Optimization and Learning Methods for Electric Distribution Network Management

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

Ahmed Samir Abdelaal Zamzam

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Nicholas D. Sidiropoulos, Advisor

June, 2019

Acknowledgements

Their are a lot of people who have earned my gratitude for their contribution to my time in graduate school. First, I would like to thank my adviser, Prof. Nikos Sidiropoulos, for providing me the opportunity to work with him. From my first day at the University of Minnesota, Prof. Sidiropoulos believed in me like nobody else and gave me endless support. I believe his valuable guidance as well as his perpetual encouragement made me not only a qualified researcher but also a better person. I would give him most of the credit for becoming the kind of scientist I am today.

My sincere respect and gratitude goes to Prof. Georgios Giannakis. I would like to thank him for being supportive during the toughest time of my life. My conversations with Prof. Giannakis and Prof. Sidiropoulos helped me gain better understanding of important technical matters, and gave me a different perspective on many aspects of life.

Due thanks go to Profs. Yousef Saad, Shuzhong Zhang, Mingyi Hong for agreeing to serve on my committee as well as their valuable comments and feedback on my research and thesis. Thanks also go to Professors Sairaj Dhople, Mehmet Akçakaya, Jarvis Haupt, and Zhi-Quan Luo of the Electrical and Computer Department whose graduate courses helped me build the foundations which made this thesis possible.

During my PhD time, I had the opportunity to collaborate with several excellent researchers, and I have greatly benefited from their brilliant ideas and vision. In particular, I would like to thank Prof. Emiliano Dall'Anese who helped me gain better understanding of the field of power systems. He also gave me the opportunity to be an intern at the National Renewable Energy Laboratory which turned out to be life changing. Thanks go to Prof. Xiao Fu, Dr. Changhong Zhao, and Prof. Josh Taylor for their insightful input to our fruitful collaborations.

The material of this thesis has also benefited from discussions with current and former members of our research group: Magda Amiridi, Faisal Almutairi, Kejun Huang, Charilaos Kanatsoulis, Nikos Kargas, Aritra Konar, Cheng Qian, Mikael Sørensen, John Tranter, and Bo Yang. I would also like to thank these people for their support. I also thank Vassilis Ioannidis, Yanning Shen, and Gang Wang who helped me through collaborations and technical discussions. Special thanks go to Mohamed Salah for setting the bar so high and making it impossible to find another friend as good as him.

My sincere gratitude is due to my brothers, Ayman, Magdy, Omar, and Mostafa, for supporting me, for their precious time and sincerity to care about me. My deepest gratitude and appreciation goes to my friends for their pleasant company and sharing of many exciting moments. In particular, I would like to thank my Egyptian friends in Minnesota, Ahmed Medhat, Ibrahim Sabek, Hamada Aboubakr, Mohamed Hagar, Tamer Sharafeldin, Tamer Mansour, Eman Ramadan, Hebatallah El-Dakiki, Hesham Zakareya, Amr Magdy, Sara Morsy, and Ahmed Saber, whose company made my days more enjoyable. Special thanks to my mother; your love is always the essential strength for me to face up everything coming to me and rejoice upon everything rejoicing me. I sincerely express my deepest gratitude to the love of my life, Sara, whose unconditional love and unwavering support provides me with the strength to continue on my journey in life.

Finally, one person has been the most caring and encouraging, and has been my ultimate source of wisdom and support. I owe *my father* everything in my life.

Ahmed S. Zamzam

Minneapolis, June 20, 2019

Dedication

To my father and his beloved granddaughter, Hana.

Abstract

Distribution system state estimation (DSSE) is a core task for monitoring and control of distribution networks. Widely used Gauss-Newton approaches are not suitable for real-time estimation, often require many iterations to obtain reasonable results, and sometimes fail to converge. Learning-based approaches hold the promise for accurate real-time estimation. This dissertation presents the first data-driven approach to 'learn to initialize' – that is, map the available measurements to a point in the neighborhood of the true latent states (network voltages), which is used to initialize Gauss-Newton. In addition, a novel learning model is also presented that utilizes the electrical network structure. The proposed neural network architecture reduces the number of coefficients needed to parameterize the mapping from the measurements to the network state by exploiting the separability of the estimation problem. The proposed approach is the first that leverages electrical laws and grid topology to design the neural network for DSSE. It is shown that the proposed approaches yield superior performance in terms of stability, accuracy, and runtime, compared to conventional optimization-based solvers.

The second part of the dissertation focuses on the AC Optimal Power Flow (OPF) problem for multi-phase systems. Particular emphasis is given to systems with large-scale integration of renewables, where adjustments of real and reactive output power from renewable sources of energy are necessary in order to enforce voltage regulation. The AC OPF problem is known to be nonconvex (and, in fact, NP-hard). Convex relaxation techniques have been recently explored to solve the OPF task with reduced computational burden; however, sufficient conditions for tightness of these relaxations are only available for restricted classes of system topologies and problem setups. Identifying feasible power-flow solutions remains hard in more general problem formulations, especially in unbalanced multi-phase systems with renewables. To identify feasible and optimal AC OPF solutions in challenging scenarios where existing methods may fail, this dissertation leverages the Feasible Point Pursuit - Successive Convex Approximation algorithm a powerful approach for general nonconvex quadratically constrained quadratic programs. The merits of the approach are illustrated using several multi-phase distribution networks with renewables.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The North American power grid has been hailed as the most important engineering achievement of the 20^{th} century [108]. However, over the course of the past decades, it has become evident that significant aspects of the operational capabilities of the power grid, encompassing power generation, transmission, distribution and consumption, will have to be overhauled in order to meet new challenges imposed by the 21^{st} century demands. The smart grid paradigm leverages recent breakthroughs in sensing and signal processing, machine learning, computational optimization, and control of dynamical systems to enable an unprecedented leap in our ability to monitor, control, optimize, and learn from power grid operations.

The sheer size and complexity of such an interconnected network poses multiple challenges in realizing the smart grid vision: i) low-latency communication networks enable meters installed across the grid to continuously generate and transmit measurements at voluminous aggregate rates, demanding accurate, real-time processing in order to facilitate real-time network management; ii) centralized approaches to system operations are increasingly rendered impractical by the scale of interconnections, complexity issues beyond the capabilities of a single control center, confidentiality concerns amongst regional operators and susceptibility to failure/cyberattacks – thus underscoring the need for developing decentralized approaches; and iii) grid operators need to account for the increasingly higher penetration of renewable energy sources (RESs), the ever improving cost effectiveness of installing energy storage (ES) units, and the changing electricity demand patterns.

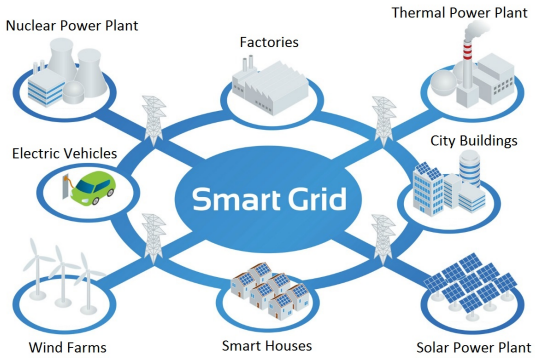


Figure 1.1: Main components of smart grid.[1]

Typically, a *smart grid* consists of electric power generators, smart meters, and various controllable energy devices; see Fig. 1.1 for an illustration. Smart meters are currently being installed on buildings that enable two-way communication between the utility and end customers. Owing to the complexity of the smart grid, heterogeneity of its components, and the high volume of meter readings to be processed, machine learning and mathematical optimization techniques are well motivated for enhancing the operation and performance of the electric distribution networks.

1.1 Context and Motivation

Due to the nonlinear nature of the equations governing the power flow in electrical grids, numerous approximation approaches have been proposed to tackle the associated monitoring and control tasks. Commonly used solvers include the Gauss-Seidel and Newton-Raphson iterative algorithms [3, 31, 67, 74, 75]. These iterative approaches seek improved approximation to the solutions of the nonlinear optimization problems, featuring quadratic convergence whenever the initial point lands within a small neighborhood of the optimal solution. As convergence of these algorithms hinges on the initial point, they may diverge if the initialization is not reliable.

Recently, convex relaxation approaches have attracted great attention [5, 23, 47, 51, 56, 71, 72, 121, 123] due to their ability to identify globally optimal solutions under certain conditions. Among those, semidefinite relaxation (SDR) was shown to be able to find

[1] Image courtesy of `SmartCitiesWorld.net` [89].

the global optimal solution of many control problems. SDR relies on matrix-lifting and rank relaxation to convexify the feasible set of the problem [5, 23, 56, 63, 65, 117]; the resulting relaxed problem can be solved in polynomial time. Optimality of the SDR solution can always be tested *a posteriori* by checking the rank of the SDR solution matrix; but it is very useful to know *a priori* in which cases SDR will yield an optimal solution for the original nonconvex problem. These are the cases when SDR yields a solution that is rank-one, or can be easily transformed to rank-one. In those cases, SDR is not a relaxation after all; we say that SDR is *tight*. Tightness of SDR relaxation was proved for a number of network setups under restrictive conditions. However, in most practical scenarios, the relaxation is not tight, and hence, does not provide optimal (and sometimes not even feasible) solutions. The other main drawback is complexity is still $\mathcal{O}(n^{6.5})$ due to variable lifting from vector to matrix.

The key novelty of this thesis is to pioneer machine learning and nonconvex optimization approaches tailored for large-scale power system monitoring and control tasks. In this context, a critical observation is that the nonconvex approaches outperform their convex counterparts when initialized judiciously. In addition, the robustness of control strategies relies mainly on precise characterization of the uncertainty in the forecasted generation and consumption quantities. Machine learning approaches can cope with uncertainty, and are therefore expected to play an important role in enable robust power system operations and control. The central goal is to put forth algorithmic foundations and performance analyses for optimally handling energy systems monitoring and control tasks.

This thesis establishes novel approaches for distribution systems state estimation and resource management benefiting from recent advances in mathematical optimization and machine learning research.

1.1.1 Distribution System State Estimation

Distribution system state estimation (DSSE) is an important task for monitoring and control of distribution networks. DSSE takes as input a set of measurements of physical quantities in the network and provides an estimate of the system state, i.e., nodal voltages. Due to the rapid introduction of volatile renewable energy sources and controllable

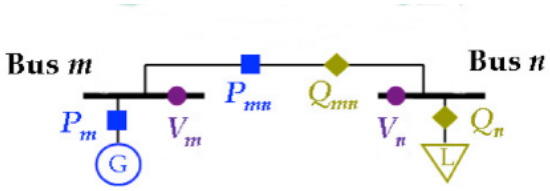


Figure 1.2: SCADA measurement system.

loads, modern distribution grids are challenged by unusual fluctuations in the operational conditions. Therefore, accurate real-time monitoring of distribution networks becomes increasingly pivotal in order to ensure reliable and optimal operation of the grid. This is possible, under certain conditions, because the different quantities in the network are related by physical laws (e.g., Kirchoff's voltage, current, power conservation laws). Supervisory control and data acquisition (SCADA) systems monitor and control plants (such as a power plant) and other complex systems that are spread out over large geographic areas. SCADA systems periodically collect several types of measurements at some buses and lines of the grid. Fig. 1.2 illustrates various types of measurements that can be acquired by the SCADA systems. These measurements include the electrical power flow and injections and voltage and current magnitudes. The need for adaptive monitoring is further amplified in future distribution networks, where volatile smart grid components such as renewable energy sources, demand-response scheduling, and smart meter infrastructure challenge situational awareness and control. The nonlinear relationship between the measured quantities and the state variables complicates any DSSE approach. In addition, owing to higher penetration of renewable energy sources, and the introduction of demand-response and other efficiency- and agility-improving measures, the grid will become more responsive and hence more volatile – necessitating equally swift and robust DSSE routines.

Traditional SE solvers are based on the weighted least-squares (WLS) criterion, and the Gauss-Newton algorithm which is used as the workhorse for solving the associated nonconvex problem. Largely depending on the initialization, this gradient-type iterative solver often converges to uninteresting points, or even diverges. This is particularly problematic when system states vary significantly between measurements, or, when bad data are present. On the other hand, when the initialization is relatively close to the optimal solution, the method enjoys quadratic convergence.

Many alternative approaches have been proposed for solving the SE task by exploiting hidden convexity in certain cases, which motivates relaxing the more general problem to a convex one [5, 23, 40, 123]. However, convex relaxations for PSSE do not perform as desired in challenging scenarios, in particular when the measured data size is relatively small, or when the available data do not include the voltage magnitude at all buses [122]. The latter is often the case in practice.

Building upon the recently proposed feasible point pursuit (FPP) algorithm for tackling nonconvex QCQPs [68] and its application to the optimal power flow problem [111], we have recently proposed an FPP-based framework for solving the PSSE problem. The key point is to put the nonlinear WLS formulation into a QCQP form. Then, the nonconvex QCQP is tackled using the FPP approach. The method was shown to be very effective in terms of finding better estimates for the voltage profile than the relaxation methods and the plain-vanilla Gauss-Newton method. However, the FPP approach solves a second order cone program in each iteration, which is a heavy computational burden for highly dynamic systems.

1.1.2 Optimal Power Flow for Multi-Phase Distribution Networks

The OPF problem aims at minimizing an appropriate operational cost while respecting the network's physical and engineering constraints. The AC-OPF problem is a predominant task in optimizing the performance of energy grids. However, due to the quadratic nature of the power flow equations, the OPF problem is known to be nonconvex and NP-hard in general [56, 58].

Several approaches have been proposed to tackle its non-convex nature, such as DC-OPF [2, 19, 92, 93], which is a linear program obtained via a linearization of the power flow equations. However, the solution of DC-OPF is not guaranteed to be even feasible for the original problem. For the actual non-linear problem, (i.e., AC-OPF), a slew of methods ranging from classical Newton-Raphson (NR), to Lagrangian relaxation, genetic algorithms and interior point methods, have been proposed. Unfortunately, these methods do not provide optimality or feasibility guarantees except in certain cases, and are quite sensitive to the initial guess. In a series of pioneering papers, necessary and sufficient conditions were provided for guaranteeing the tightness of SDR in solving AC-OPF to globally optimality [5, 56], which is satisfied by certain networks

under restrictive assumptions [56, 57, 90, 117]. The inexactness of SDR for a general network was demonstrated in [59] using a simple 3-bus network. In cases where SDR is not tight, it is not easy to recover a physically meaningful solution from the solution matrix; only a lower bound on the optimal cost is provided. An approach to recover an OPF solution from the SDR solution was pursued in [63], but it lacks guarantees. Moment-based relaxations which generalize SDR using the Lassarre Hierarchy have also been proposed for OPF [70, 72, 73] in cases where standard SDR fails. The downside of this approach is the high complexity incurred in solving a sequence of very large-scale semidefinite programming (SDP) problems. Decentralized OPF approaches for balanced transmission networks based on augmented Lagrangian schemes have been proposed in [6, 44, 50, 77], while distributed SDR-based approaches for balanced and unbalanced distribution networks have been adopted in [55] and [23] respectively. To summarize, NR methods do not exhibit reliable performance for AC-OPF in general and are lacking even basic convergence guarantees. Meanwhile, the potential of SDR in attaining globally optimal solutions appears limited.

1.2 Thesis Outline and Contributions

In Chapter 2, the distribution system state estimation problem is introduced and its mathematical formulation is presented. First, the electrical distribution network representation is put forth. Then, the WLS formulation of the problem is introduced. In addition, the chapter presents the types and synthesizing functions of measurements usually available in distribution systems which can be utilized while solving the DSSE problem. The mathematical definitions and formulations introduced in Chapter 2 are essential for the DSSE solvers presented in Chapter 3 and Chapter 4.

Chapter 3 presents a data-driven machine learning-assisted optimization approach for the DSSE problem. Since the Gauss-Newton solver is known to be very sensitive to initializations, it is therefore natural to ask if there is a smart way of initializing Gauss-Newton that will avoid such pitfalls. This chapter present a framework that utilizes historical or simulated data in order to *learn to initialize* the Gauss-Newton solver. As the requirement for the machine learning model is relaxed by mapping the measurement to a state just in the neighborhood of the underlying system state, the

complexity requirements of the neural network are reduced accordingly. We analyze the required complexity of a shallow neural network that is required to approximate the SE mapping with certain accuracy. Also, the massive reduction in the neural network complexity reduces the required number of training samples.

Key to our approach is a judicious design of the cost function used to train the shallow neural network which matches the task of the machine learning model. The performance of the developed machine learning-assisted optimization approach is assessed using the IEEE-37 distribution feeder. We show that the proposed training approach helps by providing better initial estimates to the Gauss-Newton solver. In addition, the presented framework provides superior performance results in terms of estimation accuracy, computational complexity, and stability of the Gauss-Newton solver. The results of this chapter have been presented in [114].

In Chapter 4, a physics-aware neural network model is introduced to tackle the DSSE problem. It is shown that the huge discrepancy in the accuracy of different measuring technologies in distribution networks enables effective separability in the DSSE problem. A novel graph-based neural network is introduced which utilizes the knowledge of the distribution system. Therefore, the proposed neural network is termed '*Physics-Aware Neural Network*' (PAWNN). This design reduces the number of trainable coefficients of the mapping between the measurements to the state of the system significantly. In addition to preventing overfitting, this physics-aware design reduces the training time and size of training data needed.

Additionally, a novel method for placing μPMUs in distribution grids is presented in Chapter 4. This algorithm builds on the separability of the SE mapping that results from installing μPMUs. Therefore, the complexity of the resulting machine learning depends on the installation pattern of μPMUs. The numerical results suggest that the greedy algorithm achieves near optimal solution at a linear cost. Superior performance in terms of accuracy and running time of state estimation using the PAWNN is also presented. The material of Chapter 4 has been reported in [115].

Chapter 5 focuses on the AC OPF problem for multi-phase systems featuring renewable energy sources. First, the power flow model is presented for general multi-phase systems. Then, an extension that accounts for delta-connected loads is presented. The AC OPF problem is formulated in both settings where the output of the renewable

energy sources is controllable. Similar to various AC OPF renditions, the resultant problem is a nonconvex quadratically-constrained quadratic program. However, the so-called Feasible Point Pursuit - Successive Convex Approximation (FPP-SCA) algorithm, pioneered in our group, is leveraged to obtain a feasible and locally-optimal solution.

The main idea behind FPP-SCA is to allow a controllable amount of constraint violations to enable the algorithm to make progress toward feasibility in its initial stages. The overall approach is neither restriction nor relaxation, but rather judicious *approximation* of the OPF problem in each iteration, the solution of which is subsequently used as the approximation point for the next iteration. Upon finding a feasible voltage profile, successive convex approximation of the feasible set is used to find a KKT point of the OPF problem. In addition, a modified version of the FPP-SCA approach is shown to be able to identify problematic constraints in infeasible AC OPF instances. Simulations of the approach on several distribution and transmission networks show the efficacy of the proposed approach in identifying feasible optimal solutions. The material of Chapter 5 has been reported in [111, 112].

1.3 Notational Conventions

The following notational conventions will be adopted throughout the subsequent chapters. Upper (lower) boldface letters will be used for matrices (column vectors); $(\cdot)^T$ denotes transposition; $(\cdot)^H$ denotes complex-conjugate transposition; $\mathrm{Tr}(\cdot)$ denotes the matrix trace; $(\cdot)^\dagger$ denotes the matrix MoorePenrose inverse; $\mathrm{vec}(\cdot)$ denote the column-wise matrix vectorization; $\mathbf{1}_{m\times n}$ the $m \times n$ matrix of all ones; $\mathbf{0}_{m\times n}$ the $m \times n$ matrix of all zeros; $\mathbf{I}_n$ the $n \times n$ identity matrix; $\|\|_F$ the matrix Frobenius norm; $\mathrm{Re}\{\cdot\}$ the real part; and $\mathrm{Im}\{\cdot\}$ the imaginary part.

Chapter 2

Distribution System State Estimation Problem

State estimation (SE) techniques are used to monitor power grid operations in real-time. Accurately monitoring the network operating point is critical for many control and automation tasks, such as Volt/VAr optimization, feeder reconfiguration and restoration. SE uses measured quantities like nodal voltages, injections, and line flows, together with physical laws in order to obtain an estimate of the system state variables, i.e., bus voltage magnitudes and angles [49, 102] across the network. SE techniques have also proven to be useful in network 'forensics', such as spotting bad measurements and identifying gross modelling errors [61].

Unlike transmission networks where measurement units are placed at almost all network nodes, the SE task in distribution systems is particularly challenging due to the scarcity of real-time measurements. This is usually compensated by the use of so-called *pseudo-measurements*. Obtained through short-term load and renewable energy forecasting techniques, these pseudo-measurements play a vital rule in enabling distribution system state estimation (DSSE) [30, 39, 66]. Several DSSE solvers based on weighted least squares (WLS) transmission system state estimation methods have been proposed [8,48,60,88,99]. A three-phase nodal voltage formulation was used to develop a WLS-based DSSE solver in [8, 60]. Recently, the authors of [29] used Wirtinger calculus to devise a new approach for WLS state estimation in the complex domain. In order to

reduce the computational complexity and storage requirements, the branch-based WLS model was proposed in [9,103]. However, such gains can be only obtained when the target power system features only wye-connected loads that are solidly grounded. It is also recognized that incorporating phasor measurements in DSSE improves the observability and the estimation accuracy [80,125]. Therefore, the DSSE approaches developed in this thesis consider the case where classical (quadratic) and phasor (linear) measurements are available, as well as pseudo-measurements provided through short term forecasting algorithms.

