tag{4.24a}$$

$$\text{s.t. } \underline{M} \leq \sum\nolimits_{i=1}^{m} \phi_i \leq \overline{M} \tag{4.24b}$$

$$\sum\nolimits_{i=1}^{m} \phi_i^x \geq 1 \qquad \forall x \in \mathcal{X} \tag{4.24c}$$

$$\sum_{r=1}^{m_{\text{ib}}^{ij}} R_{kr}^{ij} Z_{rl}^{ij} = S_{kl}^{ij} \phi[\tilde{\mathcal{M}}_{\text{db}}^{ij}(l)] \qquad \forall k \in [n_{\text{b}}^{ij}], \forall l \in [m_{\text{db}}^{ij}], \forall (i,j) \in \mathcal{E} \tag{4.24d}$$

$$Y_{rl}^{ij} \geq \max\{-Z_{rl}^{ij}, Z_{rl}^{ij}\} \qquad \forall r \in [m_{\text{ib}}^{ij}], \forall l \in [m_{\text{db}}^{ij}], \forall (i,j) \in \mathcal{E} \tag{4.24e}$$

$$Y_{rl}^{ij} \leq C\phi[\tilde{\mathcal{M}}_{\text{ib}}^{ij}(r)] \qquad \forall r \in [m_{\text{ib}}^{ij}], \forall l \in [m_{\text{db}}^{ij}], \forall (i,j) \in \mathcal{E} \tag{4.24f}$$

$$\sum\nolimits_{l=1}^{m_{\text{db}}^{ij}} Y_{rl}^{ij} \leq \beta^{ij} \qquad \forall r \in [m_{\text{ib}}^{ij}] \tag{4.24g}$$

which is converted to a MILP by introducing binary variables α^{ij} corresponding to the indicators $\mathbf{1}\{\beta^{ij} \geq 1\}$ and using the big-M method to recast the constraints in linear form.

Because existing MILP methods do not scale well with the number of binary variables, we can see that solving Problem (4.23) or (4.24) on a large network will be computationally burdensome. Thus, we can always fix some of the measurements and consider a subset of measurement choices to modify. Even by just considering some of the measurements, we can improve the network robustness for SE in terms of mutual incoherence. Another possible

method for dealing with the mixed-integer formulation is to consider a relaxation of the binary variables, as discussed in the following section.

4.4.2 Formulation of Measurement Choice Relaxations

Instead of solving the MILPs in (4.23) or (4.24) which involve nonconvexity in the form of the binary variables $\phi \in \{0,1\}^m$ (and $\alpha^{ij} \in \{0,1\}$ for all $(i,j) \in \mathcal{E}$ for (4.24)), we can solve convex relaxations of the problems. The relaxation of (4.23) into an LP is given by solving for $\phi \in [0,1]^m$, and the relaxation of (4.24) into an LP is given by solving for $\phi \in [0,1]^m$ and $\alpha^{ij} \in [0,1]$ for all $(i,j) \in \mathcal{E}$. For the SDP relaxation of (4.23), we note that the binary constraint $\phi \in \{0,1\}^m$ is equivalent to:

$$\phi_i(\phi_i - 1) = 0, \qquad \forall i \in \{1, \dots, m\} \tag{4.25}$$

If we define $\Phi \triangleq \phi\phi^T \in \mathbb{S}^m$, then constraint (4.25) becomes:

$$\Phi_{ii} - \phi_i = 0, \qquad \forall i \in \{1, \dots, m\} \tag{4.26}$$

Then, SDP relaxation of $\Phi = \phi\phi^T$ is given by:

$$\Phi \succeq \phi\phi^T \tag{4.27}$$

Combining these, we have the SDP relaxation of (4.23):

$$\begin{aligned}
\min_{\substack{\beta\in\mathbb{R} \\ Z^{ij},Y^{ij},\forall(i,j)\in\mathcal{E} \\ \phi\in[0,1]^m,\Phi\in\mathbb{S}^m}} \quad & \beta && && \text{(4.28a)}\\
\text{s.t.} \quad & \underline{M} \le \textstyle\sum_{i=1}^m \phi_i \le \overline{M} && && \text{(4.28b)}\\
& \textstyle\sum_{i=1}^m \phi_i^x \ge 1 && \forall x \in \mathcal{X} && \text{(4.28c)}\\
& \Phi_{ii} - \phi_i = 0 && \forall i \in [m] && \text{(4.28d)}\\
& \Phi \succeq \phi\phi^T && && \text{(4.28e)}\\
& \sum_{r=1}^{m_{\text{ib}}^{ij}} R_{kr}^{ij} Z_{rl}^{ij} = S_{kl}^{ij} \phi[\tilde{\mathcal{M}}_{\text{db}}^{ij}(l)] && \forall k \in [n_{\text{b}}^{ij}], \forall l \in [m_{\text{db}}^{ij}], \forall (i,j) \in \mathcal{E} && \text{(4.28f)}\\
& Y_{rl}^{ij} \ge \max\{-Z_{rl}^{ij}, Z_{rl}^{ij}\} && \forall r \in [m_{\text{ib}}^{ij}], \forall l \in [m_{\text{db}}^{ij}], \forall (i,j) \in \mathcal{E} && \text{(4.28g)}\\
& Y_{rl}^{ij} \le C\phi[\tilde{\mathcal{M}}_{\text{ib}}^{ij}(r)] && \forall r \in [m_{\text{ib}}^{ij}], \forall l \in [m_{\text{db}}^{ij}], \forall (i,j) \in \mathcal{E} && \text{(4.28h)}\\
& \textstyle\sum_{l=1}^{m_{\text{db}}^{ij}} Y_{rl}^{ij} \le \beta && \forall r \in [m_{\text{ib}}^{ij}], \forall (i,j) \in \mathcal{E} && \text{(4.28i)}
\end{aligned}$$

Similarly, for the α^{ij} variables in (4.24), we can define a matrix inequality of the form $A \succeq \alpha\alpha^T$ for some matrix $A \in \mathbb{S}^{|\mathcal{E}|}$ and the corresponding SDP problem of (4.24).

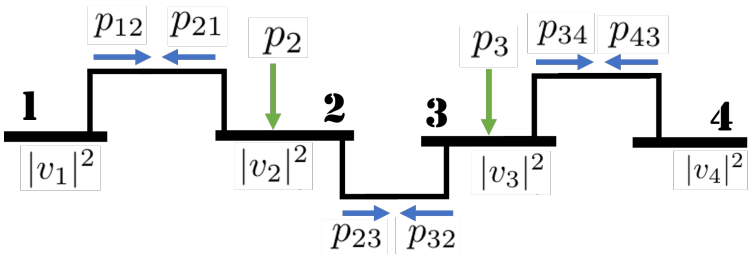

Figure 4.2: Four bus network showing all possible voltage magnitude measurements, real power flow measurements, and real power injection measurements. Note that because real power injections at buses 1 and 4 are equivalent to p_{12} and p_{43} respectively, these measurements are not considered.

4.5 Simulations

The simulations are run on a standard laptop using the Pyomo modeling language in Python 3.8. The MILPs given by (4.23) and (4.24) are solved with the Gurobi solver, which uses a branch-and-bound method to determine the binary variables. Note that for all examined test cases both (4.23) and (4.24) yield sensing matrices that are full rank.

4.5.1 Four-Bus Test Case

We first consider the four-bus test network shown in Figure 4.2. In [26], the authors considered this network and showed that different combinations of measurement choices yielded mutual incoherence metrics that were greater than 1 for certain lines in the network. By considering Problems (4.23) and (4.24), we formalize their guess-and-check process.

For the four-bus network, the set of variables is given by

$$\mathcal{X} = \{x_1^{\text{mg}}, x_2^{\text{mg}}, x_3^{\text{mg}}, x_4^{\text{mg}}, x_{12}^{\text{re}}, x_{12}^{\text{im}}, x_{23}^{\text{re}}, x_{23}^{\text{im}}, x_{34}^{\text{re}}, x_{34}^{\text{im}}\} \tag{4.29}$$

And the set of all possible measurements is given by:

$$\tilde{\mathcal{M}} = \{|v_1|^2, |v_2|^2, |v_3|^2, |v_4|^2, p_{12}, q_{12}, p_{21}, q_{21}, p_2, q_2, p_{23}, q_{23}, p_{32}, q_{32}, p_3, q_3, p_{34}, q_{34}, p_{43}, q_{43}\} \tag{4.30}$$

We take the line parameters to be $G_{ij} = 5$, $B_{ij} = -20$, and $B_{ij}^{\text{sh}} = 0.5$ in per unit values. If we set $\underline{M} = 3|\mathcal{V}| - 2 = 10$, and $\overline{M} = m = 20$, then for the four-bus network, we find that it is impossible to recover a set of measurements such that the mutual incoherence condition is satisfied in both directions for every line $(i, j) \in \mathcal{E}$, as shown in the second column of Table 4.1. The choice of measurement set given by (4.23) is

$$\mathcal{M} = \{|v_1|^2, |v_2|^2, |v_3|^2, |v_4|^2, p_{12}, p_{43}, q_{12}, q_{23}, q_{34}, q_{21}, q_{32}, q_{43}, p_2, p_3\} \tag{4.31}$$

Table 4.1: Mutual incoherence metric for four-bus network

Line $i \to j$	Mutual incoherence ρ^{ij} from solving (4.23)	Mutual incoherence ρ^{ij} from solving (4.24)
$1 \to 2$	0.91	0.00
$2 \to 3$	1.02	0.90
$3 \to 4$	1.39	1.52
$2 \to 1$	1.39	1.52
$3 \to 2$	1.02	0.93
$4 \to 3$	0.91	0.19

which yields a full rank matrix A. The values for mutual incoherence over the network with this measurement set are given in the second column of Table 4.1. We can see that in this case, there is no choice of measurements which yields all mutual incoherence metrics below 1. Instead, we can solve (4.24) to yield a choice of measurements that minimizes the number of violations of the mutual incoherence metric. Using this problem, we find the optimal set of measurements to be:

$$\mathcal{M} = \{|v_1|^2, |v_2|^2, |v_3|^2, |v_4|^2, p_{12}, p_{23}, p_{32}, p_{43}, q_{12}, q_{23}, q_{32}, q_{43}\} \tag{4.32}$$

By solving (4.24), we see that it is possible to create a measurement set such that 2 out of 3 of the lines are robust in both directions, as shown in the third column of Table 4.1. We see that a mutual incoherence of 0 is obtained for line $1 \to 2$. This occurs because the chosen measurement set has no coupling between attack variables and the rest of the variables, resulting in $\mathcal{X}_{\mathrm{b}}^{12} = \emptyset$. Note that for this test case both (4.23) and (4.24) yield sensing matrices that are full rank, thus the network variables are fully defined in terms of the measurements.

In Figures 4.3 and 4.4, we visualize two attack scenarios and how the choice of measurement set affects the mutual incoherence condition, thereby affecting which states are recoverable during the attacks. From these simulated attacks, we can see that the measurement set in (4.32) is superior to the measurement set in (4.31) in terms of recovering the maximum unknown state in SE during a cyberattack.

4.5.2 IEEE Test Cases with MILP

Then, we solve Problem (4.23) for some IEEE test cases [49], finding that there is no choice of measurements such that mutual incoherence is below 1 for every line on the network. The results are given in Table 4.2. However, we can still solve problem (4.24) to yield the optimal choice of measurements for the mutual incoherence robustness condition. The results of (4.24) are given in Table 4.3. Note that if the data for a part of the network is under attack, having more lines satisfy the mutual incoherence condition guarantees a reduction in the impact of the attack on the SE for nodes far away from the attacked region [26].

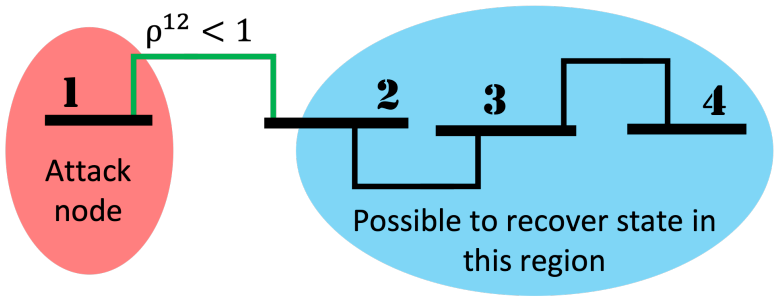

Figure 4.3: Visualization of data cyberattack on bus 1 in four-bus network. In this case, for either measurement set given by (4.31) or (4.32), we can contain the data attack to node 1 since $\rho^{12} < 1$ (see Table 4.1).

Table 4.2: Solution to (4.23) for various IEEE test cases

Network	Fraction of chosen measurements	Max ρ^{ij} for $(i,j) \in \mathcal{E}$	Solve time (s)
case5	29 / 39	1.26	0.69
case9	42 / 57	1.48	1.12
case14	95 / 120	1.61	9.33
case30	193 / 248	1.61	39.3

Table 4.3: Solution to (4.24) for various IEEE test cases

Network	Fraction of chosen measurements	Fraction of lines with $\rho^{ij} < 1$	Solve time (s)
case5	30 / 39	6 / 12	1.89
case9	36 / 57	12 / 18	1.49
case14	92 / 120	18 / 40	120.5
case30	190 / 248	37 / 82	831.1

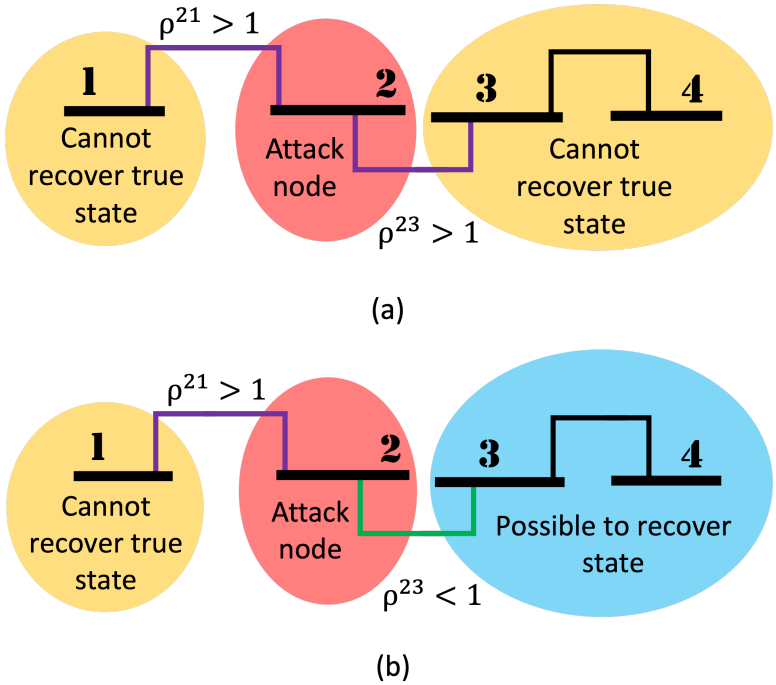

Figure 4.4: Visualization of data cyberattack on bus 2 in four-bus network. The top figure (a) corresponds to the measurement set (4.31), and the bottom figure (b) corresponds to the measurement set (4.32). The choice of measurement set affects the parameters ρ^{ij}, which in turn affect which data we can recover in case of cyberattack. In the case of a cyberattack on bus 2, we can only recover the true state for nodes 3 and 4 if we have chosen measurement set (4.32) and cannot recover the true state for any nodes if we have chosen measurement set (4.31). This is due to the fact that $\rho^{23} < 1$ for measurement set (4.32) and $\rho^{23} > 1$ for measurement set (4.31) as seen in Table 4.1.

4.5.3 Test Cases with Relaxations

Next, we test relaxations of the MILPs (4.23) and (4.24) into LPs and SDPs. In Table 4.4, we can see that the objective value of the SDP relaxation of (4.24) will be 0.0 for the four-bus network shown in Figure 4.2. We note that the solution ϕ to the SDP relaxation is fractional, which does not make physical sense when we consider that ϕ corresponds to placing or not placing a sensor. Thus, we will round the solution to find a binary $\tilde{\phi}$. However, there are no guarantees that the new ϕ will satisfy the constraints on the number of measurements or the dependencies on variables. For the four-bus network, the SDP with rounding method gives a good solution, shown in Table 4.5. Note that this solution satisfies the condition that A has full rank.

In Tables 4.6 and 4.7, we have the results from the MILPs (4.23) and (4.24) and their LP

Table 4.4: Solution to SDP relaxation of (4.24) on four-bus network

Line $i \to j$	β^{ij} from solving SDP relaxation of (4.24)
$1 \to 2$	0.70
$2 \to 3$	0.43
$3 \to 4$	0.86
$2 \to 1$	0.86
$3 \to 2$	0.43
$4 \to 3$	0.70

Table 4.5: Mutual incoherence metric for four-bus network from solving SDP relaxation of (4.24) and rounding ϕ

Line $i \to j$	Mutual incoherence ρ^{ij} from solving SDP of (4.24) and rounding ϕ
$1 \to 2$	0.00
$2 \to 3$	0.24
$3 \to 4$	0.93
$2 \to 1$	0.93
$3 \to 2$	0.24
$4 \to 3$	0.00

Table 4.6: Objective values of MILP (4.23) and its relaxations

Network	Obj. of MILP (4.23)	Obj. of LP of (4.23)	Obj. of SDP (4.28)
4-bus	1.39	0.458	0.458
case5	1.26	0.263	0.263
case9	1.48	0.334	0.318
case14	1.61	0.245	0.318
case30	1.61	0.271	0.271

and SDP relaxations on various networks. In Tables 4.8 and 4.9, we compare the solution times of the different methods. From these tables, we see that the LP and SDP are more computationally scalable than the MILP for larger networks, with the LP being the most efficient method for large networks.

4.6 Conclusions

This chapter presented an original framework for optimizing the choice of measurements in a power system to protect against false data injection. By examining a local metric for robust

Table 4.7: Comparison of solution quality for MILP (4.24) and its relaxations

Network	Fraction of lines with $\rho^{ij} < 1$ for MILP (4.24)	Fraction of lines with $\rho^{ij} < 1$ for LP of (4.24)	Fraction of lines with $\rho^{ij} < 1$ for SDP of (4.24)
4-bus	4 / 6	2 / 6	6 / 6
case5	6 / 12	6 / 12	6 / 12
case9	12 / 18	8 / 18	18 / 18
case14	18 / 40	18 / 40	10 / 40
case30	37 / 82	25 / 82	82 / 82

Table 4.8: Solve times for MILP (4.23) and its relaxations

Network	Sol. time of MILP (4.23)	Sol. time of LP of (4.23)	Sol. time of SDP (4.28)
4-bus	0.18	0.98	19.11
case5	0.69	7.69	87.58
case9	1.12	7.86	130.26
case14	9.33	56.50	731.13
case30	39.3	168.89	2941.22

Table 4.9: Solve times for MILP (4.24) and its relaxations

Network	Sol. time of MILP (4.24)	Sol. time of LP of (4.24)	Sol. time of SDP of (4.24)
4-bus	0.11	0.89	4.23
case5	1.89	7.18	11.64
case9	1.49	7.79	17.26
case14	120.5	55.29	82.33
case30	831.1	181.68	607.16

SE, we were able to define a coupled optimization problem over all lines of the network. We showed that for some test cases, there is no choice of measurements such that every line can be certified as robust in both directions. However, this framework allows us to find subsets of measurements that are more optimal than others in terms of SE robustness. Having more lines satisfy the mutual incoherence condition guarantees a reduction in the impact of the attack on SE for nodes far from the attacked region.

This measurement choice method could be used to design new smart grids such that measurements are optimally placed for SE robustness. Another application could be to place new sensors in an existing legacy power network in order to improve SE robustness, by considering a subset of possible measurement placements. This method could also be used to identify measurements that are the most immune or most susceptible for error propagation in order to classify critical measurements.

4.7 Bibliography

[1] E. Glista and S. Sojoudi, "A MILP for optimal measurement choice in robust power grid state estimation," in *2022 IEEE Power & Energy Society General Meeting (PESGM)*, 2022, pp. 1–5.

[2] A. Abur and A. G. Exposito, *Power System State Estimation: Theory and Implementation*. CRC Press, 2004.

[3] A. Monticelli, "Electric power system state estimation," *Proceedings of the IEEE*, vol. 88, no. 2, pp. 262–282, 2000.

[4] F. C. Schweppe and J. Wildes, "Power system static-state estimation, part I: Exact model," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-89, no. 1, pp. 120–125, 1970.

[5] U.S.-Canada Power System Outage Task Force, *Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations*. U.S. Department of Energy, 2004.

[6] National Academies of Sciences, Engineering, and Medicine, *Enhancing the Resilience of the Nation's Electricity System*. National Academies Press, 2017.

[7] W. Wang and Z. Lu, "Cyber security in the smart grid: Survey and challenges," *Computer Networks*, vol. 57, p. 1344–1371, 04 2013.

[8] R. Deng, P. Zhuang, and H. Liang, "False data injection attacks against state estimation in power distribution systems," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 2871–2881, 2019.

[9] Zong-Han Yu and Wen-Long Chin, "Blind false data injection attack using PCA approximation method in smart grid," *IEEE Transactions on Smart Grid*, vol. 6, no. 3, pp. 1219 – 1226, May 2015.

[10] M. Jin, I. Molybog, R. Mohammadi-Ghazi, and J. Lavaei, "Scalable and robust state estimation from abundant but untrusted data," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 1880–1894, 2020.

[11] K. Madsen, H. Nielsen, and O. Tingleff, "Methods for non-linear least squares problems (2nd ed.)," p. 60, January 2004.

[12] W. F. Tinney and C. E. Hart, "Power flow solution by Newton's method," *IEEE Transactions on Power Apparatus and systems*, no. 11, pp. 1449–1460, 1967.

[13] R. van Amerongen, "A general-purpose version of the fast decoupled load flow," *IEEE Transactions on Power Systems*, vol. 4, no. 2, pp. 760–770, 1989.

[14] T. P. Vishnu, V. Viswan, and A. M. Vipin, "Power system state estimation and bad data analysis using weighted least square method," in *2015 International Conference on Power, Instrumentation, Control and Computing (PICC)*, 2015, pp. 1–5.

[15] E. J. Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes rendus mathematique*, vol. 346, no. 9-10, pp. 589–592, 2008.

[16] I. Molybog, S. Sojoudi, and J. Lavaei, "Role of sparsity and structure in the optimization landscape of non-convex matrix sensing," *Mathematical Programming*, pp. 1–37, 2020.

[17] R. Y. Zhang, J. Lavaei, and R. Baldick, "Spurious local minima in power system state estimation," *IEEE transactions on control of network systems*, vol. 6, no. 3, pp. 1086–1096, 2019.

[18] G. Wang, G. B. Giannakis, and J. Chen, "Robust and scalable power system state estimation via composite optimization," 2019.

[19] G. Wang, H. Zhu, G. B. Giannakis, and J. Sun, "Robust power system state estimation from rank-one measurements," *IEEE Transactions on Control of Network Systems*, vol. 6, no. 4, pp. 1391–1403, 2019.

[20] K. Li, "State estimation for power distribution system and measurement impacts," *IEEE Transactions on Power Systems*, vol. 11, no. 2, pp. 911–916, 1996.

[21] R. Madani, J. Lavaei, and R. Baldick, "Convexification of power flow equations in the presence of noisy measurements," *IEEE Transactions on Automatic Control*, vol. 64, no. 8, pp. 3101–3116, 2019.

[22] H. Zhu and G. B. Giannakis, "Power system nonlinear state estimation using distributed semidefinite programming," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 6, pp. 1039–1050, 2014.

[23] Y. Zhang, R. Madani, and J. Lavaei, "Conic relaxations for power system state estimation with line measurements," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 3, pp. 1193–1205, 2017.

[24] Y. Weng, Q. Li, R. Negi, and M. Ilić, "Semidefinite programming for power system state estimation," in *2012 IEEE Power and Energy Society General Meeting*. IEEE, 2012, pp. 1–8.

[25] R. Madani, M. Ashraphijuo, J. Lavaei, and R. Baldick, "Power system state estimation with a limited number of measurements," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, 2016, pp. 672–679.

[26] M. Jin, J. Lavaei, S. Sojoudi, and R. Baldick, "Boundary defense against cyber threat for power system state estimation," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 1752–1767, 2021.

[27] H. Wu and Giri, "PMU impact on state estimation reliability for improved grid security," in *2005/2006 IEEE/PES Transmission and Distribution Conference and Exhibition*, 2006, pp. 1349–1351.

[28] G. N. Korres and N. M. Manousakis, "State estimation and bad data processing for systems including PMU and SCADA measurements," *Electric Power Systems Research*, vol. 81, no. 7, pp. 1514–1524, 2011.

[29] A. Gomez-Exposito, A. Abur, P. Rousseaux, A. de la Villa Jaen, and C. Gomez-Quiles, "On the use of PMUs in power system state estimation," in *17th Power System Computation Conference*, 2011.

[30] Y. Weng, M. D. Ilić, Q. Li, and R. Negi, "Convexification of bad data and topology error detection and identification problems in AC electric power systems," *IET Generation, Transmission & Distribution*, vol. 9, no. 16, pp. 2760–2767, 2015.

[31] M. Young and A. Silverstein, "Factors affecting PMU installation costs," *US Department of Energy-Office of Electricity Delivery and Energy Reliability*, 2014.

[32] M. Netto, V. Krishnan, Y. Zhang, and L. Mili, "Measurement placement in electric power transmission and distribution grids: Review of concepts, methods, and research needs," *IET Generation, Transmission & Distribution*, vol. 16, no. 5, pp. 805–838, 2022.

[33] N. M. Manousakis, G. N. Korres, and P. S. Georgilakis, "Taxonomy of PMU placement methodologies," *IEEE Transactions on power Systems*, vol. 27, no. 2, pp. 1070–1077, 2012.

[34] N. H. Abbasy and H. M. Ismail, "A unified approach for the optimal PMU location for power system state estimation," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 806–813, 2009.

[35] T. L. Baldwin, L. Mili, M. B. Boisen, and R. Adapa, "Power system observability with minimal phasor measurement placement," *IEEE Transactions on Power systems*, vol. 8, no. 2, pp. 707–715, 1993.

[36] L. Mili, T. Baldwin, and R. Adapa, "Phasor measurement placement for voltage stability analysis of power systems," in *29th IEEE conference on decision and control*. IEEE, 1990, pp. 3033–3038.

[37] D. J. Brueni, *Minimal PMU placement for graph observability: A decomposition approach*. Virginia Polytechnic Institute and State University, 1993.

[38] S. Chakrabarti and E. Kyriakides, "Optimal placement of phasor measurement units for power system observability," *IEEE Transactions on power systems*, vol. 23, no. 3, pp. 1433–1440, 2008.

[39] N. M. Manousakis and G. N. Korres, "A weighted least squares algorithm for optimal PMU placement," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 3499–3500, 2013.

[40] A. Almunif and L. Fan, "Mixed integer linear programming and nonlinear programming for optimal PMU placement," in *2017 North American power symposium (NAPS)*. IEEE, 2017, pp. 1–6.

[41] M. Göl and A. Abur, "PMU placement for robust state estimation," in *2013 North American Power Symposium (NAPS)*, 2013, pp. 1–5.

[42] J. Hao, R. J. Piechocki, D. Kaleshi, W. H. Chin, and Z. Fan, "Sparse malicious false data injection attacks and defense mechanisms in smart grids," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 5, pp. 1–12, 2015.

[43] L. Liu, M. Esmalifalak, Q. Ding, V. A. Emesih, and Z. Han, "Detecting false data injection attacks on power grid by sparse optimization," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 612–621, 2014.

[44] F. Broussolle, "State estimation in power systems: Detecting bad data through the sparse inverse matrix method," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-97, no. 3, pp. 678–682, 1978.

[45] W. Xu, M. Wang, J.-F. Cai, and A. Tang, "Sparse error correction from nonlinear measurements with applications in bad data detection for power networks," *IEEE Transactions on Signal Processing*, vol. 61, no. 24, pp. 6175–6187, 2013.

[46] V. Abolghasemi, S. Ferdowsi, B. Makkiabadi, and S. Sanei, "On optimization of the measurement matrix for compressive sensing," in *2010 18th European Signal Processing Conference*. IEEE, 2010, pp. 427–431.

[47] M. Elad, "Optimized projections for compressed sensing," *IEEE Transactions on Signal Processing*, vol. 55, no. 12, pp. 5695–5702, 2007.

[48] L. C. Potter, E. Ertin, J. T. Parker, and M. Cetin, "Sparsity and compressed sensing in radar imaging," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1006–1020, 2010.

[49] "EGRET: Electrical grid research and engineering tools." [Online]. Available: https://github.com/grid-parity-exchange/Egret

Chapter 5

Physics-Informed Support Vector Regression for Power System Topology Identification

In this chapter, we present a constrained support vector regression (SVR) method that can be used to accurately recover a power network's unknown topology.[1]

5.1 Introduction

With the adoption of new smart grid technologies such as smart meters, distributed and renewable generation, and an increased prevalence of phasor measurement units (PMUs) in power networks, the optimization and monitoring methods of legacy power systems will need to evolve [2]. Additionally, as the risk of cyberattacks grows, many of the classical optimization problems such as state estimation (SE) take on increased importance in ensuring the reliability of the electric grid [3, 4]. The safe and effective operation of the power grid depends on solving a variety of optimization and control problems, including SE, power flow (PF), optimal power flow (OPF), unit commitment (UC), false data detection (FDD), and voltage control [5–10]. The fundamental bases for many of these problems are the nonlinear AC power flow equations, which define the physics of power flow in a network. Using the power flow equations in most applications relies on an awareness of the power network topology and line parameters. However, knowledge of the network topology and/or the line parameters can be limited due to cyberattacks, real-time topology switching, variable environmental conditions affecting material properties or other data collection deficiencies and inaccuracies [11, 12]. In order to deal with these uncertainties, we consider a system identification problem that aims to learn the power flow mapping. By learning the power flow mapping, the goal is to recover the true system parameters and topology.

[1] Chapter 5 includes materials from [1].

A variety of machine learning and optimization methods have been proposed to learn the power flow mapping in a generic network, including compressed sensing [13], maximum likelihood estimation [14], neural networks (NN) [15–20], and support vector regression (SVR) [11, 21]. Many of these data-driven methods exploit the abundance of power system data from both traditional SCADA measurements and PMUs to learn the mapping [16]. However, these black- and gray-box methods suffer from overfitting and lack a physical representation in the power network. These methods do not explicitly make use of the sparsity inherent in power networks, which has been effectively exploited in other power applications such as OPF to efficiently solve hard, nonconvex problems [22–25].

While recent papers show that SVR can be effective at learning forward and inverse mappings between power system inputs and outputs, these papers apply classic SVR methods that fail to recover the true system parameters because they do not account for power network sparsity [11, 21]. In [11], it is shown that the power flow equations can be exactly written as a quadratic kernel within the reproducing kernel Hilbert space (RKHS). However, we will demonstrate later that the feature vector corresponding this RKHS contains many features that do not contribute to the power flow equations and should be associated with parameters equal to zero. However, in [11] and [21], the authors' methods do not ensure that these parameters are zero and will thus recover a dense parameter set that has no realistic physical meaning. In this work, we fix these shortcomings in the existing literature by proposing a new constrained SVR method that considers the actual power network sparsity. We show that this method can recover the true physical parameters and topology of a power system with high accuracy.

5.1.1 Support Vector Regression (SVR) and Prior Knowledge

SVR was developed in the 1990s as an extension of the support vector machine (SVM) classification learning algorithm [26, 27]. The idea of SVM is to find a hyperplane decision boundary that maximizes the margin between differently classified sets of data [26]. The fast implementation of SVM in nonlinear settings relies on the kernel trick which maps nonlinear features into a high-dimensional space that corresponds to a linear classifier. SVR is the regression extension of SVM. While SVM aims to find a classification hyperplane that minimizes data proximity to the plane, SVR aims to find a linear estimator that maximizes data proximity around the estimator, penalizing data points outside of an ϵ-tube around the estimator [28]. Like SVM, SVR makes use of the kernel trick to efficiently estimate nonlinear functions.

SVM and SVR have been shown to be effective in a variety of nonlinear applications, including image classification [29–31] and load forecasting [32, 33]. However, SVR fails to take into account prior information known about the space of the estimator which may result in estimators that do not accurately represent the corresponding systems. For various types of estimators such as the least-squares estimator, some proposed methods to include prior knowledge in the estimator calculation include adding regularization terms such as a Lasso term to encourage constraint satisfaction [34–36]. Another method is to add constraints to

the estimation problem, such as in constrained least-squares estimation [37–39]. However, these constraints can often increase the difficulty of solving the problem, e.g. the least-squares estimator has a closed-form solution while the non-negative least-squares (NNLS) estimator does not although there exist many fast methods to solve the NNLS problem [40–42].

Recent research has considered a constrained SVR problem in the case where both the kernel and the constraints are linear [43]. The authors of [43] show that their constrained SVR method can recover better estimators in terms of root-mean-square error (RMSE) compared to both classic SVR and other constrained regression methods, such as constrained least squares, in various biomedical and weather data settings where information is known *a priori* about the space of the estimator. Building on the work in [43], we propose a constrained SVR method that incorporates a nonlinear kernel and linear constraints. While our main focus is the power flow mapping application, the proposed constrained SVR methodology could be applied to other network mapping problems as well as more general supervised learning settings in which information is known *a priori* about the system.

5.1.2 Contributions

In this chapter, we propose a data-driven approach to learn the true topology and parameters of a power network. By leveraging a modified SVR formulation, we show that sparsity-enforcing constraints allow the SVR model to capture the true physics of the power system. We show that the dual of the constrained SVR problem can be written as a quadratic program and solved with efficient convex methods. In simulations on realistic test cases, we show that our method outperforms existing state-of-the-art methods in terms of accuracy in recovering the system parameters and in terms of solution time.

5.1.3 Notations

The symbols $\mathbb{R}$ and $\mathbb{C}$ denote the sets of real and complex numbers, respectively. $\mathbb{R}^N$ and $\mathbb{C}^N$ denote the spaces of N-dimensional real and complex vectors, respectively. The symbol $\mathbb{R}^N_+$ denotes the space of real vectors with non-negative entries. The symbol $\mathbb{S}^N$ denotes the space of $N \times N$ symmetric real matrices. The symbols $(\cdot)^T$ and $(\cdot)^*$ denote the transpose and conjugate transpose of a vector or matrix. $\mathrm{Re}\{\cdot\}$ and $\mathrm{Im}\{\cdot\}$ denote the real and imaginary part of a given scalar or matrix. The symbol $|\cdot|$ is the absolute value operator if the argument is a scalar, vector, or matrix; otherwise, it is the cardinality of a measurable set. The imaginary unit is denoted by $\mathbf{j} = \sqrt{-1}$. Given a function $f(x, \cdot)$, $\nabla_x f(x, \cdot)$ and $\nabla_x^2 f(x, \cdot)$ denote the Jacobian and Hessian of f with respect to x, respectively.

5.2 Problem Background

In this section, we present the mathematical formulation of the power flow mapping problem.

5.2.1 Alternating Current (AC) Power Flow Mapping

Let the power network be defined by a graph $\mathcal{N}(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of buses and $\mathcal{E}$ is the set of transmission or distribution lines. Let $\mathcal{G} \subseteq \mathcal{V}$ be the buses that are attached to generators. The equations that govern AC power flow between buses in the network are given as:

$$p_{ij} = |v_i||v_j|\big(G_{ij}\cos\theta_{ij} + B_{ij}\sin\theta_{ij}\big), \qquad \forall (i,j) \in \mathcal{E} \tag{5.1a}$$

$$q_{ij} = |v_i||v_j|\big(G_{ij}\sin\theta_{ij} - B_{ij}\cos\theta_{ij}\big), \qquad \forall (i,j) \in \mathcal{E} \tag{5.1b}$$

where the complex voltage at each bus i is given as $v_i \triangleq |v_i|e^{\mathbf{j}\theta_i} \in \mathbb{C}$. The expressions p_{ij} and q_{ij} respectively represent the real and reactive power flows between buses i and j. The expression θ_{ij} is the difference in voltage angle between buses i and j, given as $\theta_{ij} \triangleq \theta_i - \theta_j$. The network parameters G_{ij} and B_{ij} are respectively the conductance and susceptance for the line between buses i and j, where the complex admittance is given as $Y_{ij} = G_{ij} + \mathbf{j}B_{ij}$.