WLS DSSE is a non-convex problem that may have multiple local minima, and working with a limited number of measurements empirically aggravates the situation, as it may introduce multiple local minima as well. Furthermore, Gauss-Newton type algorithms may need many iterations, or even fail to converge. In this chapter, we present the formulation of the distribution system state estimation (DSSE) problem. Also, we present the available types of measurements avaiable in distribution systems.

2.1 Network Representation

Consider a multi-phase distribution network consisting of $N+1$ nodes and L edges represented by a graph $\mathcal{G} := (\mathcal{N}, \mathcal{L})$, whose set of multi-phase nodes (buses) is indexed by $\mathcal{N} := \{0, 1, \ldots, N\}$, and $\mathcal{L} \subseteq \mathcal{N} \times \mathcal{N}$ represents the lines in the network. Let the node 0 be the substation that connects the system to the transmission grid. The set of phases at bus n and line (l, m) are denoted by $\boldsymbol{\varphi}_n$ and $\boldsymbol{\varphi}_{lm}$, respectively. Let the voltage at the n-th bus for phase ϕ be denoted by $v_{n,\phi}$. Then, define $\mathbf{v}_n := [v_{n,\phi}]_{\phi \in \boldsymbol{\varphi}_n}$ to collect the voltage phasors at the phases of bus n. In addition, let the vector $\mathbf{v}$ concatenate the vectors $\mathbf{v}_n$ for all $n \in \mathcal{N}$. Lines $(l, m) \in \mathcal{L}$ are modeled as π-equivalent circuit, where phase impedance and shunt admittance are denoted by $\mathbf{Z}_{lm} \in \mathbb{C}^{|\boldsymbol{\varphi}_{lm}| \times |\boldsymbol{\varphi}_{lm}|}$ and $\check{\mathbf{Y}}_{lm} \in \mathbb{C}^{|\boldsymbol{\varphi}_{lm}| \times |\boldsymbol{\varphi}_{lm}|}$, respectively.

2.2 Problem Formulation

The DSSE task amounts to recovering the voltage phasors of buses given measurements related to real-time physical quantities, and the available pseudo-measurements. Actual

measurements are acquired by smart meters, PMUs, and μPMUs that are placed at some locations in the distribution network. The measured quantities are usually noisy and adhere to

$$\tilde{z}_\ell = \tilde{h}_\ell(\mathbf{v}) + \xi_\ell, \qquad 1 \leq \ell \leq L_m \tag{2.1}$$

where ξ_ℓ accounts for the zero-mean measurement noise with known variance $\tilde{\sigma}_\ell^2$. The functions $\tilde{h}_\ell(\mathbf{v})$ are dependent on the type of the measurement, and can be either linear or quadratic relationships. In the next section, the specific form of $\tilde{h}_\ell(\mathbf{v})$ will be discussed. In addition, load and generation forecasting methods are employed to obtain pseudo-measurements that can help with identifying the network state. The forecasted quantities are modeled as

$$\check{z}_\ell = \check{h}_\ell(\mathbf{v}) + \zeta_\ell, \qquad 1 \leq \ell \leq L_s \tag{2.2}$$

where ζ_ℓ represents the zero-mean forecast error which has a variance of $\check{\sigma}_\ell^2$. Since $\check{z}_\ell$'s represent power-related quantities, they are usually modeled as quadratic functions of the state variable $\mathbf{v}$. While the value of the measurement noise variance $\tilde{\sigma}_\ell^2$ depends on the accuracy of the measuring equipment, the variance of the forecast error can be determined using historical forecast data.

Let $\mathbf{z}$ be a vector of length $L = L_m + L_s$ containing the measurements and pseudo-measurements, and $\mathbf{h}(\mathbf{v})$ the equation relating the measurements to the state vector $\mathbf{v}$, which will be specified in the next section. Adopting a weighted least-squares formulation, the problem can be cast as follows

$$\begin{aligned} \min_{\mathbf{v}} \; J(\mathbf{v}) &= \sum_{\ell=1}^{L_m} \tilde{w}_\ell \big(\tilde{z}_\ell - \tilde{h}_\ell(\mathbf{v})\big)^2 + \sum_{\ell=1}^{L_s} \check{w}_\ell \big(\check{z}_\ell - \check{h}_\ell(\mathbf{v})\big)^2 \\ &= \; (\mathbf{z} - \mathbf{h}(\mathbf{v}))^T \mathbf{W} (\mathbf{z} - \mathbf{h}(\mathbf{v})) \end{aligned} \tag{2.3}$$

where the values of $\tilde{w}_\ell$ and $\check{w}_\ell$ are inversely proportional to σ_ℓ^2 and $\check{\sigma}_\ell^2$, respectively. The optimization problem (2.3) is non-convex due to the nonlinearity of the measurement mappings $\mathbf{h}(\mathbf{v})$ inside the squares.

2.3 Available Measurements for DSSE

As indicated in the previous subsection, only few real-time measurements are usually available in distribution networks, relative to the obtainable measurements in transmission systems. Therefore, pseudo-measurements are used to alleviate the issue of solving an underdetermined problem. There are always different latencies for different sources of measurements which bring up the issue of time skewness. Many approaches have been proposed in the literature to tackle the issue [120]. In this work, assume that the issue is resolved using one of the solutions in the literature, and hence, the measurements are assumed to be synchronized. First, the measurements function $\tilde{h}(\mathbf{v})$ will be introduced for all types of available measurements. Then, the construction of the pseudo-measurements mappings $\check{h}(\mathbf{v})$ will be explained.

The measurement functions consist of:

• *phasor measurements* which represent the complex nodal voltages $\mathbf{v}_n$, or current flows $\mathbf{i}_{lm}$. The corresponding measurement function is linear in the state variable $\mathbf{v}$. These measurements are usually obtained by the PMUs and μPMUs. Each measurement of this type is handled as two measurements, i.e., the real and imaginary parts of the complex quantities are handled as two measurements. For the nodal voltages, the real and imaginary parts are given as follows

$$\Re\{v_{n,\phi}\} = \frac{1}{2}\, \mathbf{e}_{n,\phi}^T\, (\mathbf{v}_n + \overline{\mathbf{v}}_n), \tag{2.4}$$

$$\Im\{v_{n,\phi}\} = \frac{1}{2j}\, \mathbf{e}_{n,\phi}^T\, (\mathbf{v}_n - \overline{\mathbf{v}}_n) \tag{2.5}$$

where $\mathbf{e}_\phi$ is the ϕ-th canonical basis in $\mathbb{R}^{|\boldsymbol{\varphi}_n|}$. In addition, the current flow measurements can be modeled as

$$\Re\{i_{lm,\phi}\} = \frac{1}{2}\, \mathbf{e}_{lm,\phi}^T\, \big(\mathbf{Y}_{lm}(\mathbf{v}_l - \mathbf{v}_m) + \overline{\mathbf{Y}}_{lm}(\overline{\mathbf{v}}_l - \overline{\mathbf{v}}_m)\big) \tag{2.6}$$

$$\Im\{i_{lm,\phi}\} = \frac{1}{2j}\, \mathbf{e}_{lm,\phi}^T\, \big(\mathbf{Y}_{lm}(\mathbf{v}_l - \mathbf{v}_m) - \overline{\mathbf{Y}}_{lm}(\overline{\mathbf{v}}_l - \overline{\mathbf{v}}_m)\big) \tag{2.7}$$

where $\mathbf{Y}_{lm}$ is the inverse of $\mathbf{Z}_{lm}$, and $\mathbf{e}_{lm,\phi}$ is the ϕ-th canonical basis in $\mathbb{R}^{|\boldsymbol{\varphi}_{lm}|}$.

• *real-valued measurements* which encompass voltage magnitudes $|v_{n,\phi}|$, current magnitudes $|i_{lm,\phi}|$, and real and reactive power flow measurements $p_{lm,\phi}, q_{lm,\phi}$. These measurements are obtained by SCADA systems, Distribution Automation, Intelligent Electronic Devices, and PMUs. The real-valued measurements are nonlinearly related to the state variable $\mathbf{v}$. The measured voltage magnitude square, and active and reactive power flows can be represented as quadratic functions of the state variable $\mathbf{v}$, see [23]. The current magnitude squared can be written as follows

$$|i_{lm,\phi}|^2 = (\mathbf{v}_l - \mathbf{v}_m)^H \mathbf{y}_{lm,\phi}^H \mathbf{y}_{lm,\phi} (\mathbf{v}_l - \mathbf{v}_m) \tag{2.8}$$

where $\mathbf{y}_{lm,\phi}$ is the ϕ-th row of the admittance matrix $\mathbf{Y}_{lm}$. Therefore, all the real-valued measurements can be written as quadratic measurements of the state variable $\mathbf{v}$.

The available real-time measurements are usually insufficient to 'pin down' the network state, as we have discussed. In this case, the system is said to be unobservable. Hence, pseudo-measurements that augment the real-time measurements are crucial in DSSE as they help achieve network observability. Pseudo-measurements are obtained through load and generation forecast procedures that aim at estimating the energy consumption or generation utilizing historical data and location-based information. They are considered less accurate than real-time measurements, and hence, assigned low weights in the WLS formulation. The functions governing the mapping from the state variable to the forecasted load and renewable energy source injections can be formulated as quadratic functions [23, 111].

Therefore, any measurement synthesizing function $h_\ell(\mathbf{v})$ can be written in the following form

$$h_\ell(\mathbf{v}) = \overline{\mathbf{v}}^T \mathbf{D}_\ell \mathbf{v} + \mathbf{c}_\ell^T \mathbf{v} + \overline{\mathbf{c}}_\ell{}^T \overline{\mathbf{v}} \tag{2.9}$$

where $\mathbf{D}_\ell$ is a Hermitian matrix, which is usually of rank-1 [100]. This renders $J(\mathbf{v})$ a fourth order function of the state variable, which is very challenging to optimize.

The Gauss-Newton algorithm linearizes the first order optimality conditions to iteratively update the state variables until convergence. The algorithm is known to perform well in practice given that the algorithm is initialized from a point in the vicinity of the true network state, albeit lacking provable convergence result in theory. Several variants of the algorithm have been proposed in the literature using polar [8], rectangular [78]

and complex [29] representations of the state variables. All these algorithms work to a certain extent, but failure cases are also often observed. Again, stable convergence performance is only observed when the initialization is close enough to the optimal solution of (2.3). This is not entirely surprising—given the non-convex nature of the DSSE problem.

Chapter 3

Machine Learning-Assisted Optimization Approach for DSSE

The weighted least-squares formulation of DSSE (2.3) is highly nonconvex. Therefore, any local algorithm can converge to one of many local minima. Gauss-Newton type algorithms behave very differently when using different initializations—the algorithms may need many iterations, or even fail to converge. It is therefore natural to ask if there is a smart way of initializing Gauss-Newton that will avoid these pitfalls?

In this chapter, we propose a novel learning architecture for the DSSE task. Our idea is as follows. A wealth of historical data is often available for a given distribution system. This data is usually stored and utilized in various other network management tasks, such as load and injection forecasting. Even without detailed recording of the network state, we can reuse this data to simulate network operations off-line. We can then think of network states and measurements as (output,input) training pairs, which can be used to train a neural network (NN) to 'learn' a function that maps measurements to states. After the mapping function is learned, estimating the states associated with a fresh set of measurements only requires very simple operations—passing the measuremnts through the learned NN. This would greatly improve the efficiency of DSSE, bringing real-time state estimation within reach. Accurate and cheap DSSE using an NN may sound too good to be true, and in some sense (in its raw form) it is; but there is also silver lining, as we will see.

Known as universal function approximators, neural networks have made a remarkable comeback in recent years, outperforming far more complicated (and disciplined) methods in several research fields; see [42, 45, 54, 97] for examples. One nice feature of neural networks and other machine learning approaches is that they alleviate the computational burden at the operation stage—by shifting computationally intensive 'hard work' to the off-line training stage.

However, accurately learning the end-to-end mapping from the available measurements to the exact network state is very challenging in our context—the accuracy achieved by convergent Gauss-Newton iterates (under good initializations) is hard to obtain using learning approaches. The mapping itself is very complex, necessitating a wide and/or deep NN that is hard to train with reasonable amounts of data. In addition, training a deep NN (DNN) is computationally cumbersome requiring significant computing resources. Also, DNN slows down real-time estimation, as passing the input through its layers is a sequential process that cannot be parallelized. To circumvent these obstacles, we instead propose to train a *shallow* neural network to 'learn to initialize'—that is, map the available measurements to a point in the neighborhood of the true latent states, which is then used to initialize Gauss-Newton; see Fig. 3.1 for an illustration. When the Gauss-Newton solver is initialized at a point in the vicinity of the optimal solution, it enjoys quadratic convergence [34]; otherwise, divergence is possible. We show that such a hybrid machine learning / optimization approach yields superior performance compared to conventional optimization-only approaches, in terms of stability, accuracy, and runtime efficiency. We demonstrate these benefits using convincing experiments with the benchmark IEEE-37 distribution feeder with several renewable energy sources installed and several types of phasor and conventional measurements, as well as pseudo-measurements. The key to success is *appropriate design of the NN training cost function* for the 'neighborhood-finding' NN. As we will see, the proposed cost function serves our purpose much better than using a generic cost function for conventional NN training. In addition, owing to the special design of the training cost function, the experiments corroborate the resiliency of the proposed approach in case of modest network reconfiguration events.

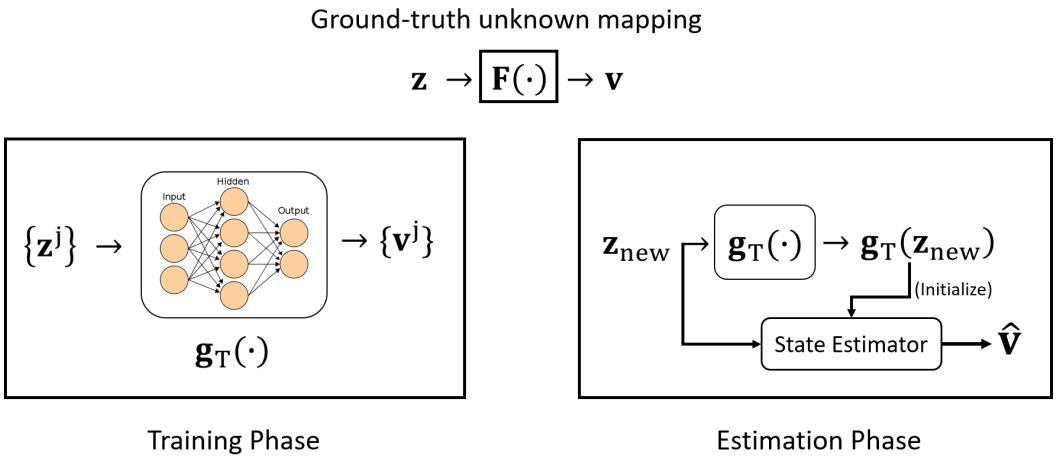


Figure 3.1: The proposed learning-based DSSE

3.1 Machine Learning Approaches for DSSE

Machine learning approaches are not entirely new in the power systems / smart grid area. For instance, an online learning algorithm was used in [79] to shape residential energy demand and reduce operational costs. In [32], a multi-armed bandit online learning technique was employed to forecast the power injection of renewable energy sources. An early example of using NNs in estimation problems appeared in [4] as part of damage-adaptive intelligent flight control. Closer to our present context, [66] proposed the use of an artificial neural network that takes the measured power flows as input and aims to estimate the bus injections, which are later used as pseudo-measurements in the state estimation. In contrast to our approach where the NN is used to approximate the network state given the conventional measurements as well as the pseudo-measurements, the authors of [66] designed an artificial NN to generate pseudo-measurements from the available power flow measurements. In addition, artificial neural networks have been used for the prediction step in dynamic state estimation [84], and for forecasting-aided state estimation [21, 35] where the state of the network is estimated from the previous sequence of states, without using conventional measurements to anchor the solution. This is better suited for transmissions systems, which are more predictable relative to distribution systems with time-varying loads. As the installation of renewable energy sources in the distribution grid surges, the huge volatility brought by these

energy sources [87] induces rapid changes in network state, and thus the previously estimated state is often bad initialization for the next instance of DSSE. To the best of our knowledge, machine learning approaches have not yet been applied to the core DSSE optimization task, which is the focus of our work.

3.2 Proposed Approach: Learning-aided DSSE Optimization

Assume that there exists a mapping $\mathbf{F}(\cdot)$ such that

$$\mathbf{F}(\mathbf{z}) = \mathbf{v};$$

i.e., $\mathbf{F}(\cdot)$ maps the (noiseless) measurements to the ground-truth states. An example of such mapping is an optimization algorithm that can optimally solve the DSSE problem in the noiseless case, assuming that the solution is unique. The algorithm takes $\mathbf{z}$ as input and outputs $\mathbf{v}$. In reality the actual (and the virtual) measurements will be noisy, so we can only aim for

$$\mathbf{F}(\mathbf{z}) \approx \mathbf{v};$$

which is also what optimization-based DSSE aims for in the noisy case.

Inspired by the recent successes of machine learning, it is intriguing to ask whether it is possible to learn mapping $\mathbf{F}(\cdot)$ from historical data. If the answer is affirmative and the learned $\hat{\mathbf{F}}(\cdot)$ is easy to evaluate, then the DSSE problem could be solvable in a very efficient way *online*, after the mapping $\hat{\mathbf{F}}(\cdot)$ is learned *offline*.

In machine learning, neural networks are known as universal function approximators. In principle, a three-layer (input, hidden, output) NN can approximate any continuous multivariate function down to prescribed accuracy, if there are no constraints on the number of neurons [20]. This motivates us to consider employing an NN for approximating $\mathbf{F}(\mathbf{z})$ in the DSSE problem. An NN with vector input $\mathbf{z}$, vector output $\mathbf{g}$, and one hidden layer comprising T neurons synthesizes a function of the folowing form

$$\mathbf{g}_T(\mathbf{z}) = \sum_{t=1}^{T} \boldsymbol{\alpha}_t \sigma(\mathbf{w}_t^T \mathbf{z} + \beta_t), \tag{3.1}$$

where $\mathbf{w}_t$ represents the linear combination of the inputs in $\mathbf{z}$ that is fed to the tth neuron (i.e., the unit represented by $\sigma(\mathbf{w}_t^T\mathbf{z}+\beta_t)$), β_t the corresponding scalar bias, and the vectors $\boldsymbol{\alpha}_t$'s combine the outputs of the neurons in the hidden layer to produce the vector output of the NN. The parameters $(\boldsymbol{\alpha}_t, \mathbf{w}_t, \beta_t)_{t=1}^T$ can be learned by minimizing the training cost function

$$\min_{\{\boldsymbol{\alpha}_t, \mathbf{w}_t, \beta_t\}_{t=1}^T} \quad \sum_j \|\mathbf{v}^j - \mathbf{g}_T(\mathbf{z}^j)\|_2^2, \tag{3.2}$$

where the pair $(\mathbf{z}^j, \mathbf{v}^j)$ is a training sample of measurements and the corresponding underlying voltages to be estimated, in our context.

The above training cost function ideally seeks an NN that works perfectly—at least over the training set. This approach is similar in spirit to the one in [96], which considered a problem in wireless resource allocation with the objective of 'learning to optimize'—meaning, training an NN to learn the exact end-to-end input-output mapping of an optimization algorithm. Our experience has been that, for DSSE, such an approach works to some extent, but its performance is not ideal. Trying to learn the end-to-end DSSE mapping appears to be too ambitious, requiring very large T or a deep NN, and very high training sample complexity. To circumvent these obstacles, we instead propose to train a *shallow* neural network, as above, to 'learn to initialize'—that is, map the available measurements to a point in the neighborhood of the true latent state, which is then used to initialize Gauss-Newton as depicted in Fig. 3.2.