In the power flow mapping problem, the conductance G_{ij} and susceptance B_{ij} are taken to be part of the unknown parameter set, and some subset of possible SCADA measurements $\mathcal{M}$ are assumed to be available. These measurements can consist of real power flows p_{ij}, reactive power flows q_{ij}, real power injections p_i, and reactive power injections q_i. Additionally, we have PMU readings at some buses that provide estimates for voltage magnitudes $|v_i|$ and voltage angles θ_i.

Note that while this formulation does assume some awareness of the power network topology, provided by the graph $(\mathcal{V}, \mathcal{E})$, we can model uncertainty in some portion of the topology using only the line parameters. To accomplish this, we consider $\mathcal{E}$ to include all possible connected lines in the network. If the line (i, j) is switched off or does not actually exist in the network, the line parameters G_{ij} and B_{ij} will be zero. Thus, we can model uncertainty in the network topology by considering only the line parameters as unknowns. However, we will assume that there is some baseline understanding of the network topology, including some awareness of the number of buses in the network and how they are interconnected. This is a reasonable assumption for most realistic test cases in which a system operator would have full understanding of the baseline network topology but might not be aware of lines that have switched open or closed due to information delay or cyberattacks.

In order to formulate the power flow equations (5.1) as a kernel within the RKHS so that the kernel trick can be used for SVR, we introduce new variables $d_i \triangleq |v_i|\cos\theta_i$ and $e_i \triangleq |v_i|\sin\theta_i$ for all $i \in \mathcal{V}$ that correspond to the rectangular coordinates of the complex voltage. Then, using trigonometric identities, we can rewrite the power flow equations as:

$$p_{ij} = G_{ij}(d_i d_j + e_i e_j) + B_{ij}(e_i d_j - d_i e_j), \qquad \forall (i,j) \in \mathcal{E} \tag{5.2a}$$

$$q_{ij} = G_{ij}(e_i d_j - d_i e_j) - B_{ij}(d_i d_j + e_i e_j), \qquad \forall (i,j) \in \mathcal{E} \tag{5.2b}$$

We can do the same for real and reactive power injections to get the following relations:

$$p_i = G_{ii}d_i^2 + G_{ii}e_i^2 + \sum_{j \neq i} G_{ij}(d_i d_j + e_i e_j) + B_{ij}(e_i d_j - d_i e_j), \qquad \forall i \in \mathcal{V} \tag{5.3a}$$

$$q_i = -B_{ii}d_i^2 - B_{ii}e_i^2 + \sum_{j\neq i} G_{ij}(e_i d_j - d_i e_j) - B_{ij}(d_i d_j + e_i e_j), \qquad \forall i \in \mathcal{V} \tag{5.3b}$$

where G_{ii} is the self conductance at bus i, composed of the shunt conductance G_i^{sh} and mutual conductances G_{ij} as $G_{ii} = G_i^{sh} - \sum_{j\neq i} G_{ij}$. Similarly, B_{ii} is the self susceptance at bus i, composed of the shunt susceptance B_i^{sh} and mutual susceptances B_{ij} as $B_{ii} = B_i^{sh} - \sum_{j\neq i} B_{ij}$. These parameters are also assumed to be unknown for the network.

5.2.2 Availability of PMU Data

There is strong interest in using increasingly available PMU data in state estimation and false data detection [44–46]. PMUs provide measurements of voltage and current phasors and have been shown to be helpful in improving the reliability of grid monitoring and detection tools [44]. While PMU penetration in the grid is growing but still somewhat limited due to high PMU installation costs, PMUs provide a larger quantity of real-time data than existing SCADA systems. While SCADA systems collect samples about every 4 seconds, PMUs typically collect about 30-60 samples per second [44, 47]. The large amount of PMU is well-suited for machine learning applications such as the power flow mapping problem presented in the preceding section.

We consider the case where we obtain voltage magnitude $|v_i|$ and angle data θ_i from PMUs on some subset of buses in the network. The PMU data serve as the input data for our mapping problem (appearing as $\mathbf{x}$ in Equation (5.12)), and the SCADA data serve as the output (appearing as $\mathbf{y}$ in Equation (5.12)). In order to reconcile the synchronization gap between PMU and SCADA data, we associate some set of PMU data to each SCADA measurement, i.e. we take 10 PMU samples collected both before and after a given SCADA measurement and associate those to the SCADA measurement. Thus, for one time step of SCADA data, we have created duplicate $\mathbf{y}_t$ SCADA measurements corresponding to each $\mathbf{x}_t$ PMU measurement. Other types of time synchronization smoothing methods could also be used to reconcile the PMU and SCADA datasets such as averaging the PMU data or estimation fusion approaches [48, 49].

5.3 Constrained SVR Problem Formulation

In this section, we present the primal SVR formulation of the power system mapping problem and show how sparsity-enforcing constraints can be added to the classic SVR problem. Then, we find the dual of this constrained SVR problem and show that it is a convex quadratic program.

5.3.1 Power Flow as Represented by the Quadratic Kernel

In [11], it was shown that AC power flow can be represented exactly by the quadratic kernel, which is given as:

$$K(\mathbf{x}_1, \mathbf{x}_2) = (\langle \mathbf{x}_1, \mathbf{x}_2 \rangle)^2, \qquad \forall \mathbf{x} \in \mathcal{X} = \mathbb{R}^{2n} \tag{5.4}$$

where $n \leq |\mathcal{V}|$ corresponds to the number of buses in the network where PMU data is available and $\mathbf{x} \in \mathbb{R}^{2n}$ corresponds to the real and reactive components in the complex voltage vector $\mathbf{v} \in \mathbb{C}^n$, where $\mathbf{x}$ is given as:

$$\mathbf{x} \triangleq \begin{bmatrix} d_1 & d_2 & \dots & d_n & e_1 & e_2 & \dots & e_n \end{bmatrix}^T \tag{5.5}$$

Then, the feature mapping corresponding to the quadratic kernel $\phi(\mathbf{x}) \in \mathbb{R}^D$ is given by:

$$\phi(\mathbf{x}) = [d_1^2 \ \dots \ d_n^2 \ e_1^2 \ \dots \ e_n^2 \ \sqrt{2}d_1d_2 \ \dots \ \sqrt{2}d_1e_2 \ \dots \ \sqrt{2}e_{n-1}e_n]^T \tag{5.6}$$

where we have that $D = \binom{2n+1}{2} = 2n^2 + n$. By construction, we have that $K(\mathbf{x}_1, \mathbf{x}_2) = \phi(\mathbf{x}_1)^T \phi(\mathbf{x}_2)$, a relation that allows for the kernel trick in the dual SVR formulation, i.e. the replacement of $\phi(\mathbf{x}_1)^T \phi(\mathbf{x}_2)$ terms by $K(\mathbf{x}_1, \mathbf{x}_2)$.

Based on this formulation, we can rewrite each of the real and reactive power flow and injection measurement relations given by Equations (5.2) and (5.3) as a dot product of the quadratic feature mapping $\phi(\mathbf{x})$ given in (5.6) and a specific parameter vector with known structure and unknown values. The power flow measurement equations (5.2a) and (5.2b) can be written as:

$$p_{ij} = \langle \mu_{p_{ij}}, \phi(\mathbf{x}) \rangle, \quad \forall (i,j) \in \mathcal{E} \tag{5.7a}$$
$$q_{ij} = \langle \mu_{q_{ij}}, \phi(\mathbf{x}) \rangle, \quad \forall (i,j) \in \mathcal{E} \tag{5.7b}$$

where $\mu_{p_{ij}}, \mu_{q_{ij}} \in \mathbb{R}^D$ with the k^{th} entries of $\mu_{p_{ij}}$ and $\mu_{q_{ij}}$ are defined as:

$$(\mu_{p_{ij}})_k \triangleq \begin{cases} G_{ij}, \text{ if } \phi(\mathbf{x}^{ij})_k = \sqrt{2}d_id_j \text{ or } \sqrt{2}e_ie_j \\ B_{ij}, \text{ if } \phi(\mathbf{x}^{ij})_k = \sqrt{2}e_id_j \\ -B_{ij}, \text{ if } \phi(\mathbf{x}^{ij})_k = \sqrt{2}d_ie_j \\ 0, \text{ otherwise} \end{cases} \tag{5.8a}$$

$$(\mu_{q_{ij}})_k \triangleq \begin{cases} -B_{ij}, \text{if } \phi(\mathbf{x}^{ij})_k = \sqrt{2}d_id_j \text{ or } \sqrt{2}e_ie_j \\ G_{ij}, \text{ if } \phi(\mathbf{x}^{ij})_k = \sqrt{2}e_id_j \\ -G_{ij}, \text{ if } \phi(\mathbf{x}^{ij})_k = \sqrt{2}d_ie_j \\ 0, \text{ otherwise} \end{cases} \tag{5.8b}$$

Similarly, the power injection measurement equations (5.3a) and (5.3b) can be written as:

$$p_i = \langle \mu_{p_i}, \phi(\mathbf{x}) \rangle, \quad \forall i \in \mathcal{V} \tag{5.9a}$$

$$q_i = \langle \mu_{q_i}, \phi(\mathbf{x}) \rangle, \quad \forall i \in \mathcal{V} \tag{5.9b}$$

where $\mu_{p_i}, \mu_{q_i} \in \mathbb{R}^D$ with the k^{th} entries μ_{p_i} and μ_{q_i} are defined as:

$$(\mu_{p_i})_k \triangleq \begin{cases} G_{ii}, & \text{if } \phi(\mathbf{x})_k = d_i^2 \text{ or } e_i^2 \\ (\mu_{p_{ij}})_k, & \text{otherwise} \end{cases} \tag{5.10a}$$

$$(\mu_{q_i})_k \triangleq \begin{cases} -B_{ii}, & \text{if } \phi(\mathbf{x})_k = d_i^2 \text{ or } e_i^2 \\ (\mu_{p_{ij}})_k, & \text{otherwise} \end{cases} \tag{5.10b}$$

The dot product relations (5.7) and (5.9) can be shown to be respectively equivalent to (5.2) and (5.3) by expanding the equations (5.7) and (5.9) with the defined μ-parameters given by (5.8) and (5.10) and using trigonometric relations to simplify the equations. Note that the G_{ij} and B_{ij} entries should be equal to zero in the case where there is no line that connects buses i and j. However, in the case where the line topology is only partially known, we can associate G_{ij} and B_{ij} parameters to any possible line in the network.

We can observe that due to the sparsity inherent in power networks [22], most of the entries in the μ-parameters should be zero. We formalize this observation in the following lemma.

Lemma 3. *The maximum ratio of non-zero entries to zero entries in $\mu_{p_{ij}}$ or $\mu_{q_{ij}}$ is $4 : 2n^2 + n - 4$, and the maximum ratio of non-zero entries to zero entries in μ_{p_i} or μ_{q_i} is $4|\mathcal{E}(i)| + 2 : 2n^2 + n - 4|\mathcal{E}(i)| - 2$ where $\mathcal{E}(i)$ is the set of lines attached to bus i.*

Proof. In the case of real (or reactive) power flows, only 4 of the entries in each $\mu_{p_{ij}}$ (or $\mu_{q_{ij}}$) could possibly be non-zero by definition in (5.8). Similarly, in the case of real (or reactive) power injections, only $4|\mathcal{E}(i)| + 2$ of the entries in each μ_{p_i} or μ_{q_i} could possibly be non-zero by definition in (5.10). The minimum number of zero entries is found by subtracting these from the number of features, given by $D = 2n^2 + n$. □

We will exploit the inherent sparsity of the μ-parameters in order to learn the network line parameters and topology via a constrained SVR method described in the sections below.

5.3.2 Constrained SVR with Multiple Measurement Types

For a complete background on support vector regression methods, see [28]. Most of the SVR methods in the literature try to learn a mapping between vectors $\mathbf{x}_t$ and scalars y_t, for multiple time steps $t = 1, \ldots, T$. However, for power flow mapping, we want to learn the relationship between vectors $\mathbf{x}_t \in \mathbb{R}^{2n}$ and vectors $\mathbf{y}_t \in \mathbb{R}^M$, for multiple time steps $t = 1, \ldots, T$, where the vector $\mathbf{y}_t$ contains multiple types of p_{ij}, q_{ij}, p_i, and q_i SCADA measurements and the mapping $\mathbf{x}_t \to \mathbf{y}_t$ is given by the equations (5.7) and (5.9). Given a set of measurements $\mathcal{M}$, where $M \triangleq |\mathcal{M}|$, and a set of data collection time steps $1, \ldots, T$, we can write the corresponding state equation model in concise form as:

$$\mathbf{y}_t = W\phi(\mathbf{x}_t), \qquad \forall t \in \{1, \ldots, T\} \tag{5.11}$$

where the weight matrix $W \in \mathbb{R}^{M\times D}$ relates the features corresponding to PMU voltage measurements to the SCADA measurements in $\mathcal{M}$. Each row in matrix W corresponds to a μ-parameter defined in Equations (5.8) and (5.10). In Lemma 3, we observed that these μ-parameters are very sparse, which translates to W having this same sparse structure. We want to enforce this sparsity pattern, which we will take as some matrix $E \in \mathbb{R}^{M\times D}$ composed of a sparse set of variables as defined in the next section, on the weight matrix W, providing the constraint $W - E = 0$. Combining this constraint with the SVR model, we arrive at the primal version of the constrained SVR problem:

$$\min_{W,\xi,E} \ \frac{1}{2}||W||_F^2 + C\sum_{m=1}^{M}\sum_{t=1}^{T}(\xi_{m,t} + \xi^*_{m,t}) \tag{5.12a}$$

$$\text{s.t. } \mathbf{y}_t - W\phi(\mathbf{x}_t) \le \epsilon + \xi_t, \ \ \forall t \in \{1,\dots,T\} \tag{5.12b}$$

$$W\phi(\mathbf{x}_t) - \mathbf{y}_t \le \epsilon + \xi^*_t, \ \ \forall t \in \{1,\dots,T\} \tag{5.12c}$$

$$\xi_t, \xi^*_t \ge 0, \ \ \forall t \in \{1,\dots,T\} \tag{5.12d}$$

$$W - E = 0 \tag{5.12e}$$

where $\epsilon \in \mathbb{R}^M$, $\epsilon_i \ge 0$ for all $i \in \{1,\dots,M\}$ defines the ϵ-tube around the estimator inside which errors are not penalized. The variables $\xi_t, \xi^*_t \in \mathbb{R}^M$ defined for all $t \in \{1,\dots,T\}$ are the penalty terms for violating the state equations (5.11) outside of the ϵ-tube. A linear penalty on these errors, scaled by the hyperparameter $C > 0$, is added to the Frobenius norm of the weight matrix W to form the objective function. The inclusion of the Frobenius norm in the objective serves to encourage flatness in the weight parameters, thus promoting parameters that are physically realistic. The hyperparameter C determines the trade-off between the bi-objectives of promoting flatness and minimizing violations of the state equations.

Our modification from the classical SVR problem that maps $\mathbf{x}_t$ to $y_t \in \mathbb{R}$ is that the weight vector becomes a matrix given by W and ξ_t, ξ^*_t, and ϵ are now vectors in $\mathbb{R}^M$. If we ignore the sparsity constraint (5.12e), we can decouple this matrix version of the SVR problem (5.12) into the standard vector-version SVR problems for each measurement type y_m, for all $m = 1,\dots,M$. Thus, the matrix $E \in \mathbb{R}^{M\times D}$ serves to couple the different SVR problems that correspond to each measurement type in the vector $\mathbf{y}_t$. Intuitively, this matrix E allows us to use information obtained about the parameters corresponding to one measurement type to learn the parameters corresponding to a different measurement type. For example, the line conductance G_{ij} of some line $(i,j) \in \mathcal{E}$ appears in the measurement equations for both p_{ij} and q_i (as well as q_{ij} and p_i), thus its predicted value should be the same whether we predict the mapping from $\mathbf{x}$ to p_{ij} or from $\mathbf{x}$ to q_i.

5.3.3 Defining the Sparsity Pattern on a Two-Bus Network

To define structure of the sparsity pattern E, let us first consider a two-bus network (shown in Figure 5.1), with one line connecting buses 1 and 2. We consider the case where $\mathbf{y}_t \in \mathbb{R}^M$ with $M = 4$ that corresponds to the following measurement vectors for each time step

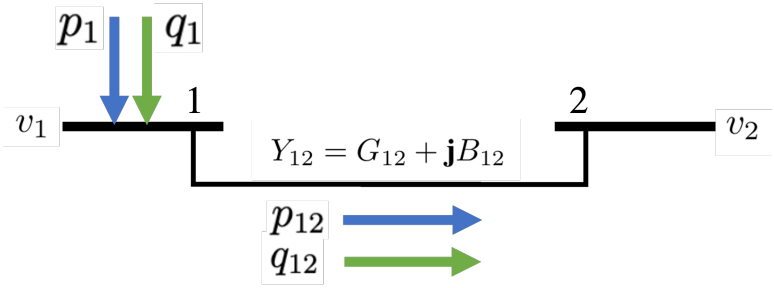

Figure 5.1: Example of two-bus network with selected SCADA measurements from Equation (5.13)

$t \in \{1, \ldots, T\}$:

$$\mathbf{y} = \begin{bmatrix} p_{12} & q_{12} & p_1 & q_1 \end{bmatrix}^T \tag{5.13}$$

Using the quadratic feature vector given by (5.6), where $\phi(\mathbf{x}) \in \mathbb{R}^D$ with $D = 10$ corresponding to the two-bus network, and the sparse relations for the μ-parameters given by (5.8) and (5.10), we know that W will have the following sparsity pattern E:

$$E \triangleq \begin{bmatrix} 0 & 0 & 0 & 0 & G_{12} & 0 & B_{12} & -B_{12} & 0 & G_{12} \\ 0 & 0 & 0 & 0 & B_{12} & 0 & -G_{12} & G_{12} & 0 & B_{12} \\ G_{11} & 0 & G_{11} & 0 & G_{12} & 0 & B_{12} & -B_{12} & 0 & G_{12} \\ -B_{11} & 0 & -B_{11} & 0 & B_{12} & 0 & -G_{12} & G_{12} & 0 & B_{12} \end{bmatrix} \tag{5.14}$$

We have used Equations (5.8a), (5.8b), (5.10a), and (5.10b) to respectively construct rows 1, 2, 3, and 4 of matrix E corresponding to the measurements p_{12}, q_{12}, p_1, and q_1. Construction of the matrix E depends solely on the choice of measurement set and the baseline network topology, both of which are known *a priori*. Note that all the G and B terms in matrix E are unknown parameters. Using this example definition of E, we can formulate the dual of (5.12) for the two-bus network example. We will then generalize our formulation to larger networks.