More specifically, we propose using the following cost function for training the NN:

$$\min_{\{\mathbf{w}_t, \beta_t, \boldsymbol{\alpha}_t\}_{t=1}^T} \sum_j \max\{\|\mathbf{v}^j - \mathbf{g}_T(\mathbf{z}^j)\|_2^2 - \epsilon^2,\ 0\} \tag{3.3}$$

where the cost function indicates that the NN parameters are tuned with the relaxed goal that $\mathbf{g}_T(\mathbf{z}^j)$ lies in the ball of radius ϵ around $\mathbf{v}^j$. Fig. 3.3 illustrates the effect of changing the value of ϵ on the empirical loss function. The high-level idea is as follows: instead of enforcing minimization of $\sum_j \|\mathbf{v}^j - \mathbf{g}_T(\mathbf{z}^j)\|_2^2$, we seek a 'lazy' solution such that $\|\mathbf{v}^j - \mathbf{g}_T(\mathbf{z}^j)\|_2^2 \leq \epsilon$ for as many j as possible—in other words, it is enough to get to the right neighborhood. As we will show, this 'lowering of the bar' can significantly reduce the complexity of the NN (measured by the number of neurons) that is needed

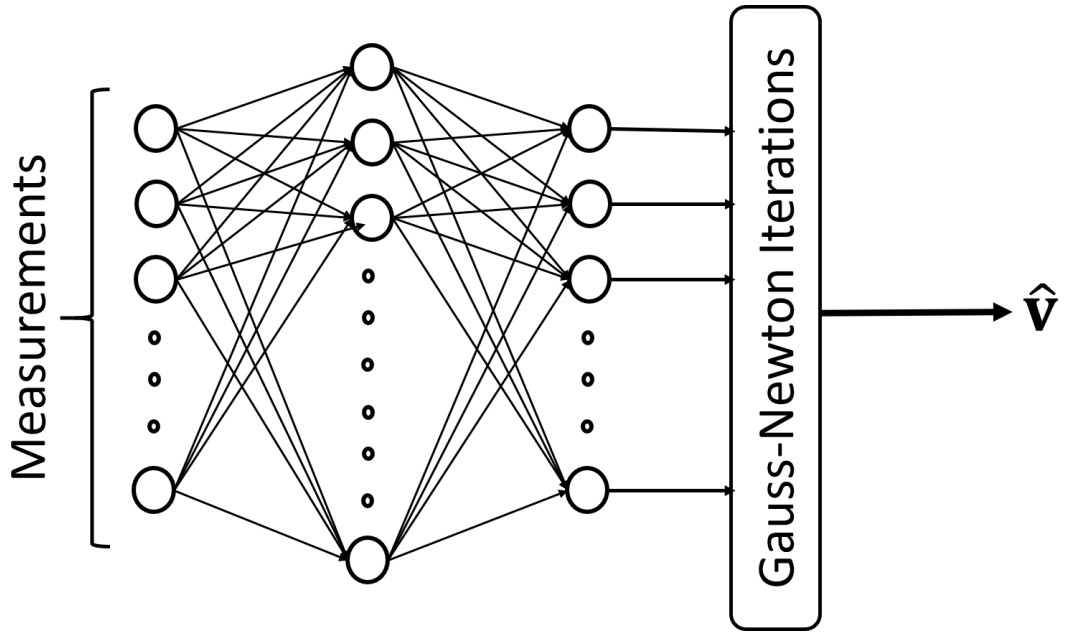


Figure 3.2: Learning-based state estimator structure

to learn such an approximate mapping, and with it also the number of training samples required for learning. These complexity benefits are obtained while still reproducing a point that is close enough to serve as a good initialization for Gauss-Newton, ensuring stable and rapid convergence. To back up this intuition, we have the following result.

Proposition 1. *Let $\sigma(\cdot)$ be any continuous sigmoidal function, and let $\mathbf{g}_T(\mathbf{z}) : \mathbb{R}^L \to \mathbb{R}^K$ be in the form*

$$\mathbf{g}_T(\mathbf{z}) = \sum_{t=1}^{T} \boldsymbol{\alpha}_t \sigma(\mathbf{w}_t^T \mathbf{z} + \beta_t).$$

Then, for approximating a continuous mapping $\mathbf{F} : \mathbb{R}^L \to \mathbb{R}^K$, the complexity for a shallow network to solve Problem (3.3) exactly (i.e., with zero cost) for a finite number of bounded training samples $\left(\mathbf{z}^j, \mathbf{v}^j = \mathbf{F}(\mathbf{z}^j)\right)$ is at least in the order of

$$T = \mathcal{O}\left(\left(\frac{\epsilon}{\sqrt{K}}\right)^{-\frac{L}{r}}\right).$$

where r is the number of continuous derivatives of $\mathbf{F}(\cdot)$.

The proof of this proposition is relegated to Appendix A. Note that the boundedness

assumption on the inputs is a proper assumption since these quantities represent voltages and powers. The implication here is very interesting, as controlling ϵ can drastically reduce the required T (and, along with it, sample complexity) while still ensuring an accurate enough prediction to enable rapid convergence of the ensuing Gauss-Newton stage. Furthermore, keeping the network shallow and T moderate makes the actual online computation (passing the input measurements through the NN to obtain the sought initialization) simple enough for real-time operation. This way, the relative strengths of learning-based and optimization-based methods can be effectively combined, and the difficulties of both methods can be circumvented.

One important remark is that Proposition 1 is derived under the assumption that $\mathbf{F}(\cdot)$ is a continuous mapping that can be parametrized with L parameters, which is hard to verify in our case. Nevertheless, we find that the theoretical result here is interesting enough and intuitively pleasing. In a case of a simple single-phase feeder, the state estimation mapping is indeed continuous and finitely parametrizable; see Appendix B. More importantly, as will be seen, this corroborating theory is consistent with our empirical results.

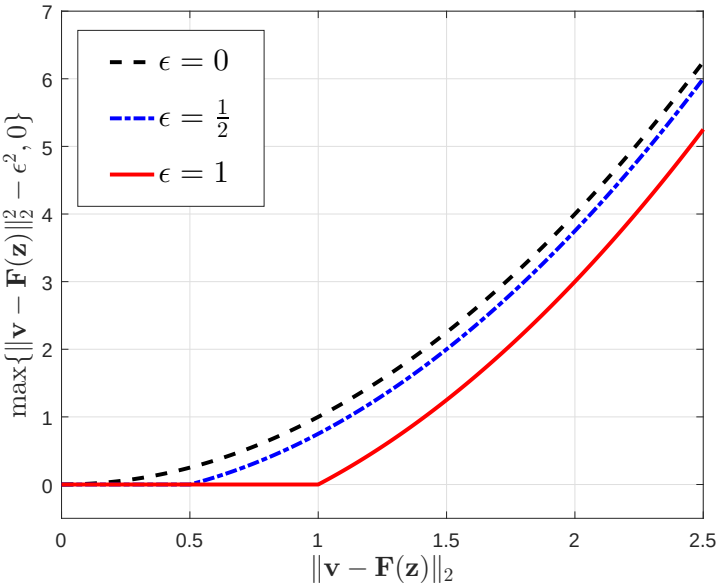


Figure 3.3: The empirical loss function used for training

In order to tune the neural network parameters, N_t training samples have to be used in order to minimize the cost function in (3.3). Two different ways can be utilized in order to obtain such training data. First, historical data for load and generation can be utilized. Note that these data are not readily available unless all the buses in the network are equipped with measuring devices, however, such load and generation profiles can be estimated using a state estimation algorithm. Then, the network power flow equations can be solved to obtain the system state which is used later to synthesize the measurements using (2.1) and (2.2). Hence, for each historical load and generation instance, a noiseless training pair $(\mathbf{z}^j, \mathbf{v}^j)$ can be generated. The second way to obtain the training data is to resort to an operating state estimation procedure. In this case, the goal of the neural network approach is to emulate the mapping of the estimator from the measurements space to the state space. The second approach suffers all the limitations of the current state estimation algorithms such as inaccuracy or computational inefficiency. In addition to providing noisy training pairs, these limitations result in a much more time consuming way of generating training data. Therefore, the first way is adopted for the rest of this work, and the detailed procedure is presented in the experiments section.

Remark 1. *One concern for data-driven approaches is that the mapping is learned from historical data under a certain network topology. What if the configuration changes for some reason (e.g., maintenance)? Is the trained mapping still useful? The answer is, surprisingly, affirmative, thanks to the 'lazy' training objective—since we do not seek exact solutions for the mapping, modest reconfiguration of the network will not destroy the effectiveness of the trained NN for initialization, as we will see in Section V.*

3.3 Experimental Results

The proposed state estimation procedure is tested on the benchmark IEEE-37 distribution feeder. This network is recommended for testing state estimation algorithms by the Test Feeder Working Group of the Distribution System Analysis Subcommittee of the IEEE PES [86]. The feeder is known to be a highly unbalanced system that has several delta-connected loads, which are blue-colored in Fig. 3.4.

The feeder has nodes that feature different types of connections, i.e., single-, two-,

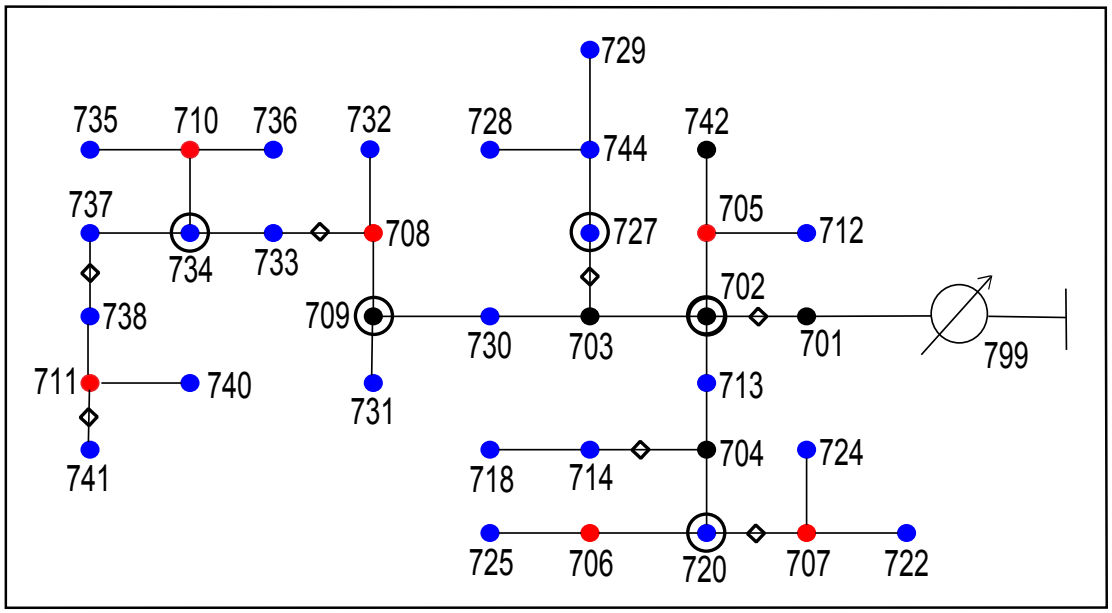


Figure 3.4: IEEE-37 distribution feeder. Nodes in blue are with loads, and red nodes represent buses with DER installed. Buses with PMUs are circled, and the links where the current magnitudes are measured have a small rhombus on them.

and three-phase connections. Additionally, distributed energy resources are assumed to be installed at six different buses, which are colored in red in Fig. 3.4. In Table 3.1, the types of the connections of all the loads and DERs are presented where (L) and (G) mean load and DER, respectively.

Historical load and generation data available in [7] modulated by the values of the loads are used to generate the training samples. Each time instance has an injection profile which is used as an input to the linearized power flow solver in [38]. The algorithm returns a voltage profile (network state variable) which is utilized to generate the value of the measurements at this point of time. A total of $100,000$ loading and generation scenarios were used to train a shallow neural network. The network has an input size of 103, 2048 nodes in the hidden layer, and output of size 210.

The available measurements are detailed as follows.

- *PMU measurements*: four PMUs are installed at buses 701, 704, 709, and 734 which are circled in Fig. 3.4. It assumed that the voltage phasors of all the phases are measured at these buses. This sums up to 12 complex measurement, i.e., 24

Table 3.1: Loads and DER Connections.

Bus	**Type**	**Connections**	**Bus**	**Type**	**Connections**
705	(G)	a-b, b-c	728	(L)	a-b, b-c, c-a
706	(G)	b-c	729	(L)	a-b
707	(G)	b-c, c-a	730	(L)	c-a
708	(G)	b-c	731	(L)	b-c
710	(G)	a-b	732	(L)	c-a
711	(G)	c-a	733	(L)	a-b
712	(L)	c-a	734	(L)	c-a
713	(L)	c-a	735	(L)	c-a
714	(L)	a-b, b-c	736	(L)	b-c
718	(L)	a-b	737	(L)	a-b
720	(L)	c-a	738	(L)	a-b
722	(L)	b-c, c-a	740	(L)	c-a
724	(L)	b-c	741	(L)	c-a
725	(L)	b-c	742	(L)	a-b, b-c
727	(L)	c-a	744	(L)	a-b

real measurements. We installed a unit at the substation, and then placed the rest to be almost evenly distributed along the network in order to achieve better observability.

- *Current magnitude measurements*: The magnitude of the current flow is measured on all phases of the lines that are marked with a rhombus in Fig. 3.4. The number of current magnitude measurements is 21 real measurements. We installed the units such that the state estimation problem can be solved without unobservability problems. We tested different installation for the current flow measuring devices with noiseless measurements, and then chose one such that the problem is not ill-posed.

- *Pseudo-measurments*: The aggregate load demand of the buses with load installed, which are blue-colored in Fig. 4.3, are estimated using a load forecasting algorithm using historical and situational data. Therefore, only two real quantities are obtained by the state estimator that relate to the active and reactive estimated load demand at the load buses. In addition, an energy forecast method is used to obtain an estimated injection from the renewable energy sources located at the

DER buses which are colored in red in Fig. 3.4. The total number of load buses in the feeder is 23, and the number of distributed energy sources is 6. Therefore, the state estimator obtains 58 real pseudo-measurements relating to the active and reactive forecasted demand/injection at these buses.

The state estimator obtains noisy measurements and inexact load demands and energy generation quantities. It is assumed that the noise in the PMU voltage measurements is drawn from a Gaussian distribution with zero mean and a standard deviation of 10^{-3}. Additionally, the noise added to current magnitudes is Gaussian distributed with a standard deviation of 10^{-2}. Finally, the differences between the pseudo-measurement and the real load demand and generations are assumed to be drawn from a Gaussian distribution with a standard deviation of 10^{-1}. The proposed learning-based state estimation approach aims at estimating the voltage phasor at all the phases of all the buses in the network.

The shallow neural network is trained using the TensorFlow [1] software library with 90% of the data used for training while the rest is used for verification. After tuning the network parameters, noisy measurements are generated and then passed to the state estimator architecture in Fig. 3.2. In order to show the effect of the modified cost function, we test the networks trained with different values of ϵ on $1,000$ loading and generation scenarios. Fig. 3.5 shows the histogram of the distance between the output of the shallow NN and the true network state. With the conventional training cost function ($\epsilon = 0$) the resulting distribution is more spread than the histogram that we obtain through the network trained with a relaxed cost function ($\epsilon = 1$).

Two performance indices (3.4)-(3.5) are introduced to quantify the quality of the estimate as well as the performance of the proposed approach. The first index, which is denoted by ν, represents the Frobenius norm square of the estimation error. Also, the cost function value at the estimate is denoted by μ.

$$\nu = \|\hat{\mathbf{v}} - \mathbf{v}^{\text{true}}\|_2^2 \tag{3.4}$$

$$\mu = \sum_{\ell=1}^{L} (z_\ell - h_\ell(\hat{\mathbf{v}}))^2 \tag{3.5}$$

Furthermore, in order to show the effect of changing the cost function used for

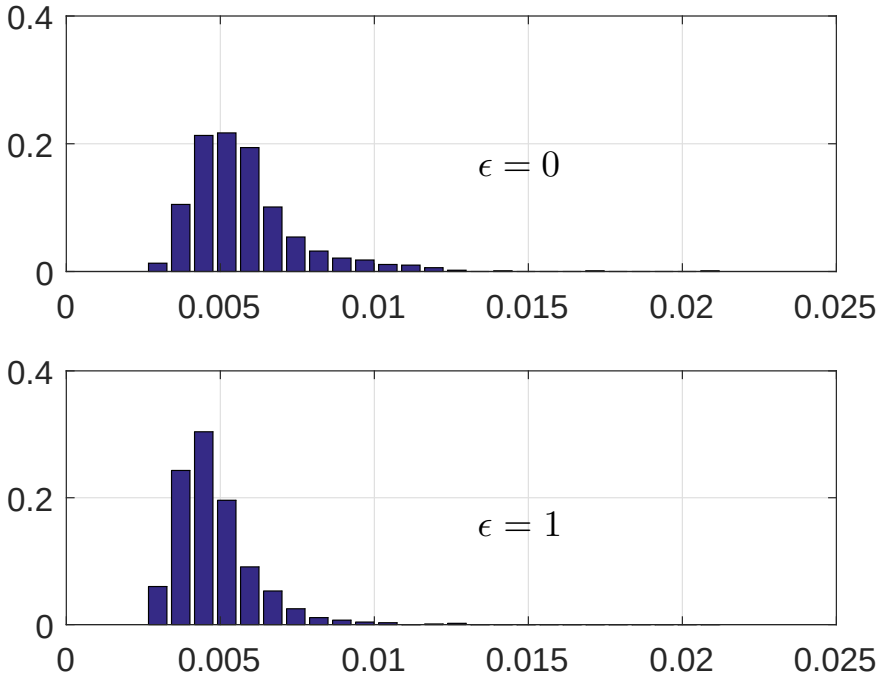


Figure 3.5: Histogram of the distance between the shallow NN output and the true voltage profile with ($\epsilon = 0$) and ($\epsilon = 1$).

training, the average cost achieved using the proposed approach is shown in Table 3.2 when different values of ϵ are used for training cost function. In addition, the average number of iterations required by the Gauss-Newton iterates to converge to the optimal estimate is also presented. Using a positive value of ϵ can lead to savings up 25% in computations, which is valuable when solving the DSSE for large systems. Also, it can be seen that choosing non-zero values for ϵ enhances the performance of the proposed architecture. The estimation accuracy can be almost 5 times better using a positive ϵ. As the approximation requirement is relaxed while training the shallow NN, the network gains in generalization ability, accommodating more scenarios of loading and generation profiles.

To assess the efficacy of the proposed approach we compare it against the complex variable Gauss-Newton state estimator using [91] as a state-of-art Gauss-Newton solver for a real-valued optimization problem in complex variables. The shallow NN was

Table 3.2: The estimator performance with different values of (ϵ).

$\boldsymbol{\epsilon}$	# Iterations	μ
$\mathbf{0}$	7.035	8.968×10^{-3}
$\frac{\mathbf{1}}{\mathbf{8}}$	6.825	5.531×10^{-3}
$\frac{\mathbf{1}}{\mathbf{4}}$	6.095	3.417×10^{-3}
$\frac{\mathbf{1}}{\mathbf{2}}$	5.675	1.822×10^{-3}
$\frac{\mathbf{1}}{\sqrt{\mathbf{2}}}$	5.220	5.056×10^{-3}
$\mathbf{1}$	6.150	5.859×10^{-3}
$\mathbf{2}$	6.415	1.365×10^{-2}

trained with $\epsilon = \frac{1}{2}$ in the next comparisons.

Table 3.3: Performance comparison of different state estimators

Method	ν	μ
Proposed	9.558×10^{-3}	1.822×10^{-3}
G-N	9.845×10^{-2}	4.861×10^{-2}

In Table 3.3, the average accuracy achieved in estimating the true voltage profile using both the Gauss-Newton method and the proposed architecture is presented for 1000 scenarios. In the Gauss-Newton implementation, the complex voltages provided by the PMUs are used to initialize the voltage phasors corresponding to these buses. This provides a better initialization point to the Gauss-Newton algorithm which also enhances its stability. Still, the proposed approach is able to achieve almost 10 times better accuracy on average. In addition, the fitting error which represents the WLS cost function is greatly enhanced using the proposed approach.

Table 3.4: Timing and convergence of different state estimators

Method	Time (ms)	# Divergence
Proposed	347	0
G-N	2468	28

In order to assess the computational time of the proposed algorithm, we tried 1000 simulated cases for the NN-assisted state estimator and the Gauss-Newton (optimization-only) state estimator. In Table 3.4, the number of divergent cases out of the 1000 trials is presented for both approaches. While the Gauss-Newton approach failed to converge in 28 scenarios, the proposed architecture has converged for all considered cases. In addition, the time taken by the proposed learning approach is almost four times less than the Gauss-Newton algorithm. This is due to the fact that only few Gauss-Newton iterations need to be done when the proposed approach is utilized.

3.4 System Reconfiguration

In distribution systems, the network configuration may be subject to changes either for restoration [76], i.e., to isolate a fault, or for system loss reduction [10,41]. An important task is to identify the underlying topology of the feeder. In order to perform this task, several approaches have been recently developed utilizing measurement data [16,26,106]. Without access to the correct network topology, accurate state estimation is untenable, as the estimator will attempt to fit the measurements to a wrong model. In other words, the ground truth function generating $\mathbf{z}$ is different from $\mathbf{h}(\mathbf{v})$, and hence, solving (2.3) is meaningless. Therefore, in this study we only consider the case where the (new) system topology has been adequately identified. Hence, in the latter part of the proposed approach, the Gauss-Newton iterations utilize accurate system topology information.

In order to test the robustness of the proposed approach, we assume that switches are available on several lines in the feeder as depicted in Fig. 3.4. Also, we add three additional lines to the network as redundant lines that are assumed to be unenergized under normal operating conditions [22]. Specifically, switches are assumed to be present on the original lines $(710, 735)$, $(703, 730)$ and $(727, 7444)$, and 3 additional tie lines $(742, 744)$, $(735, 737)$ and $(703, 741)$ are added to the feeder. We asses the robustness of the proposed learning approach under the following three scenarios.