5.3.4 Formulating the Dual of the Constrained SVR Problem

In the primal constrained SVR problem (5.12), the constraints (5.12b) and (5.12c) will slow down a generic quadratic program solver due to the high dimension of the feature vector $\phi(\mathbf{x})$, which scales with n^2. Thus, it is more computationally useful to consider the dual form of (5.12), making use of the kernel trick to eliminate most instances of $\phi(\mathbf{x})$.

In order to take the dual of (5.12), we introduce the Lagrange multipliers $\alpha_t, \alpha_t^* \in \mathbb{R}_+^M$ corresponding to the (5.12b) and (5.12c) inequality constraints, $\beta_t, \beta_t^* \in \mathbb{R}_+^M$ corresponding to the (5.12d) inequality constraints, and $\lambda \in \mathbb{R}^{M \times D}$ corresponding to the (5.12e) equality constraints. Then, the Lagrangian of (5.12) can be written as:

$$L = \frac{1}{2}||W||_F^2 + C\sum_{m=1}^{M}\sum_{t=1}^{T}(\xi_{m,t} + \xi_{m,t}^*) - \sum_{t=1}^{T}(\alpha_t + \alpha_t^*)^T\epsilon + \sum_{t=1}^{T}(\alpha_t - \alpha_t^*)^T\left(\mathbf{y}_t - W\phi(\mathbf{x}_t)\right) - \sum_{t=1}^{T}\left(\alpha_t^T\xi_t + {\beta_t}^T\xi_t + (\alpha_t^*)^T\xi_t^* + (\beta_t^*)^T\xi_t^*\right) + \text{trace}\{\lambda^T(E - W)\} \quad (5.15)$$

For the two-bus network example, given Equation (5.14) defining E, we can expand the $\text{trace}\{\lambda^T E\}$ term as:

$$\begin{aligned}\text{trace}\{\lambda^T E\} = {} & \lambda_{1,5}G_{12} + \lambda_{1,7}B_{12} - \lambda_{1,8}B_{12} + \lambda_{1,10}G_{12} + \lambda_{2,5}B_{12} - \lambda_{2,7}G_{12} \\ & + \lambda_{2,8}G_{12} + \lambda_{2,10}B_{12} + \lambda_{3,1}G_{11} + \lambda_{3,3}G_{11} + \lambda_{3,5}G_{12} + \lambda_{3,7}B_{12} \\ & - \lambda_{3,8}B_{12} + \lambda_{3,10}G_{12} - \lambda_{4,1}B_{11} - \lambda_{4,3}B_{11} + \lambda_{4,5}B_{12} - \lambda_{4,7}G_{12} \\ & + \lambda_{4,8}G_{12} + \lambda_{4,10}B_{12} \end{aligned} \quad (5.16)$$

The equations ensuring stationarity of the Lagrangian (5.15) with respect to the G and B unknown parameters are:

$$\partial L/\partial G_{11} = \lambda_{3,1} + \lambda_{3,3} = 0 \quad (5.17a)$$
$$\partial L/\partial B_{11} = -\lambda_{4,1} - \lambda_{4,3} = 0 \quad (5.17b)$$
$$\partial L/\partial G_{12} = \lambda_{1,5} + \lambda_{1,10} - \lambda_{2,7} + \lambda_{2,8} + \lambda_{3,5} + \lambda_{3,10} - \lambda_{4,7} + \lambda_{4,8} = 0 \quad (5.17c)$$
$$\partial L/\partial B_{12} = \lambda_{1,7} - \lambda_{1,8} + \lambda_{2,5} + \lambda_{2,10} + \lambda_{3,7} - \lambda_{3,8} + \lambda_{4,5} + \lambda_{4,10} = 0 \quad (5.17d)$$

The equations (5.17) can be written in more concise form as:

$$\text{trace}\{\lambda^T L_r\} = 0, \quad \forall r \in \{1, \ldots, R\} \quad (5.18)$$

where $R = 4$ for the given two-bus example and measurement set. More generally, R is a known integer corresponding to the number of unique unknown line parameters in the network, as defined by the chosen measurement set $\mathcal{M}$ and the given baseline network topology. The matrices $L_r \in \mathbb{R}^{M\times D}$ for all $r \in \{1, \ldots, R\}$ are also explicitly known given the chosen measurement set and line topology.

For example, on the given two-bus example, Equation (5.17c) corresponding to the stationarity of the Lagrangian with respect to G_{12} can be written as $\text{trace}\{\lambda^T L_3\} = 0$ by defining L_3 as:

$$L_3 \triangleq \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \end{bmatrix} \quad (5.19)$$

The equations ensuring stationarity of the Lagrangian (5.15) with respect to the error penalty variables ξ_t, and ξ_t^* for all $t \in \{1, \ldots, T\}$ are:

$$\partial L/\partial \xi_{m,t} = C - \alpha_{m,t} - \beta_{m,t} = 0 \quad (5.20a)$$

$$\partial L/\partial \xi^*_{m,t} = C - \alpha^*_{m,t} - \beta^*_{m,t} = 0 \tag{5.20b}$$

Finding the stationarity of the Lagrangian (5.15) with respect to the weight matrix W yields the relationship between the primal weight matrix W and the dual support vectors α_t and α^*_t for all $t \in \{1, \dots, T\}$:

$$W = \sum_{t=1}^{T} (\alpha_t - \alpha^*_t)\phi(\mathbf{x}_t)^T + \lambda \tag{5.21}$$

Combining these constraints and plugging the definition of W in (5.21) back into the Lagrangian (5.15), we arrive at the dual form of the constrained SVR problem, given in the theorem below.

Theorem 7. *The dual of the constrained SVR problem (5.12) corresponding to the power mapping problem can be written as the convex quadratic program:*

$$\min_{\alpha,\alpha^*,\lambda} \; f(\alpha, \alpha^*, \lambda) \tag{5.22a}$$

$$s.t. \;\; \alpha_{m,t}, \alpha^*_{m,t} \in [0, C], \;\; \forall m \in \{1, \dots, M\}, \;\; \forall t \in \{1, \dots, T\} \tag{5.22b}$$

$$trace\{\lambda^T L_r\} = 0, \;\; \forall r \in \{1, \dots, R\} \tag{5.22c}$$

where we have $\alpha, \alpha^* \in \mathbb{R}^{M \times T}$, $\lambda \in \mathbb{R}^{M \times D}$, *and the objective function:*

$$f(\alpha_t, \alpha^*_t, \lambda) = \frac{1}{2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\alpha_t - \alpha^*_t)^T(\alpha_s - \alpha^*_s)K(\boldsymbol{x}_s, \boldsymbol{x}_t) + \sum_{t=1}^{T}(\alpha_t + \alpha^*_t)^T\epsilon$$
$$- \sum_{t=1}^{T}(\alpha_t - \alpha^*_t)^T\boldsymbol{y}_t + \frac{1}{2}||\lambda||^2_F + \sum_{t=1}^{T}(\alpha_t - \alpha^*_t)^T\lambda\phi(\boldsymbol{x}_t) \tag{5.23}$$

Proof. See Appendix 5.A. □

To see that the dual problem (5.22) is a convex quadratic program, we can introduce the variable $Z \triangleq \begin{bmatrix}\alpha & \alpha^* & \lambda\end{bmatrix}^T \in \mathbb{R}^{(2T+D)\times M}$ and rewrite the problem in stacked form as:

$$\min_{Z} \; \frac{1}{2}\text{trace}\{Z^T A_0 Z\} + \text{trace}\{A_1^T Z\} \tag{5.24a}$$

$$\text{subject to: } \underline{Z} \leq Z \leq \overline{Z} \tag{5.24b}$$

$$\text{trace}\{Z^T \tilde{L}_r\} = 0, \;\; \forall r = 1, \dots, R \tag{5.24c}$$

where the lower and upper bounds $\underline{Z}$ and $\overline{Z}$ follow from (5.22b) and the matrices $\tilde{L}_r$ are expanded versions of L_r such that $\text{trace}\{Z^T\tilde{L}_r\} = \text{trace}\{\lambda^T L_r\}$ for all $r \in \{1, \dots, R\}$. Additionally, we define the matrices $A_0 \in \mathbb{S}^{(2T+D)}$ and $A_1 \in \mathbb{R}^{(2T+D)\times M}$ as:

$$A_0 \triangleq \begin{bmatrix} Q & -Q & \Phi^T \\ -Q & Q & -\Phi^T \\ \Phi & -\Phi & I_D \end{bmatrix} \tag{5.25a}$$

$$A_1 \triangleq \begin{bmatrix} [(\epsilon^T)_{\times T}] - Y^T \\ [(\epsilon^T)_{\times T}] + Y^T \\ 0_{D\times M} \end{bmatrix} \tag{5.25b}$$

where we have $\Phi \in \mathbb{R}^{D\times T}$ as the matrix composed of columns $\phi(\mathbf{x}_t) \in \mathbb{R}^D$ and $Y \in \mathbb{R}^{M\times T}$ as the matrix composed of columns $\mathbf{y}_t \in \mathbb{R}^D$ for all $t \in \{1,\dots,T\}$. We define the kernel matrix $Q \in \mathbb{R}^{T\times T}$ as having entries $Q_{s,t} = K(\mathbf{x}_s, \mathbf{x}_t)$ for all $s,t \in \{1,\dots,T\}$, noting that $Q = \Phi^T\Phi$. The notation $[(\epsilon^T)_{\times T}]$ indicates a $T \times M$ matrix where every row equals ϵ^T, the notation I_D indicates an identity matrix of dimension D, and the notation $0_{D\times M}$ indicates a $D \times M$ matrix composed of all zeros. In this stacked formulation, we can observe that $A_0 = aa^T$ where $a \triangleq \begin{bmatrix}\Phi & -\Phi & I_D\end{bmatrix}^T$, thus $A_0 \succeq 0$ and the program is convex.

5.4 Analysis of Constrained SVR Approach

5.4.1 Strong Duality of Constrained SVR Problem

Similar to the classic SVR approach where the dual problem is used as a solution for the primal, we show that the proposed dual of the constrained SVR problem can also be used as a solution to the primal.

Theorem 8. *The dual of the constrained SVR problem given by (5.22) is exact. Therefore, solving (5.22) will recover the solution to (5.12).*

Proof. Since the primal problem (5.12) is a convex quadratic program, we just need to show that it satisfies the weak Slater's condition to prove strong duality. To show that there exists some point strictly within the feasible space of (5.12), we can set $W = E = 0$ and $\xi_{m,t} > \max\{0, (\mathbf{y}_t - \epsilon)_m\}$ and $\xi^*_{m,t} > \max\{0, (-\mathbf{y}_t - \epsilon)_m\}$ for all $t \in \{1,\dots,T\}$, $m \in \{1,\dots,M\}$. Then, we can see that the inequality constraints (5.12b),(5.12c), and (5.12d) are all strictly feasible for this point and weak Slater's holds. □

5.4.2 Effect of Measurement Availability on Constrained SVR

The availability of measurements, both in terms of SCADA measurements and PMU measurements, directly affects which line parameters we are able to accurately learn with the constrained SVR method. For example, in the case where two buses i and j both have PMUs and the SCADA measurements p_{ij} and q_{ij} are the only available measurements, we are able to exactly recover G_{ij} and B_{ij} when there is no measurement noise and are able to recover good estimates for these parameters when there is noise. This is evident by considering Equations (5.8a) and (5.8b) in relation to the constrained SVR problem (5.12) where C is sufficiently high. Similarly, if we have all possible PMU and SCADA measurements available in the network, we can recover all the exact G and B parameters in the network when there is no measurement noise and good estimates when there is noise. However, in the more realistic case where PMU measurements are not available at every bus, the formulation (5.12)

attributes the discrepancy due observed and unobserved PMU measurements to the penalty terms ξ and ξ^*. When there is no available PMU at bus i, it will be impossible to recover the parameters G_{ij}, B_{ij}, G_{ii} and B_{ii} since the features d_i^2, e_i^2, $\sqrt{2}d_id_j$, $\sqrt{2}e_id_j$, $\sqrt{2}d_ie_j$, and $\sqrt{2}e_ie_j$ do not appear in the feature vector $\phi(\mathbf{x})$. Thus, SCADA measurements p_{ij} and q_{ij} are not helpful to learn G_{ij} and B_{ij} if either bus i or bus j is not equipped with a PMU. On the contrary, the real and reactive power injection measurements at bus i can still be useful to learn some parameters even if not all buses attached to bus i are equipped with PMUs.

While limited access to PMU measurements does prevent the constrained SVR problem from accurately learning all the network parameters, there are many realistic cases where most of network parameters and topology are known but there is some uncertainty in parts of the network. For example, some line $(i, j) \in \mathcal{E}$ might often switch from open to closed in real-time such that its status is often unknown. By placing PMUs at buses i and j and solving the constrained SVR problem, this method could accurately recover G_{ij} and B_{ij} for the line and thus determine the status of the line. Future work should consider how strategic PMU placement can improve power flow mapping recovery.

5.5 Simulations

The simulations are run on a standard laptop (2.6 GHz 6-Core Intel Core i7 with 16 GB 2400MHz RAM). The software MATPOWER is used to import test networks, formulate admittance matrices, and generate sample data points based on solving power flow problems with the known system parameters [50]. Then, the constrained SVR model is formulated using the Pyomo modeling language in Python 3.8. The convex quadratic program (5.22) is solved with the Gurobi solver. Another possible algorithm based on Sequential Minimal Optimization (SMO) is presented in Appendix 5.D.

In the following simulations, we test our method on cases with various signal-to-noise (SNR) ratios in the SCADA measurements, cases with outliers in the SCADA or PMU measurements, and networks where measurements are only partially observed. For all of these test cases, we assume that the noise in PMU magnitude measurements follows a zero-mean Gaussian distribution with 0.005 p.u. standard deviation (SNR of 46 dB) and that the noise in PMU angle measurements follows a zero-mean Gaussian distribution with 0.01 p.u. standard deviation (SNR of 40 dB). These values are consistent with the existing literature on PMU errors [51]. For these simulations, we consider the case where we have 10 PMU measurements associated to each SCADA measurement at any time step, as discussed in Section 5.2.2.

5.5.1 Performance Metrics

To compare the performance of the propsed constrained SVR method with the classic SVR method, we consider the following performance metrics. Root-mean-square error (RMSE) measures how well the estimator fits the SCADA measurements $\mathbf{y}$ by penalizing the squared

error discrepancy, i.e. $RMSE = \sqrt{\frac{1}{MT}\sum_{t=1}^{T}\sum_{i=1}^{M}(y_{m,t} - \hat{y}_{m,t})^2}$, where $\hat{y}_{m,t}$ is output predicted by the estimator for some time step t and measurement type m. Mean absolute error (MAE) measures how well the estimator fits the SCADA measurements $\mathbf{y}$ by penalizing the absolute error discrepancy, i.e. $MAE = \frac{1}{MT}\sum_{t=1}^{T}\sum_{i=1}^{M}|y_{m,t} - \hat{y}_{m,t}|$. The benefit of MAE is that it is less sensitive to outliers than RMSE, so it provides another useful metric when comparing model performance. In terms of RMSE and MAE, the classic SVR method has been shown to perform well in the power flow mapping problem [11, 21]. This is because the classic SVR method has good performance learning the overall mapping between $\mathbf{x}_t$'s and $\mathbf{y}_t$'s, and RMSE and MAE do not penalize overfitting. Thus, a more useful metric to see if the method really learns the true line parameters is to consider the error between the actual line parameters and the estimated line parameters. In the case of mutual conductance, denoted by the subscript m.c., we take $\mathbf{G}_{\text{m.c.}}$ to be the vector of true mutual conductances G_{ij} for all lines (i, j) where we have obtained estimates $\hat{G}_{ij}$ using the SVR method and take $\hat{\mathbf{G}}_{\text{m.c.}}$ to be the corresponding vector of $\hat{G}_{ij}$'s. Then, the normalized estimation error for mutual conductance over all estimated lines is given by:

$$\Gamma_{\text{m.c.}} \triangleq \frac{||\mathbf{G}_{\text{m.c.}} - \hat{\mathbf{G}}_{\text{m.c.}}||_2}{||\mathbf{G}_{\text{m.c.}}||_2} \tag{5.26}$$

We have similar normalized error relations for the mutual susceptances (given as $\Gamma_{\text{m.s.}}$), self conductances (given as $\Gamma_{\text{s.c.}}$), and self susceptances (given as $\Gamma_{\text{s.s.}}$).