- Scenario A: a fault has occurred in line $(727, 744)$ and the tie-switch on line $(742, 744)$ has been turned on;
- Scenario B: a fault has occurred in line $(703, 730)$ and the line $(703, 741)$ has been energized; and

- Scenario C: the switch on line $(710, 735)$ has been turned off, and the switch on line $(735, 737)$ has been connected.

We train the NN using data that are generated from the original network topology. We assess the performance of the proposed learning-based state estimator on Scenarios A, B and C. Note that, although in these scenarios the neural network is trained on a different generating model, the proposed approach still has advantageous performance when compared with the plain Gauss-Newton approach. This can be attributed to the specially designed cost function (3.3) that was proposed to train the NN. During the course of our experiments, we noticed that the robust performance against topology reconfiguration events is more pronounced when positive ε is used for training the NN. Table 3.5 compares the performance of the proposed state estimator with ($\varepsilon = \frac{1}{2}$) against the plain Gauss-Newton approach under the three system reconfiguration events. Clearly, the proposed approach still provides performance gains even under modest topology changes. When the approach was tested under significant reconfiguration events, our simulations showed that the initialization produced by projecting the flat voltage profile onto the linear space defined by the PMU measurements performs better than the neural network initialization. The training procedure is not computationally intensive due to the simplicity of the model and the relaxed training cost function. Therefore, in case of severe system reconfigurations that are expected to last for long time, the shallow neural network can be retrained in the order of a few minutes to match the underlying physical model.

Table 3.5: Performance comparison with system reconfiguration.
(Averaged over 100 runs)

Scenario		Time (ms)	ν	μ
A	**Proposed**	476	3.894×10^{-2}	3.172×10^{-2}
	G-N	1574	7.460×10^{-2}	1.147×10^{-1}
B	**Proposed**	2339	7.963×10^{-2}	1.051×10^{-2}
	G-N	3696	9.207×10^{-2}	2.050×10^{-2}
C	**Proposed**	429	1.297×10^{-2}	3.150×10^{-3}
	G-N	1568	8.527×10^{-2}	1.289×10^{-1}

Chapter 4

Physics-Aware Neural Networks for DSSE

Exploiting valuable information from abundant real-time and historical data, data-driven approaches hold the promise to significantly enhance monitoring accuracy and improve the performance of distribution networks. To that end, neural network approaches have been used to estimate the bus injections from the real-time measurements in [66]. The estimated bus injections can be used as pseudo-measurements to compensate for the scarcity of real-time measurements. In addition, plain feed-forward neural networks (NN) were proposed to estimate the network state from the measurements in [11]. This approach reduces the complexity of the state estimation task to matrix-vector multiplications by shifting the computational burden to an off-line training stage utilizing historical or simulated data. It is often challenging to avoid exploding or vanishing gradients while training these feed-forward NNs, and thus the provided estimates are less accurate than any optimization-based approach. A joint optimization/learning approach was proposed in Chapter 3. Different from that technique the authors of [119] devised a learning approach where a deep NN is constructed by unfolding an iterative solver for the least-absolute-value formulation of the state estimation problem in transmission networks [101].

All past learning models for state estimation overlook the physics of the underlying distribution network, hence leading to over-parameterization of the mapping from the

measurements to the network states. In order to utilize our knowledge of the physical system that governs the relationship between the network states and the measurements, we propose a novel neural network architecture exploiting the distribution network structure. We start by showing that owing to disparity in the accuracy of the measurements, i.e., the μ-phasor-measuring unit (μPMU) measurements are far more accurate than any other measurements in the network, the DSSE problem can be (approximately) partitioned into smaller problems. Therefore, the estimation of the state (voltage) at a certain bus in the network can be done using the measurements taken at the partition/partitions where this bus is located. In this chapter, we formally describe the partitioning of the DSSE problem that results from installing μPMUs in the distribution network. In addition, the quality of the partitioning resulting from a certain installation of μPMUs in the network can be assessed using the diameter measure which is related to the size of partitions generated. We propose a greedy algorithm for installing μPMUs in the distribution network to minimize the diameter of the resulting partitioning. Simulations results on the IEEE-37 distribution feeder show that the proposed learning approach achieves superior performance in terms of estimation accuracy. Also, the running time is in the order of milliseconds which allows for real-time monitoring of the distribution network.

The main idea of the proposed NN architecture is to zero out the weights of the measurements taken outside a particular partition when calculating the estimate of the network states at the buses inside that partition. To do so, the structure of the network admittance matrix is embedded on the weights matrix that maps the NN iterates. The uderlying physical model that governs the operation of the distribution network is utilized to sparsify the learning model, and thus the proposed model is called *Physics-aware neural network* (PAWNN) where the pruning is done in a deterministic manner before training. In the proposed architecture, the output of a K-layer PAWNN that relates to the estimated voltage at a certain bus is only a function of the measurements taken at most K hops away from this bus. Therefore, in order to realize the mapping from the measurements to the states, the number of layers in the PAWNN has to match the diameter of the partitioning resulting from the μPMUs placement. In our simulations, we show that the proposed greedy algorithm for μPMUs placement achieves near optimal performance in terms of minimizing the diameter of the resulting partitioning.

The proposed NN architecture reduces the number of the trainable parameters, which prevents over-fitting. Also, it inherently provides robustness as any reconfiguration in the network topology will only affect the state estimates in this specific directly affected areas. For example, if the neural network has K layers, any measurement contributes only to the state estimates at buses that are at most K hops away from the location where the measurement is taken. Similarly, for the case of measurement outliers resulting from damaged meters or communication failures, the estimation of the state at distant locations from the outliers measurements will not be affected.

4.1 Partitioned DSSE

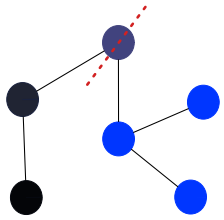
This section presents the required background for the partitioning of the DSSE problem that results from installing μPMUs that provide very accurate measurements. First, we introduce the vertex-cut partitioning which divides the edges of the graph into disjoint sets. Then, we show that installing μPMUs in the network results in a vertex-cut partitioning of the DSSE problem. Throughout this section, we use vertex, node, and bus interchangeably. Similarly, we use edge and line to refer to any connection in the graph.

Definition 1. *An* articulation vertex *of a connected graph is a vertex whose removal disconnects the graph [17, §2.4].*

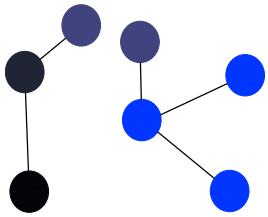
According to the definition, all the vertices in a tree graph are articulation points since removing any vertex disconnects the graph. Next, we define vertex-cut partitioning which partitions the set of the edges in the graph into multiple disjoint subsets.

Definition 2. *A* vertex-cut partitioning *refers to a partitioning of the edge set $\mathcal{L}$ into K subsets $\mathcal{L}_k$, such that $\mathcal{L}_k \in \mathcal{L}$, $\cup_{1\leq k\leq K}\ \mathcal{L}_k = \mathcal{L}$, and $\mathcal{L}_k \cap \mathcal{L}_{k'} = \phi$ for $k \neq k'$. Any vertex that holds an endpoint of an edge $(l, m) \in \mathcal{L}_k$ is placed in $\mathcal{N}_k$.*

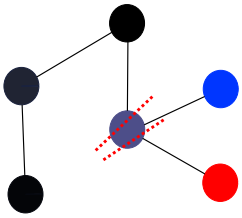
Examples of vertex-cut partitioning are depicted in Fig. 4.1. If the original graph is a tree, then the number of disjoint subsets of edges that result from choosing a vertex to be cut is equal to the number of edges connected to that vertex. In Fig. 4.1(a) and (b), we show the two graphs resulting from cutting a vertex that has two edges in the original



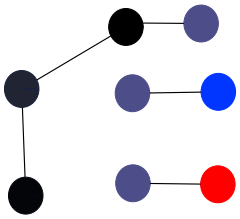
(a) An articulation point with two edges



(b) Resulting vertex-cut partitioning from Fig. 4.1a



(c) An articulation point with three edges



(d) Resulting vertex-cut partitioning from Fig. 4.1c

Figure 4.1: Vertex-cut partitioning Examples over a tree graph

graph. When the vertex to be cut has three edges as in Fig. 4.1 (c), the number of the resulting subgraphs is three where the cut-vertex is replicated in all the subgraphs.

For the purpose of our mathematical proof, we adopt the following assumption on the accuracy of the PMU measurements.

Assumption 1. *The PMU measurements are noiseless, i.e., the variance of the measurement noise associated with the PMU measured quantities is negligible.*

This assumption is not totally unrealistic as the signal to noise ratio in the PMU measurements is in the range of 40 to 50 dB [14]. Note that, this assumption is only used for our mathematical proof, but not assumed or invoked in any of the simulations.

Estimating the voltages (states) of all the buses in the network is usually done by solving an optimization problem as in (2.3). Abstracting this concept, we can say that there is a mapping $\mathbf{F}(\cdot)$ such that $\mathbf{F}(\mathbf{z}) = \hat{\mathbf{v}}$. Suppose $\mathcal{P} \subset \mathcal{N}$ is the set that comprises the buses with μPMUs installed. Since the PMU measurements are noiseless, the state estimation problem can be reduced to estimating the network voltages (states) only at the buses without PMUs ($\mathcal{N}\backslash\mathcal{P}$). Such mapping is denoted by $\mathbf{F}_r(\mathbf{z}) = \hat{\mathbf{v}}_{\mathcal{N}\backslash\mathcal{P}}$ where $\mathbf{v}_{\mathcal{N}\backslash\mathcal{P}}$ collects the voltages at all the nodes without PMUs. The next theorem shows the separability of this mapping over the partitioned graph.

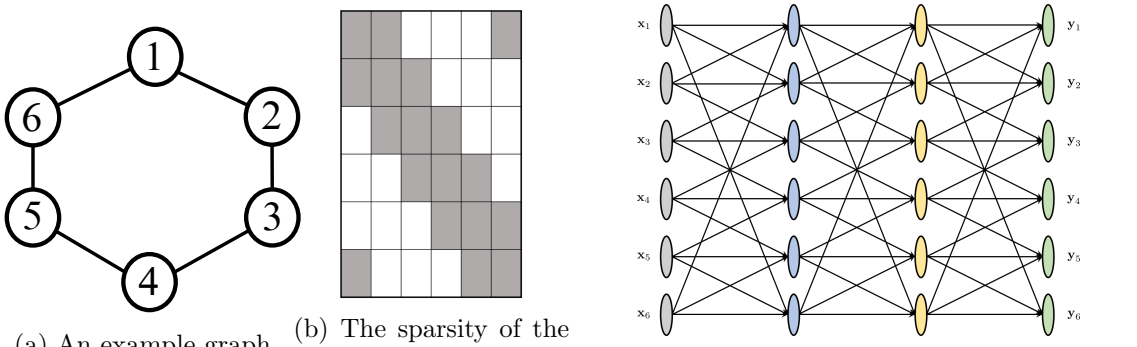
Theorem 2. *Suppose* $\{\mathcal{L}_k\}_{k=1}^{K}$ *are the disjoint partitions that result from cutting the vertices in* $\mathcal{P}$*. In addition,* $\mathcal{N}_k$ *denotes the set of nodes connected to the edges in* $\mathcal{L}_k$*, and* $\overline{\mathcal{N}}_k = \mathcal{N}_k \backslash \mathcal{P}$*. Then, the mapping* $\mathbf{F}_r(\mathbf{z}) = \mathbf{v}_{\mathcal{N}\backslash\mathcal{P}}$ *is separable over the vertex-cut partitioning, i.e., for each set* $\overline{\mathcal{N}}_k$*, the mapping* $\mathbf{F}_r(\cdot)$ *can be written as* $\mathbf{F}_r^{(k)}(\mathbf{z}^{(k)}) = \hat{\mathbf{v}}_{\overline{\mathcal{N}}_k}$*, where* $\mathbf{z}^{(k)}$ *comprises all the measurements taken at the buses* $\mathcal{N}_k$ *and the edges* $\mathcal{L}_k$*, and* $\hat{\mathbf{v}}_{\overline{\mathcal{N}}_k}$ *collects the voltages at the buses in* $\overline{\mathcal{N}}_k$*.*

Proof. Any measurement synthesizing function $h_\ell(\mathbf{v})$ taken at a bus n is a function of the state at the bus n and the buses connected to bus n. Similarly, for a current or power flow measurement taken at an edge (l, m), the synthesizing function is a function only of the state at the buses l and m. As the PMUs provide exact complex voltage measurements at the buses $\mathcal{P}$, any measurement taken at a bus $n \in \mathcal{N}_k$ on a line $(l, m) \in \mathcal{L}_k$ depends only on the state at the buses $\overline{\mathcal{N}}_k$. Therefore, the measurement synthesizing function of any measurement taken outside a certain partition $\mathcal{L}_k$ does not involve the states (voltages) at the nodes $\overline{\mathcal{N}}_k$. Hence, the state estimation mapping for $\hat{v}_{\overline{\mathcal{N}}_k}$ is function only of the measurements $\mathbf{z}^{(k)}$, which proves the theorem. □

The aforementioned theorem provides an insight regarding the separability of the state estimation problem in presence of PMUs. This separability is critical when a learning model is used to estimate the state of the network from the measurements. It is clear now that a learning model that estimates the voltage at a certain bus does not require knowledge of all the measurements in the network. Instead, by having an accurate measurement of the state (voltage) at a certain bus, all the measured quantities behind this bus can be discarded. This will play an important role in reducing the complexity of our learning model used for the task.

4.2 Graph-Pruned Neural Networks for DSSE

In this section, we present our novel learning model for DSSE. The graph-pruned neural network is composed of multiple layers whose connections reflect the distribution network connections. Let the input of the NN be denoted by $\mathbf{x}$, and the NN produces an output $\mathbf{y}$ using a stacked layered architecture in which each layer realizes a linear transformation and a point-wise nonlinearity. The vector $\mathbf{y}$ is partitioned into N parts



(a) An example graph (b) The sparsity of the weights (c) Graph-pruned NN for the graph in Fig. 4.2a

Figure 4.2: Example graph with the corresponding graph-pruned NN

that represent features of each node in the graph, e.g., the voltage at the buses. The intermediate output at the t-th layer of the NN is denoted by $\mathbf{h}_t \in \mathbb{R}^{Nd_t}$ where d_t represents the dimension of each partition in $\mathbf{h}_t$. Formally, the t-th layer output is computed using the following transformation

$$\mathbf{h}_{t+1} = \sigma_l(\mathbf{W}_t \mathbf{h}_t) \tag{4.1}$$

where σ_t is a point-wise nonlinearity, and the matrix $\mathbf{W}_t \in \mathbb{R}^{Nd_{t+1} \times Nd_t}$ is composed of $N \times N$ blocks of size $(d_{t+1} \times d_t)$. The (i, j) block in the matrix $\mathbf{W}_t$ is zeroed out (pruned) if the nodes i and j are not connected in $\mathcal{G}$, which justifies the name of the proposed learning model as *graph-pruned* NN.

Example 1. *Consider the graph in Fig. 4.2a. Suppose a three-layer graph-pruned NN is designed to estimate some features of the nodes in the graph from signals (measurements) at the network nodes. The input vector* $\mathbf{x} := [\mathbf{x}_1^T,\ \mathbf{x}_2^T,\ \ldots,\ \mathbf{x}_6^T]^T$*, and output vector* $\mathbf{y}$ *is also composed of six components* $\mathbf{y}_i$ *for* $1 \leq i \leq 6$*. The output of the NN can be written as*

$$\mathbf{y} = \sigma_3\Big(\mathbf{W}_3\ \sigma_2\big(\mathbf{W}_2\ \sigma_1(\mathbf{W}_1\mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2\big) + \mathbf{b}_3\Big)$$

where the structure of the weight matrices $\mathbf{W}_i$ *is depicted in Fig. 4.2b. In addition, let* $\mathbf{h}_{t,i}$ *denote the* i*-th block of the output of the* t*-th layer in the NN. Fig. 4.2c shows the dependency of the block of each layer in the NN on the block of the previous layer. For example,* $\mathbf{h}_{1,2}$ *is function of* $\mathbf{x}_1, \mathbf{x}_2$ *and* $\mathbf{x}_3$ *only. Notice that since the network is*

composed of three layers, each output block $\mathbf{y}_i$ *is function of the inputs related to nodes that are at most three hops away in the graph. That is, each output is function of all the inputs in this case as the pairwise distance in the graph is at most three.*

The graph convolutional NN (GCNN) [46, 52] learning approaches are designed to process data defined over graphs. These models also lead to sparsification of the weight matrices connecting the hidden layers in the neural network. In addition to being derived from the physical model governing the operations, the proposed learning model leads to a more general parameterization. For instance, the GCNN model proposed in [52] is tantamount to enforcing the blocks of the weight matrices to be scaled versions of a single matrix. For instance, if the small gray blocks in Fig. 4.2b are constrained to be scaled versions of each other, then the proposed method becomes equivalent to the GCNN in [52]. On the other hand, if the vertically aligned blocks in the weight matrices of PAWNN are chosen to be scaled versions of a fixed matrix, then the learning model of [46] emerges as a special case. For example, for the PAWNN in Fig. 2, if we constrain the blocks $(2, 1)$ and $(6, 1)$ to be scaled versions of the block $(1, 1)$ and similarly for all other vertically aligned blocks, the model reduces to the graph NN proposed in [46].

4.2.1 Required number of layers

As established in Section 4.1, the DSSE problem can be partitioned by installing PMUs that essentially break the dependencies of the estimated state at a certain bus on any measurement taken outside its partition. Therefore, the graph-pruned NN has the potential to realize mappings such that the estimated state at a certain bus is a function only of the measurements taken in the same partition. In order to characterize the number of layers required to realize the DSSE mapping for a network with PMUs installed, we introduce the following definitions.

Definition 3. *The* eccentricity *of a vertex* v *in* $\mathcal{G}(\mathcal{N}, \mathcal{L})$ *is the maximum shortest path length from* v *to all other vertices in* $\mathcal{N}$*.*

Definition 4. *The* diameter *of a graph* $\mathcal{G}(\mathcal{N}, \mathcal{L})$ *is the maximum eccentricity of all the vertices in* $\mathcal{N}$*.*

Definition 5. *The* diameter *of a vertex-cut partitioning is the maximum diameter of*

the subgraphs $\mathcal{G}_k(\mathcal{N}_k, \mathcal{L}_k)$ for $1 \leq k \leq K$, which is introduced by cutting the vertices in a set $\mathcal{P} \subseteq \mathcal{N}$, and it is denoted by dia($\mathcal{P}$).

For example, the diameter of the vertex-cut partitioning in Fig. 4.1b is 2, while the diameter of the vertex-cut partitioning in Fig. 4.1d is 3. Therefore, a two-layer graph-pruned NN can penitentially approximate the mapping between the measurements and the states if a PMU unit is installed at the vertex that was cut in Fig. 4.1b. However, with the cut in Fig. 4.1d, at least three-layer graph-pruned NN is needed. Now, it is reasonable to ask how to place available PMUs such that the diameter of the resulting vertex-cut partitioning is minimized. In the next subsection, we present a greedy algorithm that provides a simple approximate solution for this problem.

4.2.2 Greedy algorithm for PMU placement

As shown in Theorem 2, the placement of μPMUs in the distribution network makes the DSSE problem separable. In the experiments section, we will show that under realistic setup the decoupled DSSE subproblems are almost equivalent to the original formulation. Therefore, the μPMUs need to be placed in the feeder such that the resulting subproblems are balanced. In other words, we tackle the problem of minimizing the diameter of the resulting vertex-cut partitioning given a certain budget of μPMUs. We present a greedy algorithm that provides an approximate solution of this placement problem.

The proposed greedy algorithm tackles the problem of installing μPMUs one at a time. It is clear that the optimal placement of a μPMU in order to reduce the diameter of the resulting vertex-cut partitioning is to install it in the middle of the longest shortest path in all the partitions. Therefore, our algorithm starts by finding the maximum length shortest path in the network, and the μPMU is installed in the middle of this path. Then, the process continues by finding the maximum length shortest path in all the resulting subgraphs, and then placing the next μPMU along the maximum length path in all the subgraphs. The process continues until the available budget of μPMUs is exhausted. Algorithm 1 summarizes the main steps of the proposed approach.