5.5.2 Effectiveness in the Presence of Noise

We start with the case that all buses are equipped with PMUs so that the voltages in the network are fully observable. For the IEEE 14-bus test case, we generate PMU measurements based on the initial voltage state of the MATPOWER case file, adding Gaussian-distributed random noise as described in the methodology above. We consider a SCADA measurement set that consists of all real and reactive power flow measurements as well as real and reactive power injection measurements at buses 2 and 3. For this test case setup, we consider the scenario where we have zero noise (SNR=∞) and 50 simulations at various SNRs from 45 to 0 dB. The results of these experiments are given in Table 5.1. These experiments demonstrate that the constrained SVR method outperforms the classic SVR method in terms of both line parameter recovery and solution speed and has similar performance to the classic SVR method in terms of RMSE and MAE. In Figure 5.2, we have plotted $\Gamma_{\text{m.c.}}$ and $\Gamma_{\text{m.s.}}$ as a function of the SNR of the SCADA measurements, noting that similar plots were obtained for the normalized estimation errors of the other line parameters (omitted for concision). Even when SCADA measurement errors are present, the constrained SVR method can recover decent estimates for the line parameters (within 14% of the true values on average) as long as the SNR in the measurements is greater than or equal to 15 dB. Conversely, the classic SVR method consistently fails to recover the true line parameters even when no noise is present in the SCADA measurements.

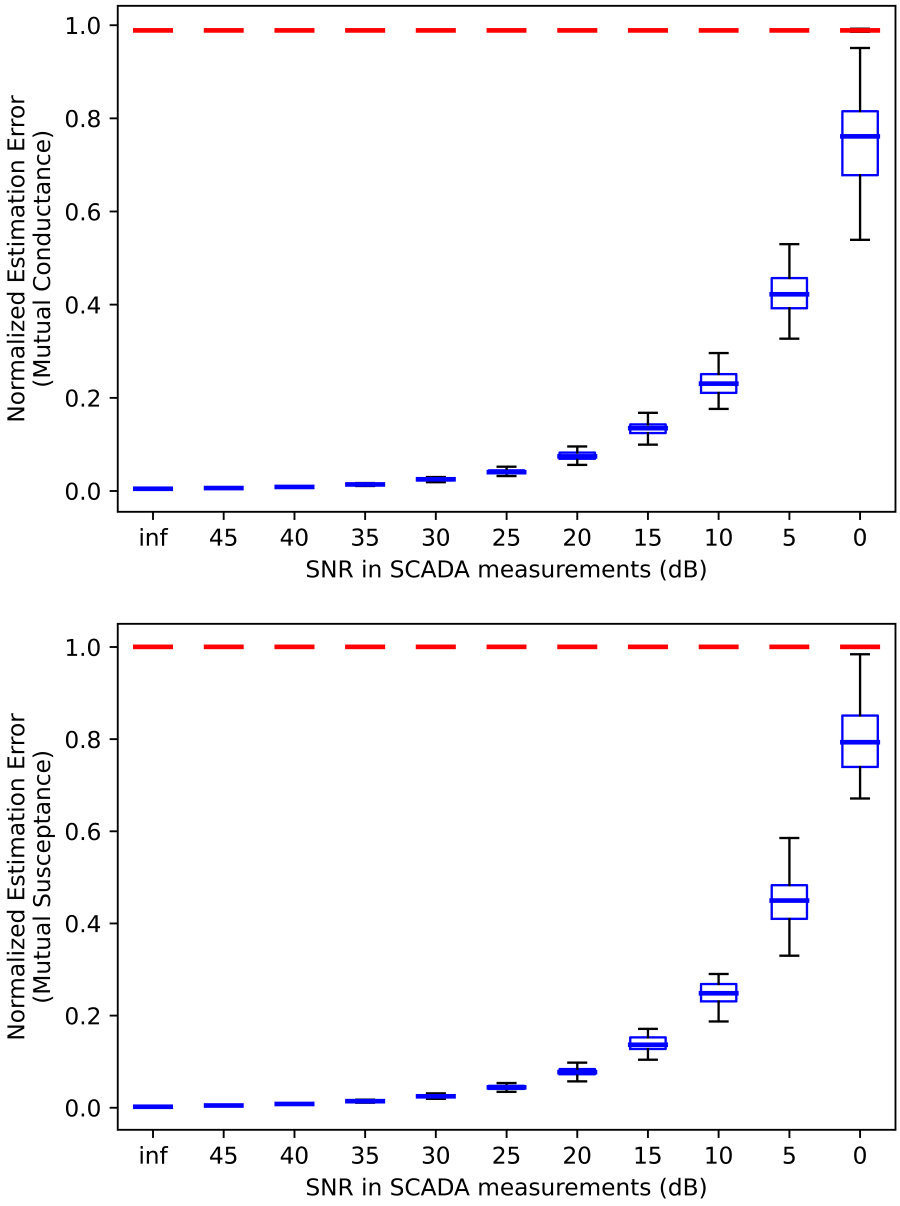

Figure 5.2: The normalized estimation error for mutual conductance (G_{ij}'s) (top figure) and mutual susceptance (B_{ij}'s) (bottom figure) as a function of signal-to-noise ratio (SNR) is presented for the 14-bus test case. Each box plot corresponds to 50 simulations of Gaussian-distributed random noise added to the SCADA measurements of this test case. The red lines are box plots of normalized estimation errors ($\Gamma_{\text{m.c.}}$ and $\Gamma_{\text{m.s.}}$) for the classic SVR method, and the blue box plots correspond to the normalized estimation errors ($\Gamma_{\text{m.c.}}$ and $\Gamma_{\text{m.s.}}$) for our proposed constrained SVR method.

Table 5.1: Comparison of Methods on 14-bus Network with Noise

Model	SNR	Avg. RMSE	Avg. MAE	Avg. solve time (s)	Avg. $\Gamma_{\text{m.c}}$	Avg. $\Gamma_{\text{s.c.}}$
Classic SVR	∞	0.0013	0.0010	46.86	0.9889	1.0003
	45	0.0009	0.0009	44.26	0.9889	1.0003
	30	0.0009	0.0008	44.49	0.9889	1.0003
	15	0.0009	0.0009	45.25	0.9889	1.0003
	0	0.0001	0.0009	42.71	0.9890	1.0003
Constr. SVR	∞	0.0535	0.0272	5.25	0.0048	0.0043
	45	0.0535	0.0272	5.21	0.0063	0.0045
	30	0.0536	0.0272	5.17	0.0246	0.0134
	15	0.0539	0.0273	5.19	0.1332	0.0813
	0	0.0664	0.0346	5.24	0.7501	0.4093

5.5.3 Robustness in the Case of Outliers

Next, we consider the same 14-bus test case as above with a SNR of 40 dB in the SCADA measurements but allow for outliers to be present in 0 to 8% in some of the PMU or SCADA measurement data. The subset of outlier-affected measurements within the PMU and SCADA datasets is chosen on a uniform distribution over all possible measurements, and the values of the outlier-affected measurements are chosen on a uniform distribution from $[-2, 2]$ per unit, scaled in relation to the chosen measurement type. The results from these simulations are given in Table 5.2 and Figure 5.3.

In Figure 5.4, we compare our proposed constrained SVR method to the classic SVR method in terms of solution time. From these simulations, it is clear that our proposed constrained SVR method outperformed existing SVR methods in terms of solution time and quality.

5.5.4 Effectiveness in Partially-Observable Networks

Finally, we consider the case where only a subset of buses in the network are equipped with PMUs. For these simulations, we consider the IEEE 14- and 30-bus networks and perform a sweep over varying PMU penetration levels in the network. For these simulations, we take the SNR of the SCADA noise as 40 dB and consider cases with and without outliers. Some results from these simulations are given in Figures 5.5, 5.6, and 5.7. From these simulations, we see that the sparse SVR model provides better average normalized estimation error and lower solution times at all PMU penetration levels than the classic SVR method, but that its benefits are more apparent at higher PMU penetrations.

Table 5.2: Performance Comparison of SVR Methods on 14-bus Network with Outliers

Model	% outlier	Avg. sol. time (s)	Avg. $\Gamma_{\text{m.c}}$	Avg. $\Gamma_{\text{s.c.}}$	Avg. $\Gamma_{\text{m.s.}}$	Avg. $\Gamma_{\text{s.s.}}$
Classic SVR	0.00	43.42	0.9906	1.0003	1.0005	1.0000
	1.02	43.89	0.9900	1.0005	1.0003	1.0000
	2.04	47.46	0.9914	1.0004	1.0004	1.0000
	3.06	48.02	0.9914	0.9997	1.0004	1.0000
	4.08	48.21	0.9922	1.0003	1.0003	1.0000
	5.10	47.81	0.9921	1.0002	1.0003	1.0001
	6.12	47.74	0.9917	1.0004	1.0004	1.0000
	7.14	48.09	0.9926	1.0003	1.0001	0.9999
	8.16	47.79	0.9925	0.9999	1.0002	0.9999
Constr. SVR	0.00	5.31	0.0089	0.0064	0.0083	0.0049
	1.02	5.38	0.4296	0.1257	0.2001	0.0629
	2.04	5.89	0.6211	0.2000	0.4078	0.0853
	3.06	5.86	0.9983	0.3181	0.5015	0.0995
	4.08	5.84	0.9450	0.3332	0.3823	0.1284
	5.10	5.85	1.3664	0.3543	0.5388	0.2383
	6.12	5.80	1.8727	0.5187	0.8018	0.2609
	7.14	5.88	1.6363	0.5128	0.7053	0.3181
	8.16	5.87	2.0223	0.5204	0.9563	0.2962

5.6 Conclusions

In this chapter, we proposed a new constrained SVR method that can learn the true power network topology of a network. Based on simulations on IEEE test cases, we showed that the proposed constrained SVR method is much better at recovering the true line parameters of a network, even in the presence of SCADA measurement noise, outliers, and missing data, than existing SVR methods.

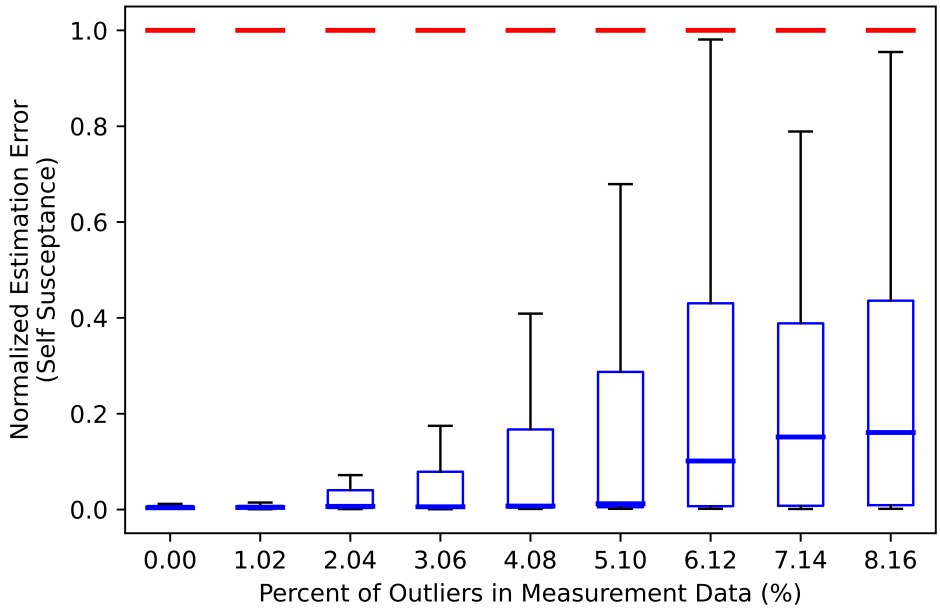

Figure 5.3: The normalized estimation error for self susceptances (B_{ii}'s) as a function of the percent of outliers is presented for the 14-bus test case. Each box plot corresponds to 50 simulations of uniformly-distributed random outliers added to some percent of the total PMU and SCADA measurements. The red lines are box plots for the classic SVR method, and the blue box plots correspond to our proposed constrained SVR method.

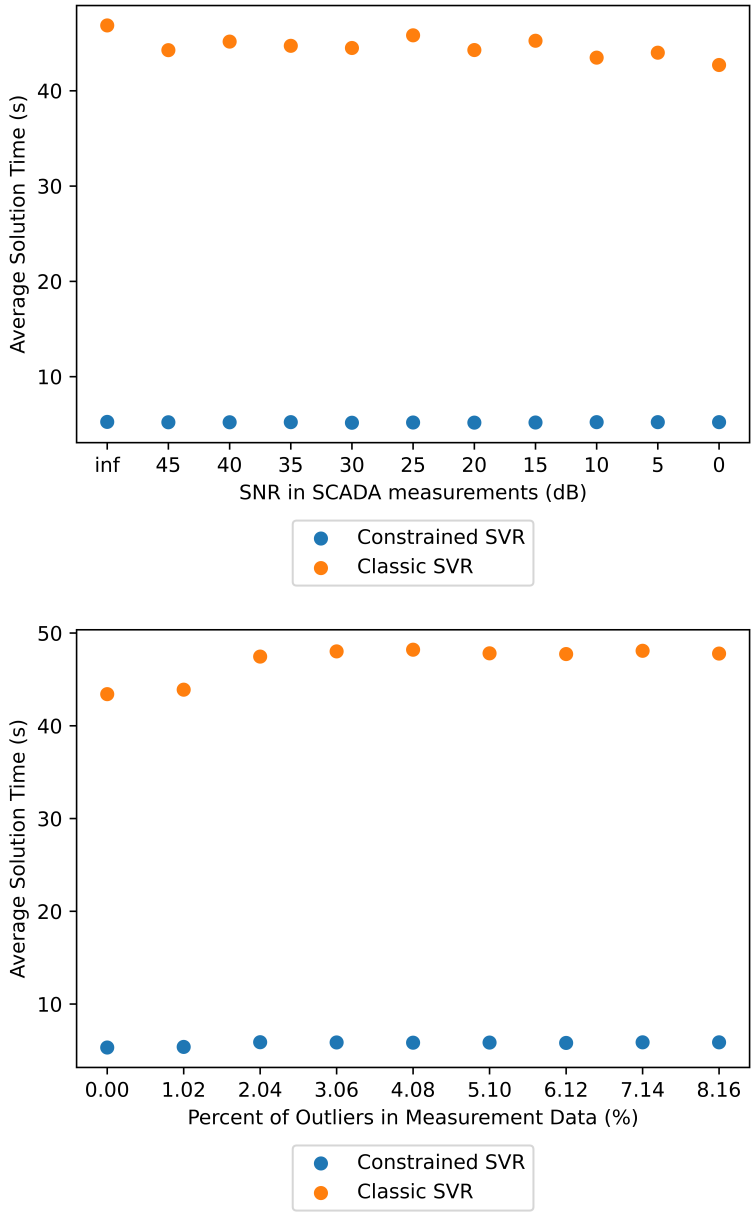

Figure 5.4: Average solution time for constrained and classic SVR methods on the 14-bus network over various signal-to-noise ratios in the SCADA data (top figure) and over various outlier levels (bottom figure). Each data point corresponds to the average over 50 simulations of randomly generated SCADA noise and/or outliers.

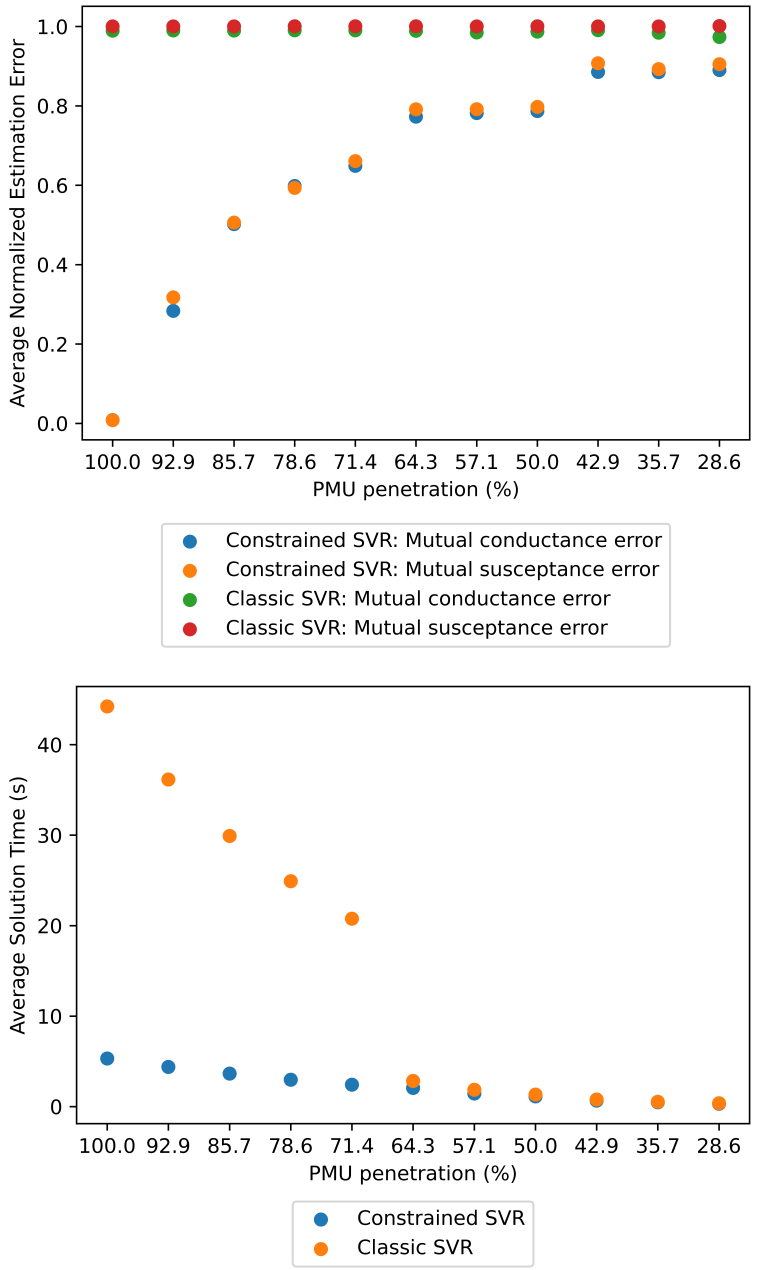

Figure 5.5: The top figure shows the average normalized estimation error for mutual conductances (G_{ij}'s) and mutual susceptances (B_{ij}'s) as a function of the PMU penetration, i.e. the percent of buses in the network from which PMU measurements are available, for the 14-bus test case with SCADA noise of 40 dB without outliers. Each data point corresponds to the average over 50 simulations. The bottom figure shows the average solution time for each set of 50 simulations.