In order to find maximum length shortest path in a tree a simple algorithm is used. Let us consider a subgraph $\mathcal{G}_k(\mathcal{N}_k, \mathcal{L}_k)$. First, we choose a random starting point $n \in \mathcal{N}_k$

Algorithm 1: Greedy Algorithm for μPMU Placement

Input: : graph $\mathcal{G}(\mathcal{N}, \mathcal{L})$ and $K \geq 1$ budget of μPMUs

Output: : set $\mathcal{S} \subseteq \mathcal{N}$ where $|\mathcal{S}| = K$

Initialization: : $\mathcal{S} = \phi$

repeat

[S1] Determine the maximum length shortest path in all the subgraphs $\mathcal{G}_k(\mathcal{N}_k, \mathcal{L}_k)$ resulting from cutting the vertices in $\mathcal{S}$

[S2] Place a μPMU in the middle of the longest path identified in [S1]

until $|\mathcal{S}| = K$

and perform depth-first search (DFS) to find the eccentricity of n which is achieved for the path from n to vertex $n' \in \mathcal{N}_k$. Now, we use DFS to find the eccentricity of node n'. The length of the path achieving maximum shortest path length from n' is the diameter of $\mathcal{G}_k$. Therefore, the complexity of each step is $\mathcal{O}(|\mathcal{N}|)$, and the process is repeated K times. Hence, the total complexity of Algorithm 1 is $\mathcal{O}(K|\mathcal{N}|)$.

4.3 Experimental Results

In this section, the proposed graph-pruned NN is utilized to estimate the state of the benchmark IEEE-37 distribution feeder. This network was recommended by the Test Feeder Working Group of the Distribution System Analysis Subcommittee of the IEEE PES for evaluating the performance of the state estimation algorithms [86]. The feeder has several delta-connected loads and is known to be highly unbalanced. The load buses are blue-colored in Fig. 4.3. In addition, some nodes in the feeder feature different types of connections, i.e., single-, two-, and three-phase connections. Renewable energy sources (RES) are installed at six different buses, which are colored in red in Fig. 4.3. The types of the connections of all the loads and RESs are presented in Table 4.1 where (L) and (G) mean load and RES, respectively.

We evaluate the performance of the proposed greedy algorithm for placing PMUs in the distribution feeder. We compare the diameter of the partitioning induced by the optimal placement of PMUs and by the partitioning resulting from our proposed approach. For the optimal placement, we use exhaustive search in order to identify

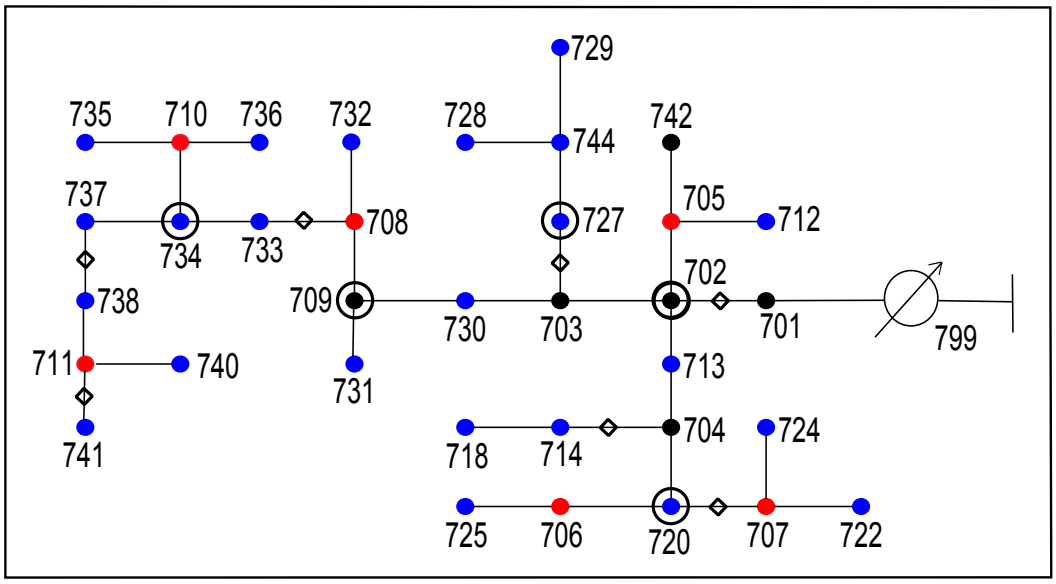


Figure 4.3: IEEE-37 distribution feeder. Nodes in blue are with loads, and red nodes represent buses with DER installed.

the placement that minimizes the diameter of the resulting vertex-cut partitioning. Table 4.2 compares the placement of PMUs using our proposed approach against the optimal placement in terms of diameter of the resulting vertex-cut partitioning. The results shows that the proposed approach produces optimal placement in all cases except for 2 and 6 PMUs where the diameter is larger by only one. The set $\mathcal{S}_g$ denote the buses with PMUs installed using the greedy algorithm, while $\mathcal{S}_o$ denote the optimal placement of PMUs.

Training samples were generated using the load and renewable generation dataset available in [7] modulated by the nominal values of the loads. The power flow solver [38] was used to find the voltage profile (network states), and then the noisy measurements were generated using the measurement synthesizing functions (2.1) and (2.2). The variance of the noise added to the PMU measurements was set to be 10^{-6}, and the variances of the current magnitude and pseudo-measurements noise were 10^{-3} and 10^{-2}, respectively. A total of $100,000$ samples were used to train graph-pruned NNs using the TensorFlow [1] software library with 90% of the data used for training and 10% for validation. The neural network used consisted of 4 layers where the widths from the first

Table 4.1: Loads and DER Connections.

Bus	**Type**	**Connections**	**Bus**	**Type**	**Connections**
705	(G)	a-b, b-c	728	(L)	a-b, b-c, c-a
706	(G)	b-c	729	(L)	a-b
707	(G)	b-c, c-a	730	(L)	c-a
708	(G)	b-c	731	(L)	b-c
710	(G)	a-b	732	(L)	c-a
711	(G)	c-a	733	(L)	a-b
712	(L)	c-a	734	(L)	c-a
713	(L)	c-a	735	(L)	c-a
714	(L)	a-b, b-c	736	(L)	b-c
718	(L)	a-b	737	(L)	a-b
720	(L)	c-a	738	(L)	a-b
722	(L)	b-c, c-a	740	(L)	c-a
724	(L)	b-c	741	(L)	c-a
725	(L)	b-c	742	(L)	a-b, b-c
727	(L)	c-a	744	(L)	a-b

Table 4.2: Optimal and greedy PMU placement.

# PMUs	$\mathcal{S}_g$	$dia(\mathcal{S}_g)$	$\mathcal{S}_o$	$dia(\mathcal{S}_o)$
1	{709}	9	{709}	9
2	{702, 709}	7	{702, 708}	6
3	{702, 709, 734}	5	{702, 709, 734}	5
4	{702, 709, 720, 734}	5	{702, 709, 720, 734}	5
5	{702, 709, 720 , 727, 734}	4	{702, 709, 720, 727, 734}	4
6	{702, 709, 720, 727, 734, 738}	4	{703, 708, 713, 720, 733, 737}	3

hidden layer to the output layer are $1680, 840, 420, 210$, respectively. The estimate of the voltages at a bus is represented using 6 outputs representing the real and imaginary parts of the voltage phasor at each phase. The graph structure of the network is imposed on the NN connectivity, i.e., the first 48 neurons in the first hidden layer, which represent bus 701, are only connected to the 24 neurons representing node 702 in the second layer.

We test the proposed graph-pruned NN on two different scenarios of measurements. In the first case (Scenario A), we employ 5 PMUs installed according to the proposed greedy algorithm, which is optimal. In Fig. 4.3, the buses where the PMUs are installed are circled, and the lines where the current magnitudes are measured have rhombuses on them. The net load and renewable energy generation at all the phases of the buses with loads or RESs installed are used as pseudo-measurements. The number of layers in the graph-pruned NN in this scenario is 4 as the diameter of the resulting partitioning of this placement is equal to 4. In this scenario, the total number of measurements is 103, which consist of 3 complex measurements of voltages at 5 buses with PMUs, 3 real measurement of current magnitudes installed at 7 locations in the network, and 26 complex pseudo-measurements at the buses with loads or RESs installed and without PMUs. In the second case (Scenario B), we evaluate the performance of the proposed graph-pruned neural network under different measurements scenario [114], where PMUs are installed at the buses {701, 704, 709, 734}, the current magnitude measurements are taken on the lines { 702→ 703, 702→ 713, 707→ 720, 703→ 727, 708→ 738, 737→ 738, 711→ 741 }, and the forecasts for loads and renewable energy are used as pseudo-measurements. We use a 5-layer graph pruned NN in this case to match the diameter of the vertex-cut partitioning resulting from the PMU placement.

In order to assess the performance of the proposed approach, we define the average estimation accuracy $\boldsymbol{\nu}$ of any algorithm as follows.

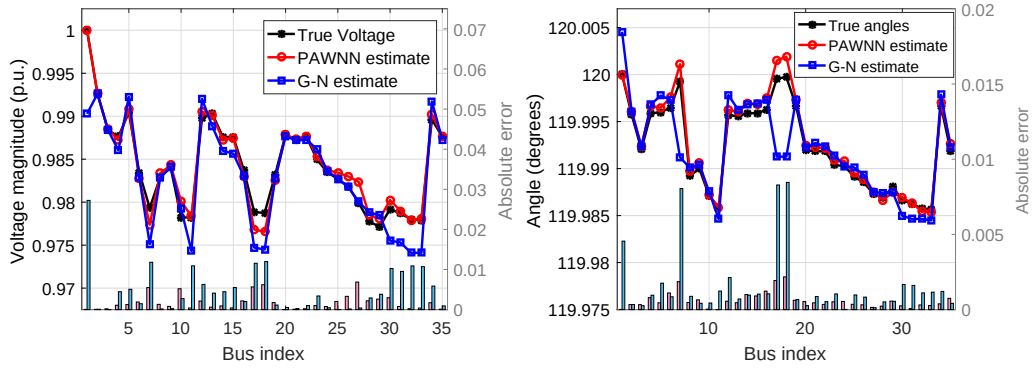
$$\boldsymbol{\nu} = \frac{1}{N} \sum_{i=1}^{N} \|\hat{\mathbf{v}}_i - \mathbf{v}_i^{\text{true}}\|_2^2 \tag{4.2}$$

where $\hat{\mathbf{v}}_i$ is the estimated voltage profile from the noisy measurements generated using $\mathbf{v}_i^{\text{true}}$. Table 4.3 shows the average performance of the proposed learning approach and the Gauss-Newton algorithm over 1000 cases. Simple feed-forward neural networks

Table 4.3: Performance comparison of different state estimators

Method	Scenario A		Scenario B	
	$\boldsymbol{\nu}$	Time (ms)	$\boldsymbol{\nu}$	Time (ms)
PAWNN	1.273×10^{-3}	1.146	2.598×10^{-3}	1.034
G-N	5.833×10^{-1}	866	4.161×10^{-1}	1076

approaches require significantly large amount of training data and computational resources. In addition, they often suffer from exploding or diminishing gradients. This results in bad estimates for the state of the network. For example, the average estimation accuracy of the state using a 4-layer feed-forward NN was 2.69×10^{-1} for noiseless measurements in Scenario A. Hence, we did not include feed-forward NN in our comparisons. The Gauss-Newton algorithm is initialized using the flat voltage profile. Clearly, the proposed learning method achieves superior performance in both scenarios, where the accuracy of estimation is an order of magnitude better than state-of-the-art Gauss-Newton approach. In addition, since the proposed learning method alleviates all the computational burden at the estimation time by shifting it to the training time, the running time of the proposed approach is three orders of magnitude higher than the optimization-based approach.



(a) Voltage magnitudes (b) Voltage angles

Figure 4.4: Estimation of the voltage magnitudes and angles at phase (b) of all buses in the IEEE-37 feeder. (The absolute estimation errors are depicted in pink and light blue for the proposed approach and the Gauss-Newton solver, respectively.)

In order to show the quality of estimates provided by the proposed Graph-pruned neural network, we present the estimate of the voltage magnitudes and angles at phase (c) at all buses. Fig. 4.4 depicts the estimated voltage magnitudes and angles using the Gauss-Newton method and the proposed Graph-pruned NN approach. Also, the absolute estimation error of the magnitudes and angles are shown in faded colors. The results show superior estimation performance for the proposed approach.

4.3.1 Robustness of PAWNN

As discussed earlier, the proposed approach is inherently robust against measurements failure or attacks. That is, erroneous measurements are not propagated in the neural network more than the number of layers. Therefore, only the estimation of voltages in the neighborhood is affected. In order to showcase the robustness of the approach, we tested the proposed learning model in a scenario where the measurements of the μPMU installed at bus 734 are corrupted with Gaussian noise with a standard deviation of 10. While the proposed approach is oblivious to the noise level of each measurement, the weight in the weighted least squares formulation used by the Gauss-Newton approach were adjusted to account for the huge noise variance of the measurements at bus 734. The estimation of the voltage magnitude along all buses at phase (b) for both approaches is depicted in Fig. 4.5. It is noticeable that the estimation of voltages around the bus indexed 27, which is the bus index of bus 734, are the only affected estimation, while the G-N estimate is totally corrupted by corrupting the measurements of only one measuring unit.

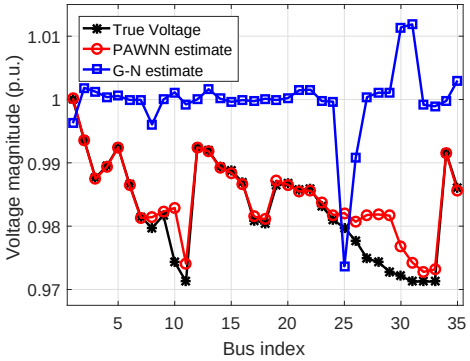


Figure 4.5: Estimation of the voltage magnitudes at phase (b) of all buses under corrupted μPMU measurements at bus 734.

Chapter 5

Optimal Power Flow for Multiphase Systems with Renewables

The AC optimal power flow (OPF) problem is a predominant task in optimizing the performance of power grids. The OPF problem aims at minimizing an appropriate operational cost while respecting the network's physical and engineering constraints. However, due to the quadratic nature of the power flow equations, the OPF problem is known to be nonconvex and NP-hard in general [56, 58]. Existing approaches to the OPF problem range from classical Newton-Raphson, to Lagrangian relaxation, genetic algorithms and interior point methods. Unfortunately, these methods do not provide optimality or feasibility guarantees except in certain cases, and are quite sensitive to the initial guess. While the Newton-Raphson method has been traditionally employed to solve AC OPF problems for transmission systems, its convergence is challenged when it is applied to multi-phase distribution networks; this is primarily due to the high resistance-to-reactance ratio of distribution lines, which can cause the Jacobian matrix to be ill-conditioned.

Many recent research efforts have been trying to approach the solution of OPF problem using relaxation techniques [5, 12, 18, 33, 37, 47, 56, 64, 117, 118]. Among those, semidefinite relaxation (SDR) was shown to be able to find the global optimal solution

of the problem in many cases. SDR relies on matrix-lifting and rank relaxation to convexify the feasible set of the OPF problem [5,23,56,63,65,117]; the resulting relaxed problem can be solved in polynomial time. OPF-optimality of the SDR solution can always be tested *a posteriori* by checking the rank of the SDR solution matrix; but it is very useful to know *a priori* in which cases SDR will yield an optimal solution for the original nonconvex OPF problem. These are the cases when SDR yields a solution that is rank-one, or can be easily transformed to rank-one. In those cases, SDR is not a relaxation after all; we say that SDR is *tight*. Tightness of SDR relaxation was proved for a number of network setups under restrictive conditions. In [56], tightness of SDR was shown for a network comprising only resistive loads, provided load over-satisfaction is allowed and the dual variables are all positive. Assuming availability of sufficient phase shifters, it was proven that SDR is exact if load over-satisfaction is allowed [90]. For tree/radial networks, under operational constraints on voltage magnitudes, line losses, and line flows, the SDP relaxation was shown to be tight if there are no lower limits on the power generation [117]. This result was extended in [57] for radial networks with lower limits only on the active power, under reasonable conditions. The inexactness of SDR for a general network was demonstrated in [59] using a simple 3-bus network. Unfortunately, in cases where SDR is not tight, it is not easy to recover a physically meaningful solution from the solution matrix; only a lower bound on the optimal cost is provided. An approach to recover an OPF solution from the SDR solution was pursued in [63], but still there is no guaranty of recovering a physically meaningful OPF solution.

Another relaxation technique was proposed in [47] for radial networks. The method eliminates the voltage angles by defining new variables representing the real and imaginary parts of the second order voltage moments, and then expresses the power flow equation in terms of the new variables. [47] solves the problem using second order cone programming (SOCP). Tightness of this SOCP relaxation is an open issue. Along the same lines, a quadratic convex (QC) relaxation approach was proposed in [18] where the network constraints are replaced by convex surrogates. Although shown to be tighter than the SOCP relaxation, the QC relaxation also lacks proof of tightness, and can return solutions that are infeasible for the original OPF problem.

As a generalization of SDR, moment-based relaxation has been proposed in [72] using the Lassarre Hierarchy. Empirically, the method has been demonstrated to be able

to find the OPF solution in cases where SDR fails. The moment-based relaxation introduces higher order voltage moments as new variables and defines the power flow quantities in terms of these moments. On the down side, considering higher order moments requires solving very large SDP instances which may not be computationally tractable. Aiming to alleviate the computational burden of moment-based relaxation, [73] exploited the structure of the OPF problem to develop a more tractable solution for low order moments. However, due to the NP-hardness of the problem, the moment order required to approach the optimal solution may be very large. Building on the same tool, a Laplacian-based approach has been proposed in [70], where an upper bound on the cost function is assumed, and the cost is replaced by a function that penalizes constraint violations.

It is also worth emphasizing that the AC OPF task is becoming increasingly important for distribution systems with high integration of renewable energy resources (RESs), where adjustments of the real and reactive output-powers from renewable sources are necessary in order to enforce voltage regulation. Particularly relevant is the case of distribution feeders with high penetration of photovoltaic (PV) systems, where reverse power flows induced by PV-systems operating according to current practices may increase the likelihood of overvoltage conditions. OPF formulations aim at minimizing the cost of real power curtailment as well as the cost for reactive power support, while concurrently pursuing utility-oriented objectives and ensuring voltage regulation [13, 24, 43, 95, 98]. However, in this particular setting, the overall cost function of the OPF task may not be strictly increasing in the power injections, which implies that relaxation methods such as SDR [23, 36, 83] are not guaranteed to be tight.

The main contributions of this chapter are as follows:

1. *Designing an efficient algorithm that can solve the OPF problem when the relaxation approaches fail to find an optimal solution.* Inspired by recent advances in solving nonconvex quadratically-constrained quadratic programs (QCQP), the OPF problem is formulated as a nonconvex QCQP. In [68], a *Feasible Point Pursuit Successive Convex Approximation* (FPP-SCA) algorithm was proposed, and empirically shown to be very effective in solving nonconvex QCQP problems in cases where SDR fails. The FPP-SCA algorithm replaces the nonconvex constraints by inner convex surrogates around a specific point to construct a convex

restriction of the original problem. Such restriction may lead to infeasibility, even if the original problem is feasible. The main idea behind FPP-SCA is to allow a controllable amount of constraint violations to enable the algorithm to make progress towards feasibility in its initial stages. Towards this end, a slack variable is added to ensure feasibility at each step, and the cost function is augmented with a term that penalizes the slack that reflects the constraints violations. The overall approach is neither restriction nor relaxation, but rather judicious *approximation* of the OPF problem in each iteration, the solution of which is subsequently used as the approximation point for the next iteration. Upon finding a feasible voltage profile, successive convex approximation of the feasible set is used to find a KKT point of the OPF problem.

2. *Fomulating the OPF for radial networks with wye- and delta connections.* A power flow model for radial multiphase systems that accommodates wye-connected and delta-connected distributed energy resources is outlined. Then, the FPP-SCA is utilized for find local solution of the OPF in these scenarios.

3. *Identifying OPF solutions when minimizing the cost of active power curtailment and reactive power support.* The modified problem is solved to obtain an optimal voltage profile that conforms to the power system's operational and economic constraints. SDR is very sensitive to the choice of the cost function (especially when the cost function is non-increasing with the power flows in the network). On the other hand, the proposed algorithm is shown to be an effective approach for solving the modified OPF problem for single-phase and multi-phase system models.

4. *Performance comparison.* The performance of the FPP-SCA is benchmarked against existing convex relaxation approaches and the IPOPT solver. Results demonstrate that the FPP-SCA algorithm is able to find solutions that are optimal or near optimal, even in cases where convex relaxation approaches and the IPOPT solver fail.

5. *Identifying problematic constraints when the AC OPF problem is infeasible.* The FPP algorithm has the ability to identify problematic constraints in cases where

the AC OPF problem is infeasible. This is a distinct feature of the proposed method that off-the-shelf solvers such as IPOPT do not offer.

5.1 Power flow model

Consider a multi-phase network comprising $N+1$ buses. The system is modeled by a graph $\mathcal{G} := (\mathcal{N}, \mathcal{L})$, where $\mathcal{N} := \{0, 1, 2, \cdots, N\}$ is the set of multi-phase buses (nodes) and $\mathcal{L} \subseteq \mathcal{N} \times \mathcal{N}$ represents the set of lines. Let bus 0 be the reference bus, whose voltages are taken as a reference for the phasorial representation. The set of phases of node k and phases of line (l, m) are denoted by $\boldsymbol{\varphi}_k$ and $\boldsymbol{\varphi}_{lm}$, respectively. Let $v_{k,\phi} \in \mathbb{C}$ and $i_{k,\phi} \in \mathbb{C}$ denote the phasor for the line-to-ground voltage and the current at node k for phase ϕ, and define $\mathbf{v}_k := [v_{k,\phi}]_{\phi \in \boldsymbol{\varphi}_k}$ and $\mathbf{i}_k := [i_{k,\phi}]_{\phi \in \boldsymbol{\varphi}_k}$. For notational simplicity, the chapter hereafter focuses on three-phase systems; however, the proposed framework is applicable to systems featuring a variety of three-, two-, and single-phase nodes and branches.

Conventional fossil-fuel generators are assumed to be located at nodes $\mathcal{G} \subseteq \mathcal{N}$, with $P^{(G)}_{k,\phi}$, $Q^{(G)}_{k,\phi}$ denoting the active and reactive power generated at phase ϕ of bus $k \in \mathcal{G}$. The load connected to phase ϕ at bus k is denoted by $P^{(L)}_{k,\phi} + jQ^{(L)}_{k,\phi} \in \mathbb{C}$. In addition, the apparent power transferred from bus $l \in \mathcal{N}$ to the rest of the network through line $(l, m) \in \mathcal{L}$ for phase ϕ is given by $S_{lm,\phi} = P_{lm,\phi} + jQ_{lm,\phi}$.

Subset $\mathcal{R} \subset \mathcal{N}$ collects nodes with installed renewable energy sources (RESs) such as PV systems. Given prevailing ambient conditions, let the available active power from the RES located at phase ϕ of bus $k \in \mathcal{R}$ be denoted by $\overline{P}^{(R)}_{k,\phi}$. Also, let $P^{(R)}_{k,\phi}$ and $Q^{(R)}_{k,\phi}$ denote the injected active power and the injected/absorbed reactive power at bus k for phase ϕ. It is assumed that both active and reactive output-powers are controllable [24, 95]. Accordingly, the allowed operating region of an RES can be described as follows:

$$\boldsymbol{\Psi}_{k,\phi} := \left\{ P^{(R)}_{k,\phi}, Q^{(R)}_{k,\phi} : \begin{array}{c} 0 \le P^{(R)}_{k,\phi} \le \overline{P}^{(R)}_{k,\phi} \\ (P^{(R)}_{k,\phi})^2 + (Q^{(R)}_{k,\phi})^2 \le \overline{S}^2_{k,\phi} \\ |Q^{(R)}_{k,\phi}| \le \tan(\overline{\theta}_{k,\phi}) P^{(R)}_{k,\phi} \end{array} \right\} \tag{5.1}$$

where $\overline{S}^{(R)}_{k,\phi}$ represents the RES-inverter capacity, and $\overline{\theta}_{k,\phi}$ capture minimum power factor requirements.

Collect voltages and currents in the vectors $\mathbf{v} := [\mathbf{v}_0^T, \mathbf{v}_1^T, \cdots, \mathbf{v}_N^T]^T$ and $\mathbf{i} := [\mathbf{i}_0^T, \cdots, \mathbf{i}_N^T]^T$ of length $3(N+1)$, respectively. Lines $(l,m) \in \mathcal{L}$ are modeled as π-equivalent circuit, where the phase impedance and shunt admittance matrices are denoted by $\mathbf{Z}_{lm} \in \mathbb{C}^{|\boldsymbol{\varphi}_{lm}|\times|\boldsymbol{\varphi}_{lm}|}$ and $\overline{\mathbf{Y}}_{lm} \in \mathbb{C}^{|\boldsymbol{\varphi}_{lm}|\times\boldsymbol{\varphi}_{lm}}$, respectively. Voltages and injected currents abide by Ohm's law and Kirchhoff's law, which lead to the compact relationship $\mathbf{i} = \mathbf{Y}\mathbf{v}$. The network admittance matrix $\mathbf{Y}$ is hermitian, has dimensions $3(N+1) \times 3(N+1)$, and is constructed as follows [23, 83]:

- The $|\boldsymbol{\varphi}_{lm}| \times |\boldsymbol{\varphi}_{lm}|$ off-diagonal block corresponding to the line $(l,m) \in \mathcal{L}$ equals $-\mathbf{Y}_{lm} \equiv -\mathbf{Z}_{lm}^{-1}$.
- The $|\boldsymbol{\varphi}_k| \times |\boldsymbol{\varphi}_k|$ diagonal block corresponding to the k-th bus is given by

$$[\mathbf{Y}]_{k,k} = \sum_{l\in\mathcal{N}_k} (\frac{1}{2}\overline{\mathbf{Y}}_{kl} + \mathbf{Y}_{kl}) \tag{5.2}$$

  where $\mathcal{N}_k := \{l : (k,l) \in \mathcal{L}\}$.

The power balance equations at node $k \in \{\mathcal{G} \cap \mathcal{R}\}$ and phase $\phi \in \boldsymbol{\varphi}_k$ are given by:

$$P_{k,\phi}^{(G)} + P_{k,\phi}^{(R)} - P_{k,\phi}^{(L)} = \mathrm{Re}\{v_k^\phi (i_k^\phi)^*\}, \tag{5.3}$$

$$Q_{k,\phi}^{(G)} + Q_{k,\phi}^{(R)} - Q_{k,\phi}^{(L)} = \mathrm{Im}\{v_k^\phi (i_k^\phi)^*\}. \tag{5.4}$$

Notice that that the balance equation for nodes without conventional generators or without RESs can be readily derived from (5.3)–(5.4) by setting $P_{k,\phi}^{(G)} = Q_{k,\phi}^{(G)} = 0$, or $P_{k,\phi}^{(R)} = Q_{k,\phi}^{(R)} = 0$. Define the vectors $\mathbf{p}_G, \mathbf{q}_G$ which collect the active and reactive powers generated by conventional generators, and let $\mathbf{p}_R, \mathbf{q}_R$ be the vectors of active and reactive output-powers from RESs at all nodes for phases.

Accordingly, a prototypical formulation of the AC-OPF problem for a multi-phase

power network with renewables is outlined next:

$$\min_{\mathbf{v},\mathbf{i},\mathbf{p}_G,\mathbf{q}_G,\mathbf{p}_R,\mathbf{q}_R} C_g(\mathbf{p}_G) + C_c(\mathbf{p}_R) + C_i(\mathbf{q}_R) \tag{5.5a}$$

subject to

- $\forall k \in \mathcal{N}, \phi \in \boldsymbol{\varphi}_k$

$$P_{k,\phi}^{(G)} + P_{k,\phi}^{(R)} - P_{k,\phi}^{(L)} = \mathrm{Re}\{v_k^{\phi}(i_k^{\phi})^*\} \tag{5.5b}$$

$$Q_{k,\phi}^{(G)} + Q_{k,\phi}^{(R)} - Q_{k,\phi}^{(L)} = \mathrm{Im}\{v_k^{\phi}(i_k^{\phi})^*\} \tag{5.5c}$$

$$\underline{P}_{k,\phi}^{(G)} \leq P_{k,\phi}^{(G)} \leq \overline{P}_{k,\phi}^{(G)} \tag{5.5d}$$

$$\underline{Q}_{k,\phi}^{(G)} \leq Q_{k,\phi}^{(G)} \leq \overline{Q}_{k,\phi}^{(G)} \tag{5.5e}$$

$$|\underline{v}_{k,\phi}| \leq |v_{k,\phi}| \leq |\overline{v}_{k,\phi}| \tag{5.5f}$$

$$(P_{k,\phi}^{(R)}, Q_{k,\phi}^{(R)}) \in \boldsymbol{\Psi}_{k,\phi} \tag{5.5g}$$

where $\underline{P}_{k,\phi}^{(G)}$ and $\overline{P}_{k,\phi}^{(G)}$ are the lower and upper bound on the real power generated at bus k for phase ϕ; $\underline{Q}_{k,\phi}^{(G)}$ and $\overline{Q}_{k,\phi}^{(G)}$ represents an upper and lower bounds on the reactive power injected/absorbed by a conventional generation unit at node k for phase ϕ; and, the constraint (5.5f) confine the range of the voltage magnitude of the network buses within predefined limits. Notice that for buses $\mathcal{N}\backslash\mathcal{G}$, the limits $\underline{P}_{k,\phi}^{(G)}$, $\overline{P}_{k,\phi}^{(G)}$, $\underline{Q}_{k,\phi}^{(G)}$, and $\overline{Q}_{k,\phi}^{(G)}$ are set to zero. In addition, for nodes $\mathcal{N}\backslash\mathcal{R}$, one has $\overline{P}_{k,\phi}^{(R)} = \overline{S}_{k,\phi}^{(R)} = 0$. The cost function (5.5a) is composed of three functions:

- Cost from conventional generation units:

$$C_g(\mathbf{p}_G) = \sum_{k\in\mathcal{G},\phi\in\boldsymbol{\varphi}_k} b_{2,k}^{\phi}(P_{k,\phi}^{(G)})^2 + b_{1,k}^{\phi} P_{k,\phi}^{(G)} \tag{5.6}$$

- Cost of curtailment from renewables:

$$C_c(\mathbf{p}_R) = \sum_{k\in\mathcal{R},\phi\in\boldsymbol{\varphi}_k} c_{2,k}^{\phi}(\overline{P}_{k,\phi}^{(R)} - P_{k,\phi}^{(R)})^2 + c_{1,k}^{\phi}(\overline{P}_{k,\phi}^{(R)} - P_{k,\phi}^{(R)}) \tag{5.7}$$

- Cost of reactive support from renewables:

$$C_i(\mathbf{q}_R) = \sum_{k\in\mathcal{R},\phi\in\boldsymbol{\varphi}_k} d^{\phi}_{2,k}(Q^{(R)}_{k,\phi})^2 + d^{\phi}_{1,k}Q^{(R)}_{k,\phi} \tag{5.8}$$

Additional terms can be considered in the cost function to minimize e.g., power losses and other operational objectives.

In order to facilitate the use of the FPP-SCA algorithm, an equivalent formulation of (5.5) will be introduced next. To this end, define $\mathbf{e}^{\phi}_k := [\mathbf{0}_{\sum_{n=0}^{k-1}|\boldsymbol{\varphi}_n|} \quad \mathbf{e}_\phi \quad \mathbf{0}_{\sum_{n=k+1}^{N}|\boldsymbol{\varphi}_n|}]$, where $\mathbf{e}_\phi$ is the ϕ-th standard canonical basis is $\mathbb{R}^{|\boldsymbol{\varphi}_k|}$. Along the lines of [23], the following matrices are defined:

$$\mathbf{Y}_{k,\phi} = \tfrac{1}{2}(\mathbf{e}_{k,\phi}\mathbf{e}^T_{k,\phi}\mathbf{Y} + \mathbf{Y}^H\mathbf{e}_{k,\phi}\mathbf{e}^T_{k,\phi}), \tag{5.9}$$

$$\tilde{\mathbf{Y}}_{k,\phi} = \tfrac{j}{2}(\mathbf{e}_{k,\phi}\mathbf{e}^T_{k,\phi}\mathbf{Y} - \mathbf{Y}^H\mathbf{e}_{k,\phi}\mathbf{e}^T_{k,\phi}), \tag{5.10}$$

$$\mathbf{M}_{k,\phi} = \mathbf{e}_{k,\phi}\mathbf{e}^T_{k,\phi}. \tag{5.11}$$

Using (5.9) and (5.10), the right hand side of the power flow equations (5.3) and (5.4) can be expressed as

$$\mathrm{Re}\{v_{k,\phi}(i_{k,\phi})^*\} = \mathbf{v}^H\mathbf{Y}_{k,\phi}\mathbf{v}, \tag{5.12}$$

$$\mathrm{Im}\{v_{k,\phi}(i_{k,\phi})^*\} = \mathbf{v}^H\tilde{\mathbf{Y}}_{k,\phi}\mathbf{v}, \tag{5.13}$$

while the magnitude square of the voltage phasor at bus k and phase ϕ can be written in the following form:

$$|v_{k,\phi}|^2 = \mathbf{v}^H\mathbf{M}_{k,\phi}\mathbf{v}\,. \tag{5.14}$$

With these definitions, the AC OPF problem (5.5) can be re-written in the following

equivalent form:

$$\min_{\mathbf{v},\mathbf{p}_R,\mathbf{q}_r,\boldsymbol{\alpha}} C_g(\boldsymbol{\alpha}) + C_c(\mathbf{p}_R) + C_i(\mathbf{q}_R) \tag{5.15a}$$

subject to

- $\forall k \in \mathcal{N}, \phi \in \boldsymbol{\varphi}_k$

$$\mathbf{v}^H \mathbf{Y}_{k,\phi} \mathbf{v} - P_{k,\phi}^{(R)} + P_{k,\phi}^{(L)} \leq \alpha_{k,\phi} \tag{5.15b}$$

$$\underline{P}_{k,\phi}^{(G)} \leq \mathbf{v}^H \mathbf{Y}_{k,\phi} \mathbf{v} - P_{k,\phi}^{(R)} + P_{k,\phi}^{(L)} \leq \overline{P}_{k,\phi}^{(G)} \tag{5.15c}$$

$$\underline{Q}_{k,\phi}^{(G)} \leq \mathbf{v}^H \tilde{\mathbf{Y}}_{k,\phi} \mathbf{v} - Q_{k,\phi}^{(R)} + Q_{k,\phi}^{(L)} \leq \overline{Q}_{k,\phi}^{(G)} \tag{5.15d}$$

$$(|\underline{v}_{k,\phi}|)^2 \leq \mathbf{v}^H \mathbf{M}_{k,\phi} \mathbf{v} \leq (|\underline{v}_{k,\phi}|)^2 \tag{5.15e}$$

$$(P_{k,\phi}^{(R)}, Q_{k,\phi}^{(R)}) \in \boldsymbol{\Psi}_{k,\phi} \tag{5.15f}$$

where $\boldsymbol{\alpha}$ is a vector that collects all $\alpha_{k,\phi}$ for all $k \in \mathcal{G}$ and $\phi \in \boldsymbol{\varphi}_k$, and $\alpha_{k,\phi}$ represents a tight upper bound on the active power generated at node k for phase ϕ. Note that, for $k \notin \mathcal{G}$, the values of $\underline{P}_{k,\phi}^{(G)}$ and $\overline{P}_{k,\phi}^{(G)}$ are set to zero for all $\phi \in \boldsymbol{\varphi}_k$. Similarly, for $k \notin \mathcal{R}$, one has $\overline{P}_{k,\phi}^{(R)} = \overline{S}_{k,\phi}^{(R)} = 0$ for all $\phi \in \boldsymbol{\varphi}_k$. The problem (5.15) is a nonconvex QCQP. Accordingly, the FPP-SCA algorithm will be utilized in Section 5.3 to identify feasible solution of (5.15) in scenarios where existing convex relaxation-based methods may fail.

5.2 Modeling Delta-Connections

In this section, we extend the proposed framework in order to accommodate delta-connected loads as well as wye-connected loads. We show how the OPF problem with delta-connections can be cast as a nonconvex QCQP. Therefore, out FPP-SCA approach can be utilized to tackle the problem.

5.2.1 Flow Model for Radial Networks with Delta-Connections

Consider a radial multiphase distribution network with nodes collected in the set $\mathcal{N} = \{0, 1, \ldots, n\}$. Let 0 represent the substation or the point of common coupling, and define $\mathcal{N}^+ := \mathcal{N} \setminus \{0\}$. Let $\mathcal{E}$ denote the set of lines connecting the buses. In particular, each line connects an ordered pair (l, m) of buses, where bus l lies between bus 0 and bus

m. We use $(l,m) \in \mathcal{E}$ and $l \to m$ interchangeably, and denote $l \sim m$ if either $l \to m$ or $m \to l$. In addition, let $\mathcal{R} \subseteq \mathcal{N}^+$ represent a set that collects the buses with RESs.

For simplicity of notation and exposition, assume that all the buses $k \in \mathcal{N}$ and lines $(l,m) \in \mathcal{E}$ have three phases a, b, c, and define the sets $\Phi := \{a,b,c\}$ and $\Phi_\Delta := \{ab, bc, ca\}$. However, the proposed approach can be straightforwardly applied to distribution networks that feature a mix of three-phase, two-phase, and single-phase nodes. Let v_n^ϕ denote the complex voltage phasor at bus $k \in \mathcal{N}$ and $\phi \in \Phi$, and let $\mathbf{v}_k$ be a column vector that collects the complex voltages at all the phases for node k, i.e., $\mathbf{v}_k := [v_k^a,\ v_k^b,\ v_k^c]$. Similarly, let the vector $\mathbf{I}_{lm}$ collect the complex current that flows in the line $(l,m) \in \mathcal{E}$ for all phases in Φ. Let $y_k \in \mathbb{C}^{3\times 3}$ denote the shunt admittance at bus k, and denote as $z_{lm} \in \mathbb{C}^{3\times 3}$ the series impedance of line $l \sim m$.

Without loss of generality, assume that every bus $k \in \mathcal{N}$ has three wye-connected net loads (one on each phase, with grounded neutral) and three delta-connected net loads (one across each pair of phases, ungrounded). Define the complex vector $\mathbf{s}_{\mathrm{Y},k}^{(L)} \in \mathbb{C}^3$ to be a vector that collects the wye-connected loads at bus $k \in \mathcal{N}$ for all phases. In a similar way, define $\mathbf{s}_{\Delta,k}^{(L)} \in \mathbb{C}^3$ to be a complex column vector that collects the delta-connected loads at bus $k \in \mathcal{N}$. Notice that both wye and delta connection may be present at the same node of the network model when different distribution transformers with either delta and/or wye primary connections are bundled together for network reduction purposes (e.g., when two transformers are connected through a short low-impedance distribution line). Based on prevailing ambient conditions, let the maximum available active powers for the wye-connected RESs at phases $\phi \in \Phi$ of node $k \in \mathcal{R}$ be collected in a column vector $\overline{\mathbf{p}}_{\mathrm{Y},k}^{(R)} \in \mathbb{R}^3$. Also, let the apparent powers injected by these RESs be collected in a vector $\mathbf{s}_{\mathrm{Y},k}^{(R)} = \mathbf{p}_{\mathrm{Y},k}^{(R)} + i\mathbf{q}_{\mathrm{Y},k}^{(R)}$. Similarly, let $\overline{\mathbf{p}}_{\Delta,k}^{(R)} \in \mathbb{R}^3$ and $\mathbf{s}_{\Delta,k}^{(R)} \in \mathbb{C}^3$ be the corresponding vectors for the delta-connected sources at bus k. Let $\mathbf{s}_{\mathrm{Y},k} := [s_{\mathrm{Y},k}^a, s_{\mathrm{Y},k}^b, s_{\mathrm{Y},k}^c]^T = \mathbf{s}_{\mathrm{Y},k}^{(L)} - \mathbf{s}_{\mathrm{Y},k}^{(R)}$ denote the net wye-connected loads at bus k. Also, let $\mathbf{s}_{\Delta,k} := [s_{\Delta,k}^{ab},\ s_{\Delta,k}^{bc},\ s_{\Delta,k}^{ca}]^T = \mathbf{s}_{\Delta,k}^{(L)} - \mathbf{s}_{\Delta,k}^{(R)}$ and $\mathbf{I}_{\Delta,k} := [I_{\Delta,k}^{ab},\ I_{\Delta,k}^{bc},\ I_{\Delta,k}^{ca}]^T$ denote the power consumptions and currents of delta-connected net loads at bus k, respectively. If neither loads nor (sources) are present at a particular phase for a particular type of connection, then the corresponding element of $\mathbf{s}_{\mathrm{Y},k}^{(L)}$, $\mathbf{s}_{\mathrm{Y},k}^{(R)}$, $\mathbf{s}_{\Delta,k}^{(L)}$, or $\mathbf{s}_{\Delta,k}^{(R)}$ is set to zero.

Voltages and line currents abide by Ohm's Law, which yields:

$$\mathbf{v}_l - \mathbf{v}_m = z_{lm}\mathbf{I}_{lm}, \quad \forall l \to m. \tag{5.16}$$

In addition, the power flow equations for the delta-connected loads at bus k can be expressed as

$$\mathbf{s}_{\Delta,k} = \begin{bmatrix} (v_k^a - v_k^b)(I_{\Delta,k}^{ab})^* \\ (v_k^b - v_k^c)(I_{\Delta,k}^{bc})^* \\ (v_k^c - v_k^a)(I_{\Delta,k}^{ca})^* \end{bmatrix}, \quad \forall k \in \mathcal{N}. \tag{5.17}$$

Power balance at bus $k \in \mathcal{N}$ then implies that:

$$\begin{aligned} \sum_{i:i\to k} \begin{bmatrix} v_k^a (I_{ik}^a)^* \\ v_k^b (I_{ik}^b)^* \\ v_k^c (I_{ik}^c)^* \end{bmatrix} = \sum_{j:k\to j} \begin{bmatrix} v_k^a (I_{kj}^a)^* \\ v_k^b (I_{kj}^b)^* \\ v_k^c (I_{kj}^c)^* \end{bmatrix} + \begin{bmatrix} v_k^a (y_k^a \mathbf{v}_k)^* \\ v_k^b (y_k^b \mathbf{v}_k)^* \\ v_k^c (y_k^c \mathbf{v}_k)^* \end{bmatrix} + \mathbf{s}_{\mathrm{Y},k} \\ + \begin{bmatrix} v_k^a (I_{\Delta,k}^{ab} - I_{\Delta,k}^{ca})^* \\ v_k^b (I_{\Delta,k}^{bc} - I_{\Delta,k}^{ab})^* \\ v_k^c (I_{\Delta,k}^{ca} - I_{\Delta,k}^{bc})^* \end{bmatrix}. \end{aligned} \tag{5.18}$$

Recalling that the network is assumed to have a tree tolopogy, the left-hand-side of (5.18) represents the power received by bus k from the rest of the network through the distribution line $(i,k) \in \mathcal{E}$. On the other hand, the first term on the right-hand-side represents the power transferred to the network through line $(k,j) \in \mathcal{E}$; the second term accounts for the power drawn to the ground at every phase through the shunt element; the third term represents the apparent power absorbed/generated by the wye-connected loads at the bus; and, the last term represents delta-connected net loads. Notice that the power flow equations (5.17) and (5.18) have quadratic terms; using these equations within an optimization task leads to nonconvex problem formulations.