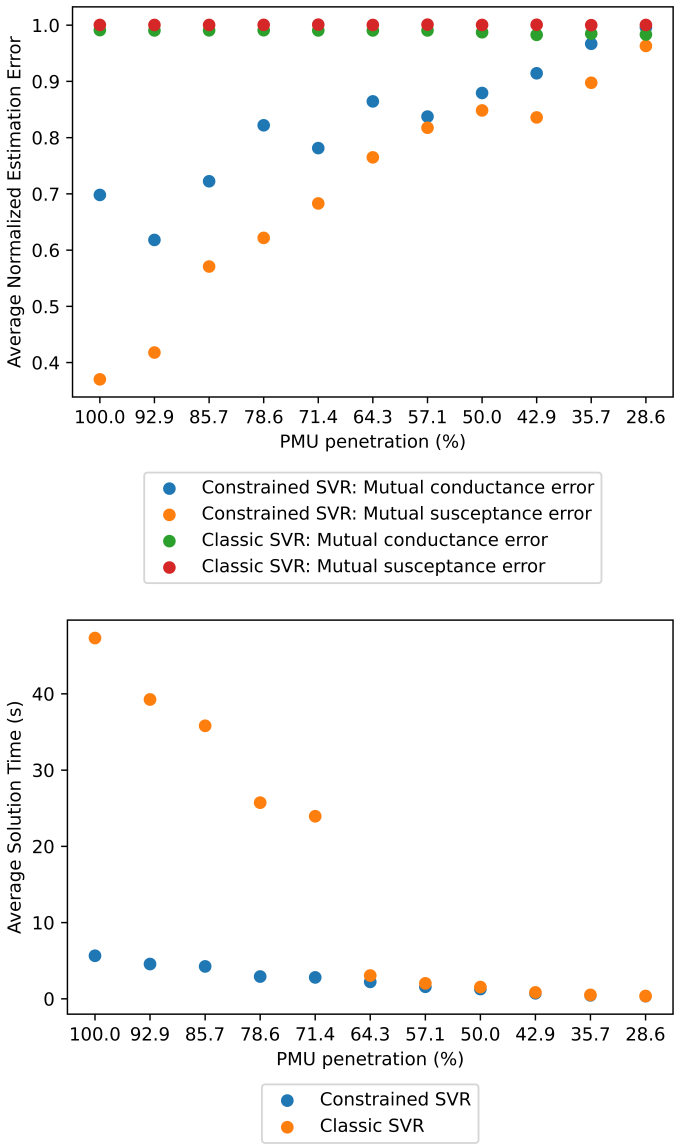

Figure 5.6: The top figure shows the average normalized estimation error for mutual conductances (G_{ij}'s and mutual susceptances (B_{ij}'s) as a function of the PMU penetration, i.e. the percent of buses in the network from which PMU measurements are available, for the 14-bus test case with SCADA noise of 40 dB and 2% outliers in all measurements. Each data point corresponds to the average over 50 simulations. In this figure, we see that the sparse SVR model provides better average normalized estimation error at all PMU penetration levels than the classic SVR method, but that its benefits are more apparent at higher PMU penetrations. The bottom figure shows the average solution time for each set of 50 simulations.

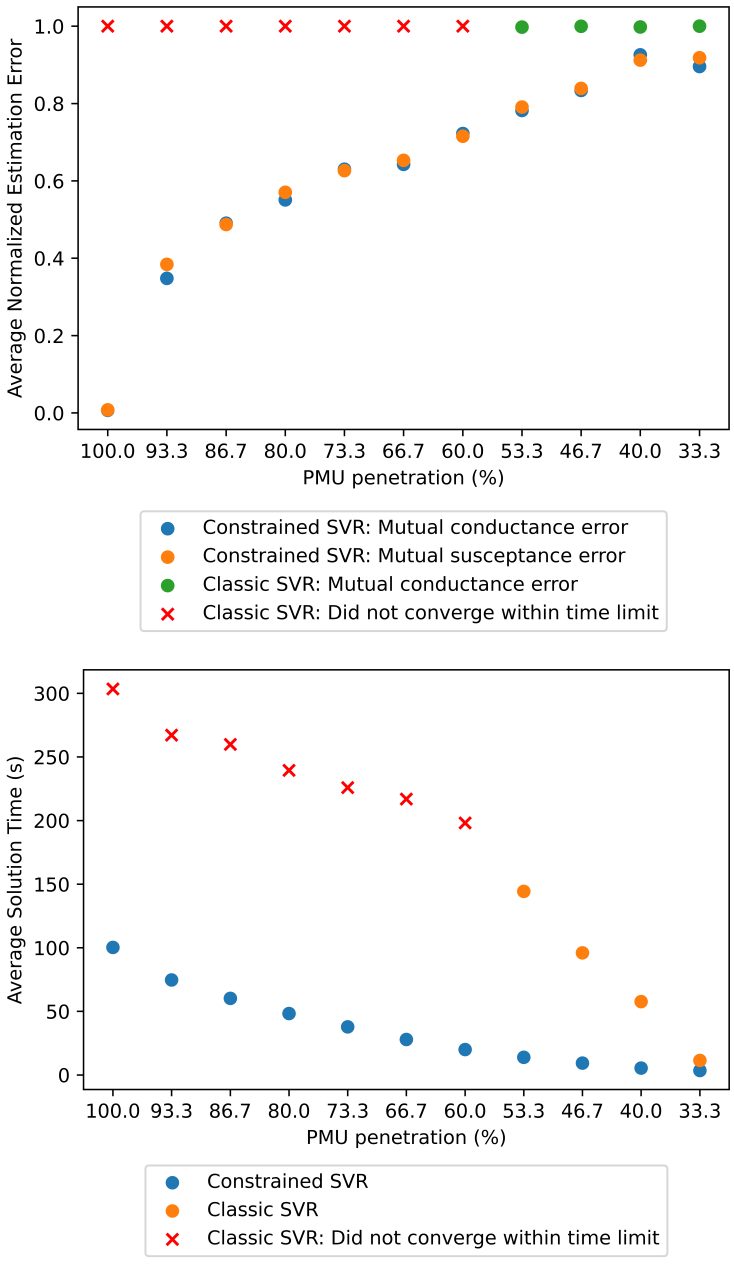

Figure 5.7: The top figure shows the average normalized estimation error of mutual conductance and mutual susceptance line parameters for the 30-bus network, with varying amounts of PMU penetration in the network. During each simulation, some subset of buses is randomly selected to be equipped with PMUs. Each data point corresponds to the average over 20 random simulations. The time limit per problem was chosen to be 3 minutes. The bottom figure shows the average solution time over the 20 simulations.

Appendix

5.A Proof of Theorem 7

The negative of the dual objective $f(\alpha_t, \alpha_t^*, \lambda)$, i.e. $g(\alpha_t, \alpha_t^*, \lambda) \triangleq -f(\alpha_t, \alpha_t^*, \lambda)$, is given by expanding (5.15) using (5.20) and (5.21):

$$\begin{aligned} g(\alpha_t, \alpha_t^*, \lambda) = \frac{1}{2}|| \textstyle\sum_{t=1}^T (\alpha_t - \alpha_t^*)\phi(\mathbf{x}_t)^T + \lambda||_F^2 - \sum_{t=1}^T (\alpha_t + \alpha_t^*)^T \epsilon \\ + \textstyle\sum_{t=1}^T (\alpha_t - \alpha_t^*)^T \mathbf{y}_t - \sum_{t=1}^T (\alpha_t - \alpha_t^*)^T \left(\sum_{t=1}^T (\alpha_t - \alpha_t^*)\phi(\mathbf{x}_t)^T + \lambda \right) \phi(\mathbf{x}_t) \\ + \text{trace}\{\lambda^T E\} - \text{trace}\left\{ \lambda^T \textstyle\sum_{t=1}^T (\alpha_t - \alpha_t^*)\phi(\mathbf{x}_t)^T \right\} - \text{trace}\{\lambda^T \lambda\} \end{aligned} \tag{5.27}$$

We expand the Frobenius norm to get:

$$\begin{aligned} \frac{1}{2}|| \textstyle\sum_{t=1}^T (\alpha_t - \alpha_t^*)\phi(\mathbf{x}_t)^T + \lambda||_F^2 = \tfrac{1}{2} \textstyle\sum_{t=1}^T \sum_{s=1}^T (\alpha_t - \alpha_t^*)^T (\alpha_s - \alpha_s^*)\phi(\mathbf{x}_s)^T \phi(\mathbf{x}_t) \\ + \text{trace}\left\{ \lambda^T \textstyle\sum_{t=1}^T (\alpha_t - \alpha_t^*)\phi(\mathbf{x}_t)^T \right\} + \frac{3}{2} \cdot \text{trace}\left\{ \lambda^T \lambda \right\} \end{aligned} \tag{5.28}$$

Combining this with the observation that (5.18) results in $\text{trace}\{\lambda^T E\} = 0$ and simplifying terms, we arrive at the relation:

$$\begin{aligned} g(\alpha_t, \alpha_t^*, \lambda) = -\frac{1}{2} \sum_{t=1}^T \sum_{s=1}^T (\alpha_t - \alpha_t^*)^T (\alpha_s - \alpha_s^*)\phi(\mathbf{x}_s)^T \phi(\mathbf{x}_t) - \frac{1}{2} \cdot \text{trace}\left\{ \lambda^T \lambda \right\} \\ - \sum_{t=1}^T (\alpha_t + \alpha_t^*)^T \epsilon + \sum_{t=1}^T (\alpha_t - \alpha_t^*)^T \mathbf{y}_t - \sum_{t=1}^T (\alpha_t - \alpha_t^*)^T \lambda \phi(\mathbf{x}_t) \end{aligned} \tag{5.29}$$

Then, we substitute $K(\mathbf{x}_s, \mathbf{x}_t)$ for $\phi(\mathbf{x}_s)^T \phi(\mathbf{x}_t)$ terms for all $s, t \in \{1, \ldots, T\}$ using the kernel trick. We have then shown that $g(\alpha_t, \alpha_t^*, \lambda) = -f(\alpha_t, \alpha_t^*, \lambda)$ as defined in Theorem 7, and we have that $\max g(\alpha_t, \alpha_t^*, \lambda) = \min f(\alpha_t, \alpha_t^*, \lambda)$. The (5.22b) constraints of the dual problem follow from the stationarity relations (5.20) plus the dual feasibility constraints $\alpha_{m,t}, \alpha_{m,t}^*, \beta_{m,t}, \beta_{m,t}^* \geq 0$ for all $t \in \{1, \ldots, T\}$, $m \in \{1, \ldots, M\}$. This completes the proof.

5.B Variation on State Equation with Linear Term

In our work, we also considered the state equation with a linear term:

$$\mathbf{y}_t = W\phi(\mathbf{x}_t) + \mathbf{r}, \qquad \forall t \in \{1, \dots, T\} \tag{5.30}$$

which is a typical form of the state equation in classical SVR. Note that this form does not make as much sense as (5.11) for power flow equations without error since they can be exactly written as (5.11) as discussed in Section 5.3.1.

However, we thought this version of the state equation (5.30) may be helpful in the case where there is measurement discrepancy due to unobserved buses, as discussed in Section 5.4.2. In this case, instead of lumping the discrepancy error into the ξ and ξ^* penalty terms in (5.12), we could consider that the discrepancy error is part of the $\mathbf{r}$ term in Equation (5.30).

For this modified formulation, the same analysis holds but the dual problem (5.22) is modified to have the following constraint:

$$\sum_{t=1}^{T} \alpha_t - \alpha_t^* = \mathbf{0} \tag{5.31}$$

where the derivation of this constraint comes from taking the derivative of the modified Lagrangian with respect to $\mathbf{r}$.

In Figure 5.8, we show simulations with this formulation on the 30-bus network in the case where we have SCADA measurement noise and varying levels of PMU penetration.

5.C Alternative Formulation with Corrupted PMU Measurements

The SVR formulation in Section 5.3.2 is flexible enough to handle other types of cyberattacks and uncertainty in the system. For example, in the case where there could be some corrupted or attacked PMU measurements and we want to explicitly model those attacks to solve for their attack values (as opposed to the formulation in Section 5.3.2 which treats these attacks as outliers), we can define the corrupted voltage magnitude as $|\hat{v}_i|$ and the corrupted voltage angle as $\hat{\theta}_i$. We relate the corrupted values to the true values for voltage magnitude $|v_i|$ and angle θ_i with the equations:

$$|v_i| = \kappa_i |\hat{v}_i| \tag{5.32a}$$

$$\theta_i = \hat{\theta}_i + \zeta_i \tag{5.32b}$$

where κ_i and ζ_i are the unknown corruption terms associated with the measurements. Note that we distinguish these corruption terms from normal noise or errors in the PMU measurements that fluctuate.

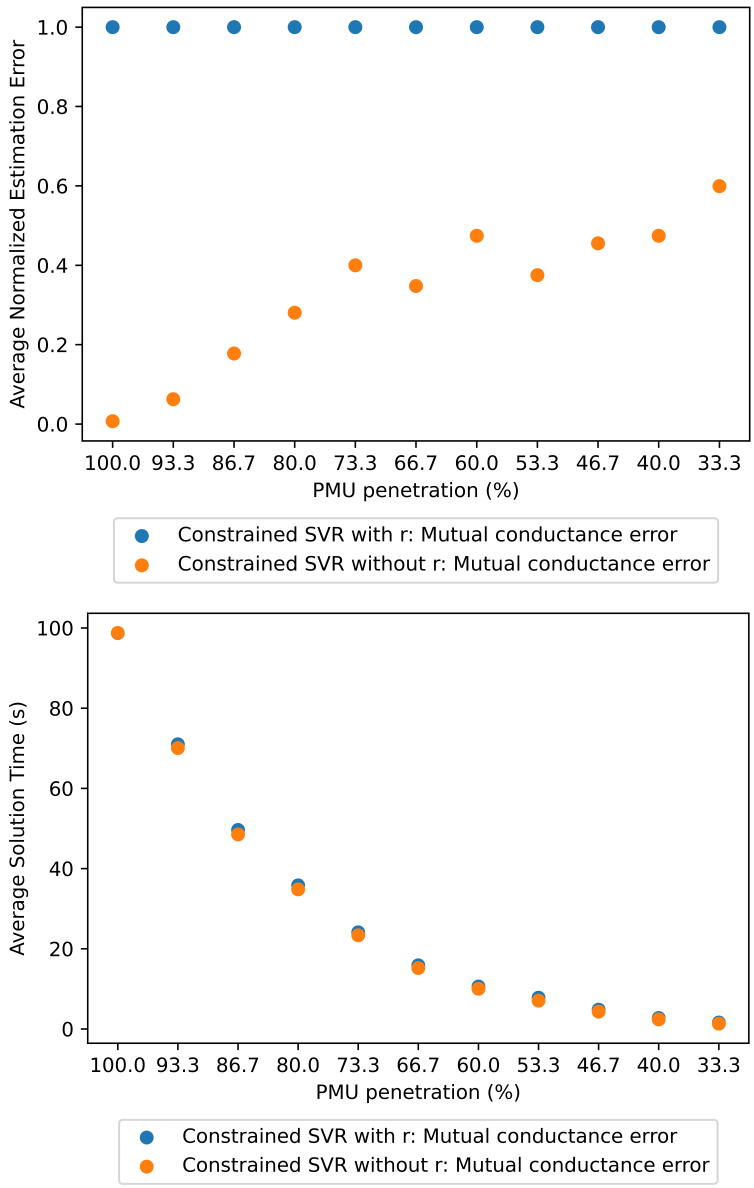

Figure 5.8: Comparison of solution quality (top figure) and solution time (bottom figure) for constrained SVR variations with and without the parameter $\mathbf{r}$ in the state equation (see Equation 5.30). From these simulations, we found that including the parameter $\mathbf{r}$ in the state equation resulted in worse outcomes in terms of line parameter recovery, even in the case of missing PMU measurements.

For this formulation, we define the state vector in relation to $\hat{d}_i = |\hat{v}_i|\cos(\hat{\theta}_i)$ and $\hat{e}_i = |\hat{v}_i|\sin(\hat{\theta}_i)$:

$$\mathbf{x} \triangleq \begin{bmatrix}\hat{d}_1 & \hat{d}_2 & \dots & \hat{d}_n & \hat{e}_1 & \hat{e}_2 & \dots & \hat{e}_n\end{bmatrix}^T \tag{5.33}$$

Then, the feature mapping corresponding to the quadratic kernel $\phi(\mathbf{x}) \in \mathbb{R}^D$ is given by:

$$\phi(\mathbf{x}) = [\hat{d}_1^2 \ \dots \ \hat{d}_n^2 \ \hat{e}_1^2 \ \dots \ \hat{e}_n^2 \ \sqrt{2}\hat{d}_1\hat{d}_2 \ \dots \ \sqrt{2}\hat{d}_1\hat{e}_2 \ \dots \ \sqrt{2}\hat{e}_{n-1}\hat{e}_n]^T \tag{5.34}$$

Similar to the formulation in Section 5.3.1, we can rewrite the real and reactive power flow and injection measurement relations given by Equations (5.2) and (5.3) as a dot product of the quadratic feature mapping $\phi(\mathbf{x})$ given in (5.34) and a specific parameter vector with known structure and unknown values. The power flow measurement equations (5.2a) and (5.2b) can be written as (5.7) when we take $\phi(\mathbf{x})$ as that in (5.34) and define $\mu_{p_{ij}} \in \mathbb{R}^D$ and $\mu_{q_{ij}} \in \mathbb{R}^D$ as having the k^{th} entries:

$$(\mu_{p_{ij}})_k \triangleq \begin{cases} \frac{1}{\sqrt{2}}G_{ij}\kappa_i\kappa_j\cos(\zeta_i-\zeta_j) + \frac{1}{\sqrt{2}}B_{ij}\kappa_i\kappa_j\sin(\zeta_i-\zeta_j), & \text{if } \phi(\mathbf{x}^{ij})_k = \sqrt{2}\hat{d}_i\hat{d}_j \text{ or } \sqrt{2}\hat{e}_i\hat{e}_j \\ -\frac{1}{\sqrt{2}}G_{ij}\kappa_i\kappa_j\sin(\zeta_i-\zeta_j) + \frac{1}{\sqrt{2}}B_{ij}\kappa_i\kappa_j\cos(\zeta_i-\zeta_j), & \text{if } \phi(\mathbf{x}^{ij})_k = \sqrt{2}\hat{e}_i\hat{d}_j \\ \frac{1}{\sqrt{2}}G_{ij}\kappa_i\kappa_j\sin(\zeta_i-\zeta_j) - \frac{1}{\sqrt{2}}B_{ij}\kappa_i\kappa_j\cos(\zeta_i-\zeta_j), & \text{if } \phi(\mathbf{x}^{ij})_k = \sqrt{2}\hat{d}_i\hat{e}_j \\ 0, & \text{otherwise} \end{cases} \tag{5.35}$$

$$(\mu_{q_{ij}})_k \triangleq \begin{cases} \frac{1}{\sqrt{2}}G_{ij}\kappa_i\kappa_j\sin(\zeta_i-\zeta_j) - \frac{1}{\sqrt{2}}B_{ij}\kappa_i\kappa_j\cos(\zeta_i-\zeta_j), & \text{if } \phi(\mathbf{x}^{ij})_k = \sqrt{2}\hat{d}_i\hat{d}_j \text{ or } \sqrt{2}\hat{e}_i\hat{e}_j \\ \frac{1}{\sqrt{2}}G_{ij}\kappa_i\kappa_j\cos(\zeta_i-\zeta_j) + \frac{1}{\sqrt{2}}B_{ij}\kappa_i\kappa_j\sin(\zeta_i-\zeta_j), & \text{if } \phi(\mathbf{x}^{ij})_k = \sqrt{2}\hat{e}_i\hat{d}_j \\ -\frac{1}{\sqrt{2}}G_{ij}\kappa_i\kappa_j\cos(\zeta_i-\zeta_j) - \frac{1}{\sqrt{2}}B_{ij}\kappa_i\kappa_j\sin(\zeta_i-\zeta_j), & \text{if } \phi(\mathbf{x}^{ij})_k = \sqrt{2}\hat{d}_i\hat{e}_j \\ 0, & \text{otherwise} \end{cases} \tag{5.36}$$

where the proof for these relations can be given by expanding (5.7) with the definitions of $\mu_{p_{ij}}$, $\mu_{q_{ij}}$, and $\phi(\mathbf{x})$ and simplifying with trigonometric identities.

Similarly, for real and reactive power injection, we can rewrite the power injection measurements (5.3) as (5.9) using the definitions of μ_{p_i} and μ_{q_i} as:

$$(\mu_{p_i})_k \triangleq \begin{cases} G_{ii}\kappa_i^2, & \text{if } \phi(\mathbf{x})_k = \hat{d}_i^2 \text{ or } \hat{e}_i^2 \\ (\mu_{p_{ij}})_k, & \text{otherwise} \end{cases} \tag{5.37}$$

$$(\mu_{q_i})_k \triangleq \begin{cases} -B_{ii}\kappa_i^2, & \text{if } \phi(\mathbf{x})_k = \hat{d}_i^2 \text{ or } \hat{e}_i^2 \\ (\mu_{q_{ij}})_k, & \text{otherwise} \end{cases} \tag{5.38}$$

Using this formulation, if the G_{ij} and B_{ij} parameters are unavailable, we would be able to learn expressions for $\frac{1}{\sqrt{2}}G_{ij}\kappa_i\kappa_j\sin(\zeta_i-\zeta_j) - \frac{1}{\sqrt{2}}B_{ij}\kappa_i\kappa_j\cos(\zeta_i-\zeta_j)$,

$-\frac{1}{\sqrt{2}}G_{ij}\kappa_i\kappa_j\sin(\zeta_i-\zeta_j)+\frac{1}{\sqrt{2}}B_{ij}\kappa_i\kappa_j\cos(\zeta_i-\zeta_j)$, $G_{ii}\kappa_i^2$, and $-B_{ii}\kappa_i^2$. While this formulation would not be able to exactly recover the true line parameters and corruption values, it would be able to learn a sparse mapping between the PMU measurements and the SCADA measurements. In the case where some G_{ij}, B_{ij}, G_{ii}, and B_{ii} parameters are available, it may be possible to solve for some of the corruption values κ_i and $\zeta_i-\zeta_j$ given estimates for $\mu_{p_{ij}}$, $\mu_{q_{ij}}$, μ_{p_i}, and μ_{q_i}.

5.D Sequential Minimal Optimization (SMO) Algorithm for Constrained SVR

In this section, we propose a SMO algorithm for the dual form of the constrained SVR problem in (5.22). Based on the existing SMO algorithm [52], we propose a method that chooses the most violating set of variables in the α and α^* blocks (where most violating is defined in relation to the first-order optimality conditions), optimize a subproblem in these blocks, and then solve a closed-form optimization problem to find the value of λ. The classical SMO algorithm was shown to converge in [53] and [54] proved its linear convergence. In some simulations, we found that the modified SMO algorithm for constrained SVR converges on small test cases, but that the built-in QP solver in Gurobi outperforms the current SMO implementation in Python. A faster implementation of SMO is needed for a true comparison of methods.

5.D.1 Derivation of Optimality Conditions for Dual (5.22)

Let $\sigma \geq 0$ and $\sigma^* \geq 0$ be the Lagrangian multipliers corresponding to the respective constraints $\alpha \geq 0$ and $\alpha^* \geq 0$ in (5.22), let $\gamma \geq 0$ and $\gamma^* \geq 0$ be the Lagrangian multipliers corresponding to the respective constraints $\alpha \leq C$ and $\alpha^* \leq C$ in (5.22), and let η_r correspond to the Lagrangian multipliers of the (5.22c) constraints. The Lagrangian of (5.22) is then:

$$L = f(\alpha,\alpha^*,\lambda) - \sum_{t=1}^{T}\sum_{m=1}^{M}\sigma_{m,t}\alpha_{m,t} - \sum_{t=1}^{T}\sum_{m=1}^{M}\sigma^*_{m,t}\alpha^*_{m,t} - \sum_{t=1}^{T}\sum_{m=1}^{M}\gamma_{m,t}(C-\alpha_{m,t}) - \sum_{t=1}^{T}\sum_{m=1}^{M}\gamma^*_{m,t}(C-\alpha^*_{m,t}) + \sum_{r=1}^{R}\eta_r \text{trace}\{\lambda^T L_r\} \quad (5.39)$$

Then, the stationarity of the Lagrangian is given by the equations:

$$\frac{\partial L}{\partial \alpha_{m,t}} = \frac{\partial f}{\partial \alpha_{m,t}} - \sigma_{m,t} + \gamma_{m,t} = 0 \quad (5.40a)$$

$$\frac{\partial L}{\partial \alpha^*_{m,t}} = \frac{\partial f}{\partial \alpha^*_{m,t}} - \sigma^*_{m,t} + \gamma^*_{m,t} = 0 \quad (5.40b)$$

$$\frac{\partial L}{\partial \lambda_{i,j}} = \frac{\partial f}{\partial \lambda_{i,j}} + \sum_{r=1}^{R} \eta_r L_{r,i,j} \qquad = 0 \tag{5.40c}$$

The equations for complementary slackness are:

$$\sigma_{m,t}\alpha_{m,t} = 0 \tag{5.41a}$$
$$\sigma^*_{m,t}\alpha^*_{m,t} = 0 \tag{5.41b}$$
$$\gamma_{m,t}(C - \alpha_{m,t}) = 0 \tag{5.41c}$$
$$\gamma^*_{m,t}(C - \alpha^*_{m,t}) = 0 \tag{5.41d}$$

We can consider different cases for various $\alpha_{m,t}$:

- Case 1: $\alpha_{m,t} = 0$ which implies that $\sigma_{m,t} \geq 0$ and $\gamma_{m,t} = 0$:

$$\frac{\partial f}{\partial \alpha_{m,t}} - \sigma_{m,t} + \gamma_{m,t} = 0 \qquad \Rightarrow \qquad \frac{\partial f}{\partial \alpha_{m,t}} \geq 0 \tag{5.42}$$

- Case 2: $\alpha_{m,t} = C$ which implies that $\sigma_{m,t} = 0$ and $\gamma_{m,t} \geq 0$:

$$\frac{\partial f}{\partial \alpha_{m,t}} - \sigma_{m,t} + \gamma_{m,t} = 0 \qquad \Rightarrow \qquad \frac{\partial f}{\partial \alpha_{m,t}} \leq 0 \tag{5.43}$$

- Case 3: $0 < \alpha_{m,t} < C$ which implies that $\sigma_{m,t} = 0$ and $\gamma_{m,t} = 0$:

$$\frac{\partial f}{\partial \alpha_{m,t}} - \sigma_{m,t} + \gamma_{m,t} = 0 \qquad \Rightarrow \qquad \frac{\partial f}{\partial \alpha_{m,t}} = 0 \tag{5.44}$$

We have similar relations for $\alpha^*_{t,i}$ cases. We will use this form of the optimality conditions of (5.22) in order to derive the SMO algorithm for the constrained SVR problem. Next, we will derive a closed-form solution for λ given α and α^* that will be useful for the algorithm.

5.D.2 Subproblem for λ

Given that α, α^* are fixed, we can derive a closed-form solution for λ in (5.22). Removing terms in (5.22) that do not depend on λ, we arrive at the problem:

$$\min_{\lambda} \ \frac{1}{2}\text{trace}\{\lambda^T\lambda\} + \sum_{t=1}^{T}(\alpha_t - \alpha^*_t)^T\lambda\phi(\mathbf{x}_t) \tag{5.45a}$$
$$\text{subject to: } \text{trace}\{\lambda^T L_r\} = 0, \ \ \forall r \in \{1, \ldots, R\} \tag{5.45b}$$

Introducing η_r corresponding to the equality constraints, we can write the Lagrangian of (5.45) as:

$$L = \frac{1}{2}\text{trace}\{\lambda^T\lambda\} + \sum_{t=1}^{T}(\alpha_t - \alpha_t^*)^T\lambda\phi(\mathbf{x}_t) + \sum_{r=1}^{R}\eta_r\text{trace}\{\lambda^T L_r\} \tag{5.46}$$

Taking the stationarity of the Lagrangian with respect to λ, we get the relation:

$$\lambda = -\sum_{t=1}^{T}(\alpha_t - \alpha_t^*)\phi^T(\mathbf{x}_t) - \sum_{r=1}^{R}\eta_r L_r \tag{5.47}$$

Plugging this back into the Lagrangian and cancelling some terms, we have:

$$g(\eta) = -\frac{1}{2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\alpha_t - \alpha_t^*)^T(\alpha_s - \alpha_s^*)\phi^T(\mathbf{x}_s)\phi(\mathbf{x}_t) - \sum_{t=1}^{T}\sum_{r=1}^{R}\eta_r(\alpha_t - \alpha_t^*)^T L_r\phi(\mathbf{x}_t) - \frac{1}{2}\sum_{r=1}^{R}\sum_{s=1}^{R}\eta_r\eta_s\text{trace}\left\{L_r^T L_s\right\} \tag{5.48}$$

Thus, the dual of (5.45) is given by:

$$\min_{\eta}\ 1/2\cdot\eta^T\tilde{A}\eta + \tilde{b}^T\eta + \tilde{c} \tag{5.49}$$

where we have defined:

- Matrix $\tilde{A} \in \mathbb{S}^R$ with entries $\tilde{A}_{s,r} \triangleq \text{trace}\left\{L_r^T L_s\right\}$
- Vector $\tilde{b} \in \mathbb{R}^R$ with entries $\tilde{b}_r \triangleq \sum_{t=1}^{T}(\alpha_t - \alpha_t^*)^T L_r\phi(\mathbf{x}_t)$
- Scalar $\tilde{c} \triangleq \frac{1}{2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\alpha_t - \alpha_t^*)^T(\alpha_s - \alpha_s^*)\phi^T(\mathbf{x}_s)\phi(\mathbf{x}_t)$

With these definitions, $g(\eta)$ can be rewritten as $g(\eta) = -\tilde{c} - \sum_{r=1}^{R}\eta_r\tilde{b}_r - \frac{1}{2}\sum_{r=1}^{R}\sum_{s=1}^{R}\eta_r\eta_s\tilde{A}_{s,r}$.

To optimize (5.49), we solve the first-order stationarity equation:

$$\tilde{A}\eta + \tilde{b} = 0 \tag{5.50}$$

which has a closed-form solution as long as A is invertible.

Thus, we see that we can solve for η, given α and α^*. To find λ from η, we simply note Equation (5.47).

5.D.3 Derivation of SMO Algorithm for Constrained SVR

Based on the optimality conditions given in Appendix 5.D.1, we will define a variable $\alpha_{m,t}$ to be a *violating variable* if one of the following is true:

$$(m,t) \in MT_{\text{up}} \text{ and } \frac{\partial f}{\partial \alpha_{m,t}} < 0 \tag{5.51a}$$

or

$$(m,t) \in MT_{\text{low}} \text{ and } \frac{\partial f}{\partial \alpha_{m,t}} > 0 \tag{5.51b}$$

where the sets MT_{up} and MT_{low} are defined as:

$$MT_{\text{up}}(\alpha) \triangleq \{(m,t) \in \{1,\dots,M\} \times \{1,\dots,T\} \ : \ \alpha_{m,t} < C\} \tag{5.52a}$$

$$MT_{\text{low}}(\alpha) \triangleq \{(m,t) \in \{1,\dots,M\} \times \{1,\dots,T\} \ : \ \alpha_{m,t} > 0\} \tag{5.52b}$$

For an exact solution in a finite number of steps, we will define τ-*violating variables* as:

$$(m,t) \in MT_{\text{up}} \text{ and } \frac{\partial f}{\partial \alpha_{m,t}} < \tau \tag{5.53a}$$

or

$$(m,t) \in MT_{\text{low}} \text{ and } \frac{\partial f}{\partial \alpha_{m,t}} > \tau \tag{5.53b}$$

for some small $\tau > 0$. We have similar equations as (5.53) for the α^* variables as well.

Then, at each iteration, we will choose the most violating variables out of MT_{up}, MT_{low}, MT^*_{up}, and MT^*_{low}:

$$\Delta_{\text{up}} \leftarrow -\min_{(m,t)\in MT_{\text{up}}} \frac{\partial f}{\partial \alpha_{m,t}} \qquad (m,t) \leftarrow \operatorname{argmin}_{(m,t)\in MT_{\text{up}}} \frac{\partial f}{\partial \alpha_{m,t}} \tag{5.54a}$$

$$\Delta_{\text{low}} \leftarrow \max_{(l,s)\in MT_{\text{low}}} \frac{\partial f}{\partial \alpha_{l,s}} \qquad (l,s) \leftarrow \operatorname{argmax}_{(l,s)\in MT_{\text{low}}} \frac{\partial f}{\partial \alpha_{l,s}} \tag{5.54b}$$

$$\Delta^*_{\text{up}} \leftarrow -\min_{(m,t)\in MT^*_{\text{up}}} \frac{\partial f}{\partial \alpha^*_{m,t}} \qquad (m^*,t^*) \leftarrow \operatorname{argmin}_{(m,t)\in MT^*_{\text{up}}} \frac{\partial f}{\partial \alpha^*_{m,t}} \tag{5.54c}$$

$$\Delta^*_{\text{low}} \leftarrow \max_{(l,s)\in MT^*_{\text{low}}} \frac{\partial f}{\partial \alpha^*_{l,s}} \qquad (l^*,s^*) \leftarrow \operatorname{argmax}_{(l,s)\in MT^*_{\text{low}}} \frac{\partial f}{\partial \alpha^*_{l,s}} \tag{5.54d}$$

For the SMO algorithm, we can select either the maximum violating variable (i.e. $\Delta = \max\{\Delta_{\text{up}}, \Delta_{\text{low}}, \Delta^*_{\text{up}}, \Delta^*_{\text{low}}\}$) and just modify that variable, or we can select some subset of the maximum violating variables and modify those. In Algorithm 5, we have presented a

form of the algorithm in which we modify all four of the maximum violating variables at each iteration of the algorithm.

In the next section, we will see that each of the subproblems for $\alpha_{m,t}$ and $\alpha_{l,s}$ are independent of each other (similarly for $\alpha^*_{m,t}$ and $\alpha^*_{l,s}$), given that $(m,t) \neq (l,s)$ which is true considering how we define the violating variables in (5.54d) as long as some variables are still considered to be τ-violating (i.e. $\max\{\Delta_{\text{up}}, \Delta_{\text{low}}, \Delta^*_{\text{up}}, \Delta^*_{\text{low}}\} > \tau$).

5.D.4 Subproblems for α and α^*

For any α update at iteration $k+1$, we will only modify the variable by up to two coordinates: (m,t) and (l,s) where $\alpha_{m,t}$ and $\alpha_{l,s}$ are taken to be τ-violating variables as defined in the preceding section. Let a be a matrix of zeros except for the $(m,t)^{\text{th}}$ entry which is δ_m and the $(l,s)^{\text{th}}$ entry which is δ_l. We can find values for these entries by minimizing the following optimization problem that rewrites the stacked form in (5.24c) using the relation $\alpha^{(k+1)} = \alpha^{(k)} + a$:

$$\min_a \ \frac{1}{2}\text{trace}\{(Z^{(k)}+a)^T A_0 (Z^{(k)}+a)\} + \text{trace}\{A_1^T (Z^{(k)}+a)\} \tag{5.55a}$$

$$\text{subject to: } \alpha^{(k)}_{m,t} + \delta_m \geq 0 \tag{5.55b}$$

$$\alpha^{(k)}_{m,t} + \delta_m \leq C \tag{5.55c}$$

$$\alpha^{(k)}_{l,s} + \delta_l \geq 0 \tag{5.55d}$$

$$\alpha^{(k)}_{l,s} + \delta_l \leq C \tag{5.55e}$$

We can write the objective of Problem (5.55) as:

$$\Psi(a) = \frac{1}{2}\text{trace}\{a^T A_0 a\} + \text{trace}\{a^T A_0 Z^{(k)}\} + \text{trace}\{A_1^T a\} + K \tag{5.56}$$

where $K \triangleq \frac{1}{2}\text{trace}\{(Z^{(k)})^T A_0 Z^{(k)}\} + \text{trace}\{A_1^T Z^{(k)}\}$ are the terms that do not depend on a.