5.2.2 QCQP Formulation of OPF for Networks with Delta-Connections

To facilitate the application of the FPP-SCA algorithm to the problem at hand, the ACOPF is presented next in a QCQP form. To this end, define the vector $\mathbf{x}_{lm} := [\mathbf{v}_l^T, \mathbf{I}_{lm}^T]^T$ for all $l \to m$, and $\mathbf{x}_{\Delta,k} := [\mathbf{v}_k^T, \mathbf{I}_{\Delta,k}^T]^T$ for all $k \in \mathcal{N}$. Consider matrices

$E_{ij} \in \mathbb{R}^{6\times 6}$ and $\tilde{E}_{ij} \in \mathbb{R}^{3\times 3}$ such that their (i,j)-th element is one (1) and all the other elements are zero (0). Then, (5.17) can be written as:

$$\mathbf{s}_{\Delta,k} = \begin{bmatrix} \mathbf{x}_{\Delta,k}^H \left(E_{41} - E_{42}\right) \mathbf{x}_{\Delta,k} \\ \mathbf{x}_{\Delta,k}^H \left(E_{52} - E_{53}\right) \mathbf{x}_{\Delta,k} \\ \mathbf{x}_{\Delta,k}^H \left(E_{63} - E_{61}\right) \mathbf{x}_{\Delta,k} \end{bmatrix}, \quad \forall k \in \mathcal{N}. \tag{5.19}$$

In (5.18), the term $v_k^a (I_{kj}^a)^*$ can be re-written as

$$v_k^a (I_{kj}^a)^* = \mathbf{x}_{kj}^H E_{41} \mathbf{x}_{kj}.$$

In a similar way, the product $v_k^a (I_{ik}^a)^*$ can be expanded as shown next:

$$\begin{aligned} v_k^a (I_{ik}^a)^* &= v_i^a (I_{ik}^a)^* - (v_i^a - v_k^a)(I_{ik}^a)^* \\ &= v_i^a (I_{ik}^a)^* - z_{ik}^a \mathbf{I}_{ik} (I_{ik}^a)^* \\ &= \mathbf{x}_{ik}^H \left(E_{41} - z_{ik}^{aa} E_{44} - z_{ik}^{ab} E_{45} - z_{ik}^{ac} E_{46} \right) \mathbf{x}_{ik} \end{aligned}$$

where $z_{ik}^a = [z_{ik}^{aa}, z_{ik}^{ab}, z_{ik}^{ac}]$ is the first row of the impedance matrix z_{ik}. Moreover, we have that

$$V_k^a (y_k^a V_k)^* = \mathbf{v}_k^H \left(y_k^{aa,*} \tilde{E}_{11} + y_k^{ab,*} \tilde{E}_{21} + y_k^{ac,*} \tilde{E}_{31} \right) \mathbf{v}_k.$$

Additionally, for delta connected units, the term $v_k^a (I_{\Delta,k}^{ab} - I_{\Delta,k}^{ca})^*$ can be re-written as

$$v_k^a (I_{\Delta,k}^{ab} - I_{\Delta,k}^{ca})^* = \mathbf{x}_{\Delta,k}^H \left(E_{41} - E_{61}\right) \mathbf{x}_{\Delta,k}.$$

Following similar steps, quadratic expressions of all the other elements in (5.18) can be obtained and, consequently, (5.18) can be written as follows:

$$\begin{aligned}
&\sum_{i:i\to k}\begin{bmatrix}\mathbf{x}_{ik}^H\left(E_{41}-z_{ik}^{aa}E_{44}-z_{ik}^{ab}E_{45}-z_{ik}^{ac}E_{46}\right)\mathbf{x}_{ki}\\ \mathbf{x}_{ik}^H\left(E_{52}-z_{ik}^{ba}E_{54}-z_{ik}^{bb}E_{55}-z_{ik}^{bc}E_{56}\right)\mathbf{x}_{ki}\\ \mathbf{x}_{ik}^H\left(E_{63}-z_{ik}^{ca}E_{64}-z_{ik}^{cb}E_{65}-z_{ik}^{cc}E_{66}\right)\mathbf{x}_{ki}\end{bmatrix}\\
=&\sum_{j:k\to j}\begin{bmatrix}\mathbf{x}_{kj}^H E_{41}\mathbf{x}_{kj}\\ \mathbf{x}_{kj}^H E_{52}\mathbf{x}_{kj}\\ \mathbf{x}_{kj}^H E_{63}\mathbf{x}_{kj}\end{bmatrix}+\begin{bmatrix}\mathbf{x}_{\Delta,k}^H\left(E_{41}-E_{61}\right)\mathbf{x}_{\Delta,k}\\ \mathbf{x}_{\Delta,k}^H\left(E_{52}-E_{42}\right)\mathbf{x}_{\Delta,k}\\ \mathbf{x}_{\Delta,k}^H\left(E_{63}-E_{53}\right)\mathbf{x}_{\Delta,k}\end{bmatrix}\\
&+\mathbf{s}_{\mathrm{Y},k}+\begin{bmatrix}\mathbf{v}_k^H\left(y_k^{aa,*}\tilde{E}_{11}+y_k^{ab,*}\tilde{E}_{21}+y_k^{ac,*}\tilde{E}_{31}\right)\mathbf{v}_k\\ \mathbf{v}_k^H\left(y_k^{ba,*}\tilde{E}_{12}+y_k^{bb,*}\tilde{E}_{22}+y_k^{bc,*}\tilde{E}_{32}\right)\mathbf{v}_k\\ \mathbf{v}_k^H\left(y_k^{ca,*}\tilde{E}_{13}+y_k^{cb,*}\tilde{E}_{23}+y_k^{cc,*}\tilde{E}_{33}\right)\mathbf{v}_k\end{bmatrix}
\end{aligned}\tag{5.20}$$

for all $k\in\mathcal{N}$, while the voltage magnitude constraints admit the following equivalent formulation:

$$(\underline{v}_k^{\phi})^2\le\mathbf{v}_k^H\tilde{E}_{\phi\phi}\mathbf{v}_k\le(\overline{v}_k^{\phi})^2,\quad\forall k\in\mathcal{N}^+,\ \forall\phi\in\Phi.\tag{5.21}$$

Next, let the vector $\mathbf{x}_v\in\mathbb{C}^{3(n+1)}$ collect $\mathbf{v}_k$ for all $k\in\mathcal{N}$. Similarly, let $\mathbf{x}_\Delta\in\mathbb{C}^{3(n+1)}$ be a vector that concatenates $\mathbf{I}_{\Delta,k}$ for all $k\in\mathcal{N}$ and let $\mathbf{x}_i\in\mathbb{C}^{3n}$ stack $\mathbf{I}_{lm}$ for all $(l,m)\in\mathcal{E}$. Define $\tilde{\mathbf{x}}$ as

$$\tilde{\mathbf{x}}=[\mathbf{x}_v^T,\ \mathbf{x}_i^T,\ \mathbf{x}_\Delta^T]^T.\tag{5.22}$$

and consider the stacked vector of its real and imaginary parts $\mathbf{x}:=[\mathrm{Re}(\tilde{\mathbf{x}})^T\,\mathrm{Im}(\tilde{\mathbf{x}})^T]^T$; notice that $\mathbf{x}$ is a rwal vector containing $(18n+12)$ elements. The quadratic functions (5.19) and (5.20) can be written as functions of $\mathbf{x}$. Particularly, (5.20) involves $6(n+1)$ equations (involving real quantities) while $6(n+1)$ equations are utilized to describe (5.19)

The OPF problem for networks with delta- and wye-connected components can then

be written as the following QCQP form:

$$\min \quad C_{loss}(p_{loss}) + \sum_{k\in\mathcal{R}} C_k(\mathbf{p}^{(c)}_{\mathrm{Y},k}, \mathbf{p}^{(c)}_{\Delta,k}) \tag{5.23a}$$

$$\text{over} \quad \mathbf{x}, \{\mathbf{s}^{(R)}_{\mathrm{Y},k}, \mathbf{s}^{(R)}_{\Delta,k}\}_{k\in\mathcal{R}}$$

$$\text{s.t.} \quad \mathbf{A}\mathbf{x} = \mathbf{d} \tag{5.23b}$$

$$\mathbf{x}^T\mathbf{B}^{\phi}_{\Delta,k}\mathbf{x} = \Re\{s^{\phi}_{\Delta,k}\} \quad \forall\phi\in\Phi_{\Delta}, \forall k\in\mathcal{N} \tag{5.23c}$$

$$\mathbf{x}^T\tilde{\mathbf{B}}^{\phi}_{\Delta,k}\mathbf{x} = \Im\{s^{\phi}_{\Delta,k}\} \quad \forall\phi\in\Phi_{\Delta}, \forall k\in\mathcal{N} \tag{5.23d}$$

$$\mathbf{x}^T\mathbf{B}^{\phi}_{\mathrm{Y},k}\mathbf{x} = \Re\{s^{\phi}_{\mathrm{Y},k}\} \quad \forall\phi\in\Phi, \forall k\in\mathcal{N} \tag{5.23e}$$

$$\mathbf{x}^T\tilde{\mathbf{B}}^{\phi}_{\mathrm{Y},k}\mathbf{x} = \Im\{s^{\phi}_{\mathrm{Y},k}\} \quad \forall\phi\in\Phi, \forall k\in\mathcal{N} \tag{5.23f}$$

$$(\underline{v}^{\phi}_k)^2 \le \mathbf{x}^T\mathbf{M}^{\phi}_k\mathbf{x} \le (\overline{v}^{\phi}_k)^2 \quad \forall\phi\in\Phi, \forall k\in\mathcal{N}^+ \tag{5.23g}$$

$$\mathbf{s}_{\Delta,k} = \mathbf{s}^{(L)}_{\Delta,k} - \mathbf{s}^{(R)}_{\Delta,k} \quad \forall k\in\mathcal{N} \tag{5.23h}$$

$$\mathbf{s}_{\mathrm{Y},k} = \mathbf{s}^{(L)}_{\mathrm{Y},k} - \mathbf{s}^{(R)}_{\mathrm{Y},k} \quad \forall k\in\mathcal{N} \tag{5.23i}$$

$$\mathbf{p}^{(R)}_{\mathrm{Y},k} \le \overline{\mathbf{p}}^{(R)}_{\mathrm{Y},k} \quad \forall k\in\mathcal{R} \tag{5.23j}$$

$$\mathbf{p}^{(R)}_{\Delta,k} \le \overline{\mathbf{p}}^{(R)}_{\Delta,k} \quad \forall k\in\mathcal{R} \tag{5.23k}$$

where $\mathbf{B}^{\phi}_{\Delta,k}$, $\tilde{\mathbf{B}}^{\phi}_{\Delta,k}$, $\mathbf{B}^{\phi}_{\mathrm{Y},k}$, and $\tilde{\mathbf{B}}^{\phi}_{\mathrm{Y},k}$ are symmetric matrices that represents the real and imaginary parts of equation (5.19) and (5.20), respectively, and their construction is explained in Appendix C. Notice that the values of $\mathbf{s}^{(R)}_{\Delta,k}$ for $k\notin\mathcal{R}$ are set to zero, and hence, they are not considered as optimization variables. The matrix $\mathbf{A}$ and the vector $\mathbf{d}$ are constructed in a way to rewrite the constraints (5.16) and constant phase at the substation in terms of $\mathbf{x}$.

5.3 Feasible Point Pursuit and Successive Convex Approximation Algorithm

The FPP-SCA is a two-step algorithm that involves solving convex optimization problems iteratively. In the first step, we solve an inner approximation of (5.15) around a particular point. In order to ensure feasibility of the approximation, we add a slack variable s to the constraints and minimize s over the approximated feasible set. We then use the solution as an approximation point for the next step. If the slack variable

becomes zero, we get a feasible point. In the second step, we solve a sequence of problems which are inner approximations of (5.15) around feasible points until convergence to a KKT point.

5.3.1 Feasible Point Pursuit

In each iteration, the non-convex feasiblity set of (5.15) is replaced by a convex inner approximation. Each non-convex quadratic constraint is replaced by a convex restriction around a specific point. For instance, consider the constraint (5.15c) which can be written as two inequalities in the following form.

$$\mathbf{v}^H \mathbf{Y}_{k,\phi} \mathbf{v} \leq -P_{k,\phi}^{(L)} + P_{k,\phi}^{(R)} + \overline{P}_{k,\phi}^{(G)}, \tag{5.24a}$$

$$\mathbf{v}^H (-\mathbf{Y}_{k,\phi}) \mathbf{v} \leq P_{k,\phi}^{(L)} - P_{k,\phi}^{(R)} - \underline{P}_{k,\phi}^{(G)}. \tag{5.24b}$$

Both constraints are non-convex as the matrices $\mathbf{Y}_k^{\phi}$ are indefinite. Consider (5.24a) where the inequality can be rewritten as

$$\mathbf{v}^H \mathbf{Y}_{k,\phi}^{(+)} \mathbf{v} + \mathbf{v}^H \mathbf{Y}_{k,\phi}^{(-)} \mathbf{v} \leq -P_{k,\phi}^{(L)} + P_{k,\phi}^{(R)} + \overline{P}_{k,\phi}^{(G)} \tag{5.25}$$

where $\mathbf{Y}_{k,\phi}^{(+)}$ and $\mathbf{Y}_{k,\phi}^{(-)}$ are the positive semidefinite and the negative semidefinite parts of the matrix $\mathbf{Y}_{k,\phi}$, respectively. For $\mathbf{Y}_{k,\phi}^{(-)}$, the following inequality holds.

$$(\mathbf{v} - \mathbf{z})^H \mathbf{Y}_{k,\phi}^{(-)} (\mathbf{v} - \mathbf{z}) \leq 0. \tag{5.26}$$

Then, expanding the left hand side, the following inequality can be obtained

$$\mathbf{v}^H \mathbf{Y}_{k,\phi}^{(-)} \mathbf{v} \leq 2\mathbf{z}^H \mathbf{Y}_{k,\phi}^{(-)} \mathbf{v} - \mathbf{z}^H \mathbf{Y}_{k,\phi}^{(-)} \mathbf{z}. \tag{5.27}$$

Hence, the surrogate function for the non-convex quadratic constraint (5.25) can be defined as

$$\mathbf{v}^H \mathbf{Y}_{k,\phi}^{(+)} \mathbf{v} + 2\mathbf{z}^H \mathbf{Y}_{k,\phi}^{(-)} \mathbf{v} \leq -P_{k,\phi}^{(L)} + P_{k,\phi}^{(R)} + \overline{P}_{k,\phi}^{(G)} + \mathbf{z}^H \mathbf{Y}_{k,\phi}^{(-)} \mathbf{z} + s \tag{5.28}$$

where the nonnegative slack variable s is added to ensure feasibility. Similarly, (5.24b) is replaced by

$$-\mathbf{v}^H\mathbf{Y}_{k,\phi}^{(-)}\mathbf{v} - 2\mathbf{z}^H\mathbf{Y}_{k,\phi}^{(+)}\mathbf{v} \leq P_{k,\phi}^{(L)} - P_{k,\phi}^{(R)} - \underline{P}_{k,\phi}^{(G)} - \mathbf{z}^H\mathbf{Y}_{k,\phi}^{(+)}\mathbf{z} + s. \tag{5.29}$$

The problem to be solved in the i-th iteration can then be written as follows, where $\mathbf{z}_i$ is the optimum $\mathbf{v}$ obtained in iteration $i-1$:

$$\min_{\mathbf{v},\mathbf{p}_R,\mathbf{q}_R,s\geq 0} \quad s \tag{5.30a}$$

subject to

- $\forall k \in \mathcal{N},\ \forall \phi \in \boldsymbol{\varphi}_k$

$$\mathbf{v}^H\mathbf{Y}_{k,\phi}^{(+)}\mathbf{v} + 2\mathbf{z}_i^H\mathbf{Y}_{k,\phi}^{(-)}\mathbf{v} \leq$$
$$-P_{k,\phi}^{(L)} + P_{k,\phi}^{(R)} + \overline{P}_{k,\phi}^{(G)} + \mathbf{z}_i^H\mathbf{Y}_{k,\phi}^{(-)}\mathbf{z}_i + s \tag{5.30b}$$
$$\mathbf{v}^H(-\mathbf{Y}_{k,\phi}^{(-)})\mathbf{v} - 2\mathbf{z}_i^H\mathbf{Y}_{k,\phi}^{(+)}\mathbf{v} \leq$$
$$P_{k,\phi}^{(L)} - P_{k,\phi}^{(R)} - \underline{P}_{k,\phi}^{(G)} - \mathbf{z}_i^H\mathbf{Y}_{k,\phi}^{(+)}\mathbf{z}_i + s \tag{5.30c}$$
$$\mathbf{v}^H\tilde{\mathbf{Y}}_{k,\phi}^{(+)}\mathbf{v} + 2\mathbf{z}_i^H\tilde{\mathbf{Y}}_{k,\phi}^{(-)}\mathbf{v} \leq$$
$$-Q_{k,\phi}^{(L)} + Q_{k,\phi}^{(R)} + \overline{Q}_{k,\phi}^{(G)} + \mathbf{z}_i^H\tilde{\mathbf{Y}}_{k,\phi}^{(-)}\mathbf{z}_i + s \tag{5.30d}$$
$$\mathbf{v}^H(-\tilde{\mathbf{Y}}_{k,\phi}^{(-)})\mathbf{v} - 2\mathbf{z}_i^H\tilde{\mathbf{Y}}_{k,\phi}^{(+)}\mathbf{v} \leq$$
$$Q_{k,\phi}^{(L)} - Q_{k,\phi}^{(R)} - \underline{Q}_{k,\phi}^{(G)} - \mathbf{z}_i^H\tilde{\mathbf{Y}}_{k,\phi}^{(+)}\mathbf{z}_i + s \tag{5.30e}$$
$$\mathbf{v}^H\mathbf{M}_{k,\phi}\mathbf{v} \leq |\overline{v}_{k,\phi}|^2 + s \tag{5.30f}$$
$$2\mathbf{z}_i^H(-\mathbf{M}_{k,\phi})\mathbf{v} \leq -|\underline{v}_{k,\phi}|^2 + \mathbf{z}_i^H(-\mathbf{M}_{k,\phi})\mathbf{z}_i + s \tag{5.30g}$$
$$(P_{k,\phi}^{(R)}, Q_{k,\phi}^{(R)}) \in \boldsymbol{\Psi}_{k,\phi} \tag{5.30h}$$

The optimization problem (5.30) can be cast as SOCP which can be solved efficiently in polynomial time. Each problem instance is feasible due to the positive slack variable. This *feasible point pursuit* is summarized in Algorithm 2.

It is clear that the value of s is nonincreasing with i as $\mathbf{z}_i(\mathbf{v}_{i-1})$ is always feasible while solving (5.30). Despite the fact that this method in not guaranteed to find a feasible point, it always converges in the simulations to a voltage profile given by $\mathbf{v}_f$ that is feasible. Therefore, $\mathbf{v}_f$ is used as a starting point for the second part of our

Algorithm 2: Feasible Point Pursuit Algorithm

Initialization: *set* $i = 0$, and choose $\mathbf{z}_0$ to be the flat voltage profile.
repeat
  $\mathbf{v}_i, s \leftarrow$ solution of (5.30).
  $\mathbf{z}_{i+1} \leftarrow \mathbf{v}_i$.
  $i \leftarrow i + 1$.
until $s < \epsilon_1$ *or* $||\mathbf{v}_i - \mathbf{v}_{i-1}|| \leq \epsilon_1$
Output: $\mathbf{v}_f \leftarrow \mathbf{v}_i$

algorithm (SCA).