We also have that $\nabla_Z f(Z^{(k)}) = A_0 Z^{(k)} + A_1$, yielding:

$$\Psi(a) = \frac{1}{2}\text{trace}\{a^T A_0 a\} + \text{trace}\{a^T \nabla_Z f(Z^{(k)})\} + K \tag{5.57}$$

Substituting in the structure of a, we have:

$$\Psi(\delta_m, \delta_l) = \frac{1}{2}\delta_m^2 [A_0]_{t,t} + \frac{1}{2}\delta_l^2 [A_0]_{s,s} + \delta_m \frac{\partial f}{\partial \alpha_{m,t}}(Z^{(k)}) + \delta_l \frac{\partial f}{\partial \alpha_{l,s}}(Z^{(k)}) + K \tag{5.58}$$

To find the unconstrained minimum, we have:

$$\frac{\partial \Psi}{\partial \delta_m} = \delta_m [A_0]_{t,t} + \frac{\partial f}{\partial \alpha_{m,t}}(Z^{(k)}) = 0 \tag{5.59a}$$

$$\frac{\partial \Psi}{\partial \delta_l} = \delta_l [A_0]_{s,s} + \frac{\partial f}{\partial \alpha_{l,s}}(Z^{(k)}) = 0 \tag{5.59b}$$

Solving these equations and clipping at the constraints, we have:

$$\delta_m = \max\left(\min\left(-\frac{1}{[A_0]_{t,t}}\frac{\partial f(Z^{(k)})}{\partial \alpha_{m,t}}, C - \alpha_{m,t}^{(k)}\right), -\alpha_{m,t}^{(k)}\right) \tag{5.60a}$$

$$\delta_l = \max\left(\min\left(-\frac{1}{[A_0]_{s,s}}\frac{\partial f(Z^{(k)})}{\partial \alpha_{l,s}}, C - \alpha_{l,s}^{(k)}\right), -\alpha_{l,s}^{(k)}\right) \tag{5.60b}$$

$$\alpha_{m,t}^{(k+1)} \leftarrow \alpha_{m,t}^{(k)} + \delta_m \tag{5.60c}$$

$$\alpha_{l,s}^{(k+1)} \leftarrow \alpha_{l,s}^{(k)} + \delta_l \tag{5.60d}$$

We can derive similar update equations for $\alpha_{m,t}^*$ and $\alpha_{l,s}^*$. Then, we can define the modified SMO algorithm for constrained SVR in Algorithm 5.

Algorithm 5 Modified SMO for Constrained SVR

Initialize: $\tau > 0$; $\alpha^{(0)}, (\alpha^*)^{(0)} \in \mathbb{R}^{M \times T}$ and set $k = 0$
while $\Delta > \tau$ **do**
$\Delta_{\text{up}} \leftarrow -\min_{(m,t) \in MT_{\text{up}}} \frac{\partial f}{\partial \alpha_{m,t}}$ $\quad (m,t) \leftarrow \operatorname{argmin}_{(m,t) \in MT_{\text{up}}} \frac{\partial f}{\partial \alpha_{m,t}}$
$\Delta_{\text{low}} \leftarrow \max_{(l,s) \in MT_{\text{low}}} \frac{\partial f}{\partial \alpha_{l,s}}$ $\quad (l,s) \leftarrow \operatorname{argmax}_{(l,s) \in MT_{\text{low}}} \frac{\partial f}{\partial \alpha_{l,s}}$
$\Delta^*_{\text{up}} \leftarrow -\min_{(m,t) \in MT^*_{\text{up}}} \frac{\partial f}{\partial \alpha^*_{m,t}}$ $\quad (m^*,t^*) \leftarrow \operatorname{argmin}_{(m,t) \in MT^*_{\text{up}}} \frac{\partial f}{\partial \alpha^*_{m,t}}$
$\Delta^*_{\text{low}} \leftarrow \max_{(l,s) \in MT^*_{\text{low}}} \frac{\partial f}{\partial \alpha^*_{l,s}}$ $\quad (l^*,s^*) \leftarrow \operatorname{argmax}_{(l,s) \in MT^*_{\text{low}}} \frac{\partial f}{\partial \alpha^*_{l,s}}$
$\alpha^{(k+1)} \leftarrow$ Solution of subproblem for $\alpha_{m,t}$ and $\alpha_{l,s}$ given by (5.60)
$(\alpha^*)^{(k+1)} \leftarrow$ Solution of subproblem for $\alpha^*_{m,t}$ and $\alpha^*_{l,s}$
Given $\alpha^{(k+1)}, (\alpha^*)^{(k+1)}$, find closed-form update of $\lambda^{(k+1)}$
$k \leftarrow k + 1$
end while

Note that for the state equation with a linear violation term given in (5.30), we can also consider a modified SMO algorithm. However, in this formulation, due to constraint (5.31), the subproblems for $\alpha_{m,t}$, $\alpha_{l,s}$, $\alpha^*_{m,t}$ and $\alpha^*_{l,s}$ are coupled. In this case, we select a maximum violating pair of α or α^* variables and solve a one-dimensional subproblem that optimizes these variables, just as in [52]. Then, we solve for a closed-form update of λ given α and α^*.

5.D.5 Testing of SMO Algorithm

The results using Algorithm 5 to solve (5.22) compared to the results of using the GUROBI QP solver to solve (5.22) are given in Figure 5.9.

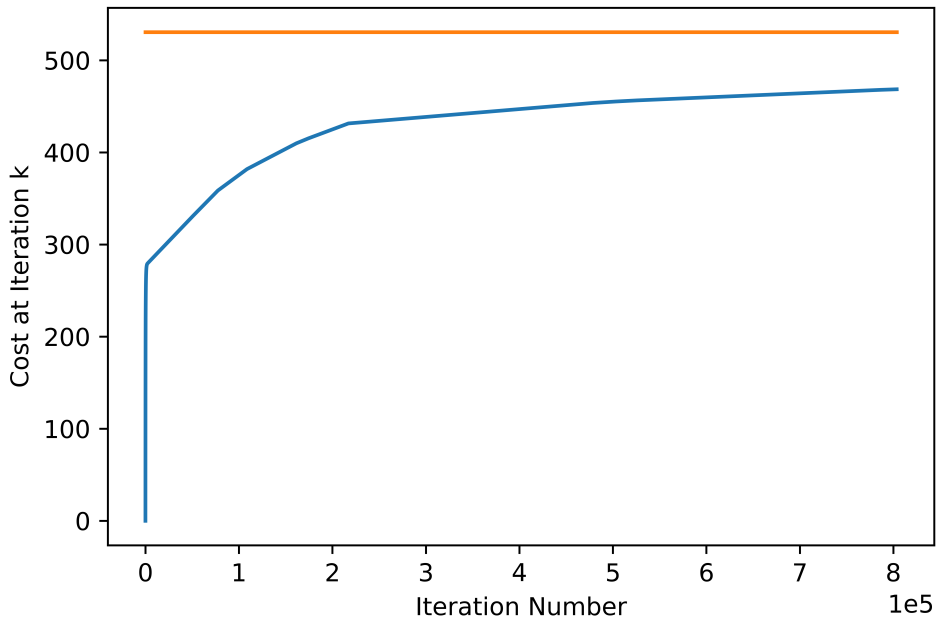

Figure 5.9: Modified Sequential Minimal Optimization (SMO) for Constrained SVR problem. The results of Algorithm 5 on the 14-bus test case with SCADA measurement noise (SNR=40dB) are shown. The blue line shows the SMO objective at iteration k, and the orange line is the objective found using GUROBI to solve the QP in (5.22).

5.5 Bibliography

[1] E. Glista and S. Sojoudi, "Leveraging the physics of AC power flow in support vector regression to identify power system topology," in *submitted to the 2023 Conference on Decision and Control (CDC)*, 2023.

[2] U.S. Department of Energy, "Grid modernization and the smart grid." [Online]. Available: https://www.energy.gov/oe/grid-modernization-and-smart-grid

[3] M. Jin, J. Lavaei, and K. H. Johansson, "Power grid AC-based state estimation: Vulnerability analysis against cyber attacks," *IEEE Transactions on Automatic Control*, vol. 64, no. 5, pp. 1784–1799, 2019.

[4] M. Jin, J. Lavaei, S. Sojoudi, and R. Baldick, "Boundary defense against cyber threat for power system state estimation," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 1752–1767, 2021.

[5] A. Wood, B. Wollenberg, and G. Sheblé, *Power Generation, Operation, and Control*. Wiley, 2013.

[6] M. Huneault and F. D. Galiana, "A survey of the optimal power flow literature," *IEEE Transactions on Power Systems*, vol. 6, no. 2, pp. 762–770, 1991.

[7] B. Knueven, J. Ostrowski, and J.-P. Watson, "On mixed-integer programming formulations for the unit commitment problem," *INFORMS Journal on Computing*, vol. 32, no. 4, pp. 857–876, 2020.

[8] S. Frank, I. Steponavice, and S. Rebennack, "Optimal power flow: A bibliographic survey I: Formulations and deterministic methods," *Energy systems*, vol. 3, pp. 221–258, 2012.

[9] A. Monticelli, "Electric power system state estimation," *Proceedings of the IEEE*, vol. 88, no. 2, pp. 262–282, 2000.

[10] G. Liang, J. Zhao, F. Luo, S. R. Weller, and Z. Y. Dong, "A review of false data injection attacks against modern power systems," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1630–1638, 2016.

[11] J. Yu, Y. Weng, and R. Rajagopal, "Robust mapping rule estimation for power flow analysis in distribution grids," in *2017 North American Power Symposium (NAPS)*, 2017, pp. 1–6.

[12] P. Zarco and A. Exposito, "Power system parameter estimation: A survey," *IEEE Transactions on Power Systems*, vol. 15, no. 1, pp. 216–222, 2000.

[13] F. Basiri, J. Casadiego, M. Timme, and D. Witthaut, "Inferring power-grid topology in the face of uncertainties," *Physical Review E*, vol. 98, no. 1, p. 012305, 2018.

[14] Y. Sharon, A. M. Annaswamy, A. L. Motto, and A. Chakraborty, "Topology identification in distribution network with limited measurements," in *2012 IEEE PES Innovative Smart Grid Technologies (ISGT)*, 2012, pp. 1–6.

[15] M. Delimar, I. Pavic, and Z. Hebel, "Artificial neural networks in power system topology recognition," in *The IEEE Region 8 EUROCON 2003. Computer as a Tool.*, vol. 2, 2003, pp. 287–291 vol.2.

[16] X. Hu, H. Hu, S. Verma, and Z.-L. Zhang, "Physics-guided deep neural networks for power flow analysis," *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 2082–2092, 2021.

[17] C. Wang, J. An, and G. Mu, "Power system network topology identification based on knowledge graph and graph neural network," *Frontiers in Energy Research*, vol. 8, p. 613331, 2021.

[18] S. B. Efe and M. Cebeci, "Power flow analysis by artificial neural network," *International Journal of Energy and Power Engineering*, vol. 2, p. 204, 2013.

[19] L. Imen, L. Djamel, S. Hassiba, D. Abdellah, and F. Selwa, "Optimal power flow study using conventional and neural networks methods," in *2015 International Conference on Renewable Energy Research and Applications (ICRERA)*, 2015, pp. 1422–1427.

[20] K. Baker, "Emulating AC OPF solvers with neural networks," *IEEE Transactions on Power Systems*, vol. 37, no. 6, pp. 4950–4953, 2022.

[21] J. Yuan and Y. Weng, "Support matrix regression for learning power flow in distribution grid with unobservability," *IEEE Transactions on Power Systems*, vol. 37, no. 2, pp. 1151–1161, 2022.

[22] S. Sojoudi and J. Lavaei, "Physics of power networks makes hard optimization problems easy to solve," *2012 IEEE Power and Energy Society General Meeting*, pp. 1–8, 2012.

[23] R. Madani, A. Kalbat, and J. Lavaei, "ADMM for sparse semidefinite programming with applications to optimal power flow problem," in *2015 54th IEEE Conference on Decision and Control (CDC)*. IEEE, 2015, pp. 5932–5939.

[24] D. K. Molzahn and I. A. Hiskens, "Sparsity-exploiting moment-based relaxations of the optimal power flow problem," *IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 3168–3180, 2014.

[25] R. A. Jabr, "Exploiting sparsity in SDP relaxations of the OPF problem," *IEEE Transactions on Power Systems*, vol. 27, no. 2, pp. 1138–1139, 2011.

[26] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. Association for Computing Machinery, 1992, p. 144–152.

[27] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," in *Advances in Neural Information Processing Systems*, vol. 9. MIT Press, 1996.

[28] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.

[29] O. Chapelle, P. Haffner, and V. Vapnik, "Support vector machines for histogram-based image classification," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 1055–1064, 1999.

[30] C. Schuldt, I. Laptev, and B. Caputo, "Recognizing human actions: A local SVM approach," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 3, 2004, pp. 32–36 Vol.3.

[31] Y. Lin, F. Lv, S. Zhu, M. Yang, T. Cour, K. Yu, L. Cao, and T. Huang, "Large-scale image classification: fast feature extraction and SVM training," in *CVPR 2011*. IEEE, 2011, pp. 1689–1696.

[32] B.-J. Chen, M.-W. Chang, and C.-J. Lin, "Load forecasting using support vector machines: A study on EUNITE competition 2001," *IEEE Transactions on Power Systems*, vol. 19, no. 4, pp. 1821–1830, 2004.

[33] W.-C. Hong, "Electric load forecasting by support vector model," *Applied Mathematical Modelling*, vol. 33, no. 5, pp. 2444–2454, 2009.

[34] H. Xu, C. Caramanis, and S. Mannor, "Robust regression and Lasso," *Advances in neural information processing systems*, vol. 21, 2008.

[35] H. Wang and C. Leng, "Unified LASSO estimation by least squares approximation," *Journal of the American Statistical Association*, vol. 102, no. 479, pp. 1039–1048, 2007.

[36] M. A. Rasmussen and R. Bro, "A tutorial on the Lasso approach to sparse modeling," *Chemometrics and Intelligent Laboratory Systems*, vol. 119, pp. 21–31, 2012.

[37] C. K. Liew, "Inequality constrained least-squares estimation," *Journal of the American statistical association*, vol. 71, no. 355, pp. 746–751, 1976.

[38] G. H. Golub and U. Von Matt, "Quadratically constrained least squares and quadratic problems," *Numerische Mathematik*, vol. 59, no. 1, pp. 561–580, 1991.

[39] B. R. Hunt, "The application of constrained least squares estimation to image restoration by digital computer," *IEEE Transactions on Computers*, vol. C-22, no. 9, pp. 805–812, 1973.

[40] V. Franc, V. Hlaváč, and M. Navara, "Sequential coordinate-wise algorithm for the non-negative least squares problem," in *Computer Analysis of Images and Patterns: 11th International Conference, CAIP 2005, Versailles, France, September 5-8, 2005. Proceedings 11.* Springer, 2005, pp. 407–414.

[41] D. Kim, S. Sra, and I. S. Dhillon, "A non-monotonic method for large-scale non-negative least squares," *Optimization Methods and Software*, vol. 28, no. 5, pp. 1012–1039, 2013.

[42] R. Bro and S. De Jong, "A fast non-negativity-constrained least squares algorithm," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 11, no. 5, pp. 393–401, 1997.

[43] Q. Klopfenstein and S. Vaiter, "Linear support vector regression with linear constraints," *Machine Learning*, vol. 110, no. 7, pp. 1939–1974, 2021.

[44] H. Wu and Giri, "PMU impact on state estimation reliability for improved grid security," in *2005/2006 IEEE/PES Transmission and Distribution Conference and Exhibition*, 2006, pp. 1349–1351.

[45] A. Ashok, M. Govindarasu, and V. Ajjarapu, "Online detection of stealthy false data injection attacks in power system state estimation," *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 1636–1646, 2018.

[46] E. Glista and S. Sojoudi, "A MILP for optimal measurement choice in robust power grid state estimation," in *2022 IEEE Power and Energy Society General Meeting (PESGM)*, 2022.

[47] L. Vanfretti, M. Baudette, and A. D. White, "Monitoring and control of renewable energy sources using synchronized phasor measurements," in *Renewable Energy Integration*, second edition ed., 2017, pp. 419–434.

[48] A. Simões Costa, A. Albuquerque, and D. Bez, "An estimation fusion method for including phasor measurements into power system real-time modeling," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1910–1920, 2013.

[49] S. Dahale and B. Natarajan, "Bayesian framework for multi-timescale state estimation in low-observable distribution systems," *IEEE Transactions on Power Systems*, vol. 37, no. 6, pp. 4340–4351, 2022.

[50] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, "MATPOWER: Steady-state operations, planning and analysis tools for power systems research and education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, February 2011.

[51] M. Brown, M. Biswal, S. Brahma, S. J. Ranade, and H. Cao, "Characterizing and quantifying noise in PMU data," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, 2016, pp. 1–5.

[52] J. Platt, "Sequential minimal optimization: A fast algorithm for training support vector machines," 1998.

[53] S. S. Keerthi and E. G. Gilbert, "Convergence of a generalized SMO algorithm for SVM classifier design," 2002.

[54] J. She and M. Schmidt, "Linear convergence and support vector identification of sequential minimal optimization," in *10th NIPS Workshop on Optimization for Machine Learning*, vol. 5, 2017, p. 50.