5.3.2 Successive Convex Approximation

Starting from a feasible point, the nonconvex feasible set is replaced at each iteration by an inner convex approximation. Similar to the FPP phase, the surrogates are formulated as convex upper bounds for the nonconvex parts of the quadratic constraints. Consequently, a monotone sequence that converges to a KKT point of the original problem (5.5) is generated. In each iteration, the following problem is solved

$$\min_{\mathbf{v},\boldsymbol{\alpha},\mathbf{p}_R,\mathbf{q}_R} \quad C_g(\boldsymbol{\alpha}) + C_c(\mathbf{p}_R) + C_i(\mathbf{q}_R) \tag{5.31a}$$

subject to

$$\text{(5.30b)} - \text{(5.30h)} \quad (\text{with } s \text{ removed} \Leftrightarrow s \text{ set to } 0)$$

$\bullet\ \forall k \in \mathcal{N},\ \forall \phi \in \boldsymbol{\varphi}_k$

$$\mathbf{v}^H \mathbf{Y}_{k,\phi}{}^{(+)} \mathbf{v} + 2\mathbf{z}_i^H \mathbf{Y}_{k,\phi}{}^{(-)} \mathbf{v} \leq - P_{k,\phi}^{(L)} + P_{k,\phi}^{(R)} + \mathbf{z}_i^H \mathbf{Y}_{k,\phi}{}^{(-)} \mathbf{z}_i + \alpha_{k,\phi} \tag{5.31b}$$

Note that, since the starting point is feasible, we do not add s to the surrogate constraints, or equivalently, the value of s is set to be zero. Therefore, the generated sequence is always feasible and the cost function is nonincreasing with the iterates. Algorithm 3 describes the steps of the SCA phase.

Claim 3 (Convergence)**.** *From [82, Theorem 1], it can be shown that every limit point generated using the proposed algorithms is a KKT point. Hence, the first phase converges*

Algorithm 3: Successive Convex Approximation Algorithm

Initialization: *set* $i = 0$, and $\mathbf{z}_0 = \mathbf{v}_f$.
repeat
  $\mathbf{v}_i \leftarrow$ solution of (5.31).
  $\mathbf{z}_{i+1} \leftarrow \mathbf{v}_i$.
  $i \leftarrow i + 1$.
until $\frac{\mathbf{v}_{i-1} - \mathbf{v}_i}{\mathbf{v}_{i-1}} < \epsilon_2$
Output: $\mathbf{v}_{opt} \leftarrow \mathbf{v}_i$

to a KKT point of (5.30)*. In addition, if we start the second phase from a feasible initialization, then the whole sequence generated will converge to the set containing all the KKT points of the OPF problem* (5.5)*.*

The first part of the claim follows directly from [82]. For the second phase, if the initialization point is feasible, then the whole generated sequence will lie in the feasibility set. Because the feasible set is compact, i.e., closed and bounded, the whole converging sequence will go to the set that comprises all the KKT points of (5.5). Note though that a KKT point of (5.30) is not guaranteed to be a feasible point of (5.5) – in fact [68] contains a counter-example – however our experience is that a feasible point is generated with high probability, if one exists (always the case in our OPF experiments).

5.3.3 Identifying Problematic Constraints

The AC OPF problem may be infeasible under a number of operational settings, where the demand cannot be satisfied without violating voltage and/or flow constraints. When convex relaxation of the OPF problem is infeasible, it provides an infeasibility certificate for the original (nonconvex) problem. However, such relaxations typically cannot provide informative feedback on the problematic constraints – something valuable to the network operator to take corrective actions. Off-the-shelf solvers such as IPOPT cannot identify the problematic constraints either.

The FPP-SCA method (5.30) seeks a feasible operating point in the first phase by minimizing the slack variable. The value of the slack variable at each iteration is in fact related to the maximum constraint violation. This method can be suitably modified to enable network operators to identify the constraints that render the overall OPF

infeasible. Particularly, consider associating a slack variable with each constraint, and minimizing a cost function that is strictly increasing in the slack variables. Specifically, consider replacing problem (5.30) with the following one:

$$\min_{\mathbf{v},\mathbf{p}_R,\mathbf{q}_R,s\geq 0} \quad \|\boldsymbol{s}\|_2^2 \tag{5.32a}$$

subject to

- $\forall k \in \mathcal{N},\ \forall \phi \in \boldsymbol{\varphi}_k$

$$\mathbf{v}^H \mathbf{Y}_{k,\phi}^{(+)} \mathbf{v} + 2\mathbf{z}_i^H \mathbf{Y}_{k,\phi}^{(-)} \mathbf{v} \leq$$
$$- P_{k,\phi}^{(L)} + P_{k,\phi}^{(R)} + \overline{P}_{k,\phi}^{(G)} + \mathbf{z}_i^H \mathbf{Y}_{k,\phi}^{(-)} \mathbf{z}_i + s_{k,\phi}^{\overline{P}} \tag{5.32b}$$
$$\mathbf{v}^H (-\mathbf{Y}_{k,\phi}^{(-)}) \mathbf{v} - 2\mathbf{z}_i^H \mathbf{Y}_{k,\phi}^{(+)} \mathbf{v} \leq$$
$$P_{k,\phi}^{(L)} - P_{k,\phi}^{(R)} - \underline{P}_{k,\phi}^{(G)} - \mathbf{z}_i^H \mathbf{Y}_{k,\phi}^{(+)} \mathbf{z}_i + s_{k,\phi}^{\underline{P}} \tag{5.32c}$$
$$\mathbf{v}^H \tilde{\mathbf{Y}}_{k,\phi}^{(+)} \mathbf{v} + 2\mathbf{z}_i^H \tilde{\mathbf{Y}}_{k,\phi}^{(-)} \mathbf{v} \leq$$
$$- Q_{k,\phi}^{(L)} + Q_{k,\phi}^{(R)} + \overline{Q}_{k,\phi}^{(G)} + \mathbf{z}_i^H \tilde{\mathbf{Y}}_{k,\phi}^{(-)} \mathbf{z}_i + s_{k,\phi}^{\overline{Q}} \tag{5.32d}$$
$$\mathbf{v}^H (-\tilde{\mathbf{Y}}_{k,\phi}^{(-)}) \mathbf{v} - 2\mathbf{z}_i^H \tilde{\mathbf{Y}}_{k,\phi}^{(+)} \mathbf{v} \leq$$
$$Q_{k,\phi}^{(L)} - Q_{k,\phi}^{(R)} - \underline{Q}_{k,\phi}^{(G)} - \mathbf{z}_i^H \tilde{\mathbf{Y}}_{k,\phi}^{(+)} \mathbf{z}_i + s_{k,\phi}^{\underline{Q}} \tag{5.32e}$$
$$\mathbf{v}^H \mathbf{M}_{k,\phi} \mathbf{v} \leq |\overline{v}_{k,\phi}|^2 + s_{k,\phi}^{\overline{V}} \tag{5.32f}$$
$$2\mathbf{z}_i^H (-\mathbf{M}_{k,\phi}) \mathbf{v} \leq -|\underline{v}_{k,\phi}|^2 + \mathbf{z}_i^H (-\mathbf{M}_{k,\phi}) \mathbf{z}_i + s_{k,\phi}^{\underline{V}} \tag{5.32g}$$
$$(P_{k,\phi}^{(R)}, Q_{k,\phi}^{(R)}) \in \boldsymbol{\Psi}_{k,\phi} \tag{5.32h}$$

where $\boldsymbol{s}$ is a vector collecting all the slack variables. It is clear that, in this setting, the individual value of each slack relates to the violation of the respective constraint.

Let $\mathbf{v}_i$, $\mathbf{p}_i^{(R)}$, $\mathbf{q}_i^{(R)}$ and $\mathbf{s}_i$ denote the solution of (5.32) at the i-th iteration of the FPP algorithm. Then, using [82], one can easily prove that the sequence generated by solving (5.32) iteratively is convergent. When $\mathbf{s}_i$ is all zeros at the i-th iteration, the corresponding $\boldsymbol{v}_i$ is a feasible solution for the original problem. On the other hand, if the problem is infeasible, then the slacks will converge to a non-zero vector and the positive elements of $\mathbf{s}$ will provide a pointer to the constraints that cannot be satisfied.

Notice that replacing the 2-norm in the cost function (5.32a) by $\|\mathbf{s}\|_\infty$ yields an optimization problem that is equivalent to (5.30).

5.4 Test Cases and Results

To demonstrate the efficacy of the proposed algorithm, four scenarios where convex relaxation techniques and existing solvers for nonlinear (nonconvex) programs are not able to reveal feasible solutions will be considered. In the first case, we consider a single-phase equivalent model for a distribution system with high PV penetration. The ability of the proposed algorithm to minimize the curtailed power while respecting the network operational constraints will be demonstrated. Then, the three-phase model of the same distribution system will be presented. Finally, several transmission systems will be used to show the ability of the FPP-SCA algorithm to solve challenging OPF problem instances where other methods fail to find feasible voltage profiles. In addition, the ability of the proposed algorithm to identify constraints that render the OPF problem infeasible will be demonstrated. Finally, we will present the results of the FPP-SCA algorithm when utilized for solving ACOPF problem for radial networks with delta-connected loads.

The proposed algorithm and the SDR one both employ the MATLAB-based optimization modeling package YALMIP [62] along with the interior-point solver SeDuMi [94] on an Intel CPU @ 3.5 GHz (16 GB RAM) computer. For the IPOPT solver, A Julia/JuMP[1] Package for Power Network Optimization[2] was adopted to solve the single-phase OPF problems for transmission systems. We initialize our algorithm with the flat voltage profile. In addition, we choose the values of ϵ_1 and ϵ_2 to be 10^{-11} and 10^{-5}, respectively.

5.4.1 Results for Systems with Only Wye-Connections

Distribution systems with high PV penetration are likely to experience overvoltage challenges. The ability to curtail active power generated by the renewables has been shown to reliably prevent overvoltages and maintain the system operational constraints. In the first scenario, a modified version of the IEEE 37-node test feeder, shown in Fig. 5.1, is considered. The model is constructed by considering a single-phase equivalent feeder. Real load data measured from feeders in Anatolia, CA in August 2012 [7] are

[1] [Online] http://julialang.org/.

[2] A Julia/JuMP Package for Power Network Optimization. [Online] https://github.com/lanl-ansi/PowerModels.jl.

used. The PV inverters are assumed to be located at the red nodes in Fig. 5.1, and their generation profiles are based on the real irradiance data available in [7].

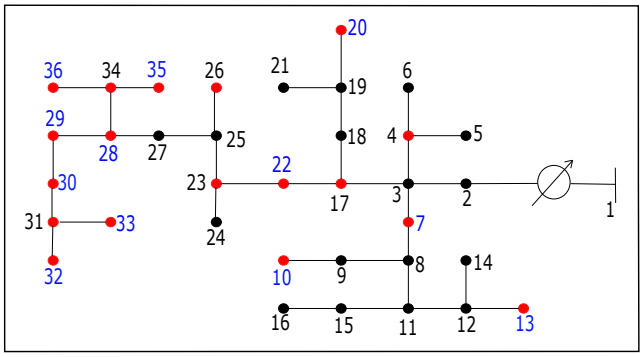


Figure 5.1: IEEE 37-node test feeder. The red nodes are the nodes with PV inverters in the single-phase model. The nodes with PV units installed in the three-phase case are indexed in blue.

In order to show the efficacy of the proposed algorithm, two different load and irradiance profiles are considered. The first profile is taken at 1 : 00 PM, where the available power from PV inverters exceeds the demand. Then, the load and irradiance data at 7 : 00 PM in considered, where the PV inverters have very low active power. In both cases, the values of $\overline{S}_{k,\phi}^{(R)}$ are set to be $2\overline{P}_{k,\phi}^{(R)}$, and the values of $\overline{\theta}_{k,\phi}$ are set such that the minimum power factor is 0.7 for all the PV units. The limits of the voltage magnitudes $\overline{v}_{k,\phi}$ and $\underline{v}_{k,\phi}$ are set to be 1.05 and 0.95, respectively. Additionally, the cost function is determined by setting $b_{2,k}^{\phi} = 0.1$, $c_{2,k}^{\phi} = 1$, $d_{2,k}^{\phi} = 0.5$, and $b_{1,k}^{\phi} = c_{1,k}^{\phi} = d_{1,k}^{\phi} = 0$ for all $k \in \mathcal{N}$ and $\phi \in \boldsymbol{\varphi}_k$.

Table 5.1 shows that the FPP-SCA algorithm is able to find a feasible voltage profile in both situations, while SDR is not able to find a meaningful solution when the PV penetration is high. The voltage profiles produced by the FPP-SCA and the SDR are shown in Fig. 5.2 in the case of low irradiance. From the depicted voltage profile, active power is drawn from node-1 to achieve the load demand at this moment. On the other hand, the voltage profile given by the proposed algorithm at 1 : 00 PM is shown in Fig. 5.3, where the excess of the active power generated by the PV is delivered to the transmission system connected at node-1. Table 5.2 lists the amount of the available power at each PV unit at 1 : 00 PM, as well as the curtailed active power resulted from the FPP-SCA solution and the injected/absorbed reactive power. Note that the power

Times	**SDR**		**FPP-SCA**	
	Feasibility	Cost	Feasibility	Cost
1 : 00 PM	times	–	✓	70486
7 : 00 PM	✓	35157	✓	35183

Table 5.1: Comparison between the FPP-SCA algorithm and the SDR.

factor constraint is achieved with equality at all the PV units.

k	$\overline{P}_{k,1}^{(R)}$	$(\overline{P}_{k,1}^{(R)} - P_{k,1}^{(R)})$	$Q_{k,1}^{(R)}$
4	98.60	5.82	−11.46
7	98.60	5.81	−11.46
10	98.60	5.78	−11.50
13	197.20	5.71	−11.57
17	197.20	9.15	−18.53
20	197.20	9.11	−18.58
22	197.20	12.06	−22.92
23	197.20	13.02	−24.38
26	197.20	14.53	−26.72
28	98.60	18.80	−33.04
29	197.20	18.52	−33.54
30	197.20	18.39	−33.81
31	197.20	18.29	−34.02
32	98.60	18.27	−34.08
33	197.20	18.26	−34.08
34	197.20	25.11	−37.89
35	197.20	41.11	−49.40
36	345.10	25.05	−37.93

Table 5.2: PV inverters data for the single-phase system

Next, we consider the three-phase model of the IEEE 37-node feeder. The PV units are assumed to installed at the nodes indexed in blue in Fig. 5.1. The PV penetration profile is adopted from the data available in [7]. An instance with high PV penetration was chosen where the SDR scheme is unable to find a feasible voltage profile. The PV penetration data is summarized in Table 5.3, where the PV units are installed at one of the phases at selected buses. Again, we use the same constraints on $(P_{k,\phi}^{(R)}, Q_{k,\phi}^{(R)})$ and the cost function from the first scenario are considered.

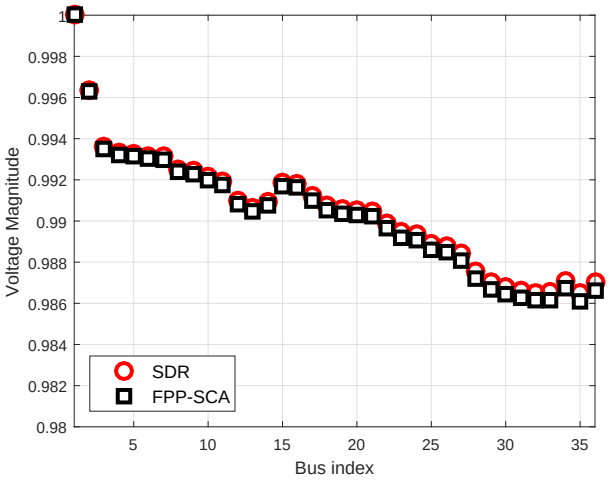


Figure 5.2: The optimal voltage profile using FPP-SCA and the SDR at 7 : 00 PM.

Fig. 5.4 depicts the optimal voltage profiles for the three phases across all the buses. It is clear that the voltage magnitude is high at the nodes with PV units which indicates the high power injection at these buses. Table 5.3 lists the amount of curtailed power at the PV units, as well as the reactive power injected/absorbed by the PV inverters.

Remark 2. *Initializing the algorithm from the flat voltage profile in high PV penetration scenarios, the method needs about* 1000 *iterations in order to converge, where the subproblem can be solved in* 5 *and* 2 *seconds on average in each iteration for the single- and multi-phase systems, respectively. However, initializing the algorithm from the optimal voltage profile of close enough preceding time instance can significantly speed up the proposed algorithm. Using this strategy of warm start, the method takes only about* 6 *iterations (i.e., 15-30 seconds) to converge.*

The ability of the proposed algorithm to solve the OPF problem instances for transmission networks is demonstrated using the test cases described in [73]. Additionally, a modified version of a 5-bus network presented in [15] is utilized. The load and generation limits are edited to the values in Table 5.4, where the real and reactive power quantities are given in MVA and MVAr, respectively. All the other network parameters correspond to the original dataset. Table 5.5 presents the lower bound provided by SDR, the cost of the solution produced by FPP-SCA, and the cost obtained by IPOPT.

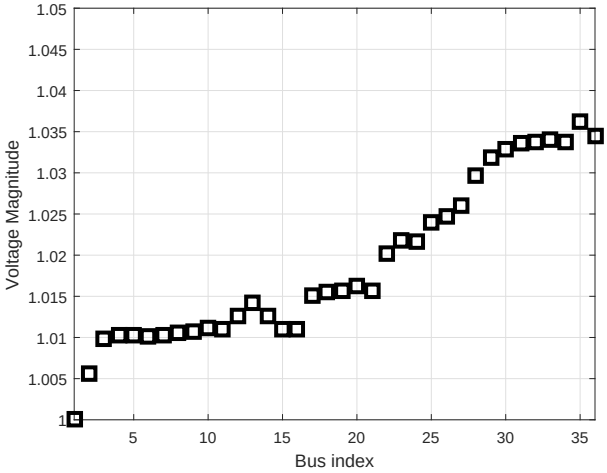


Figure 5.3: The optimal voltage profile using FPP-SCA at 1 : 00 PM.

We also compare the propose method against the moment-based relaxation [73] and the Laplacian-based approach in Table 5.6 [70].

Remark 3. *For transmission networks, there are limits on the apparent power flows on the lines. Such constraints can be written as nonconvex quadratic ones after introducing slack variables. This transformation is necessary to write the OPF in QCQP form. The resulting constraints can be handled using the same way as shown before.*

Consider the 14-,39-, 57-, 118-, and 300-bus systems (see e.g., [73]) and a modified version of the 5-bus network illustrated in [15]. These networks do not have any installed PV inverters, and hence, only traditional generation cost is considered. Even though the IPOPT is the most reliable software for solving the OPF problem for transmission systems, the modified WB5 system represents a case where IPOPT fails; on the other hand, FPP-SCA provides a feasible (and close to optimal) solution. In addition, no nonlinear solver among Trusted Region Augmented Lagrangian Multipliers (TRALM [104]), Primal Dual Interior Point Method (PDIPM [104]), and the Matlab Interior Point Solver (MIPS [124]), was able to reveal feasible solutions for all the transmission networks we tested. The solutions obtained using our algorithm are compared with the results of the algorithms in [73] and [70] in Table 5.6. The FPP-SCA algorithm yields solutions that achieve generation costs *very* close to the SDR bound, in all the problem instances

k	ϕ	$\overline{P}_{k,\phi}^{(R)}$	$(\overline{P}_{k,\phi}^{(R)} - P_{k,\phi}^{(R)})$	$Q_{k,\phi}^{(R)}$
7	3	97.86	1.23	0.32
10	1	97.86	0.14	−0.46
13	2	195.71	0	0.17
20	1	195.71	0.23	−0.81
22	3	195.71	2.82	0.85
26	3	195.71	3.55	1.14
28	3	97.86	4.77	1.60
29	1	195.71	0.02	−3.51
30	1	195.71	0.01	−3.87
32	3	97.86	6.76	2.31
33	3	195.71	6.64	2.28
35	2	195.71	0	4.82
36	3	342.5	4.78	1.63

Table 5.3: PV inverters data for the three-phase system.

Table 5.4: WB5 network data.

Node	Load		Gen. Limit			
	$P^{(L)}$	$Q^{(L)}$	$\overline{P}^{(G)}$	$\underline{P}^{(G)}$	$\overline{Q}^{(G)}$	$\underline{Q}^{(G)}$
1	0	0	350	0	300	−30
2	150	20	–	–	–	–
3	150	20	–	–	–	–
4	75	10	–	–	–	–
5	0	0	450	0	300	−30

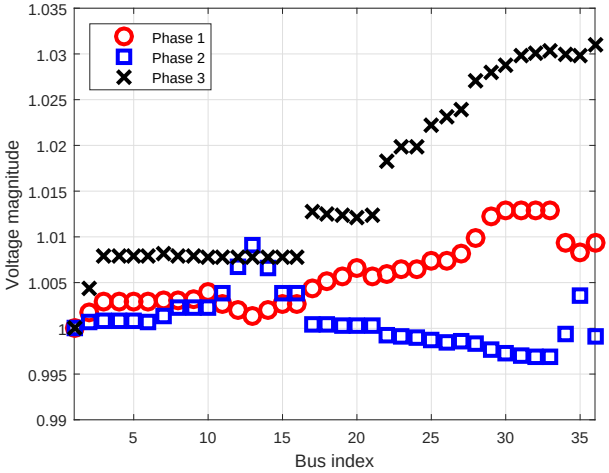


Figure 5.4: The optimal voltage profile at the three phases.

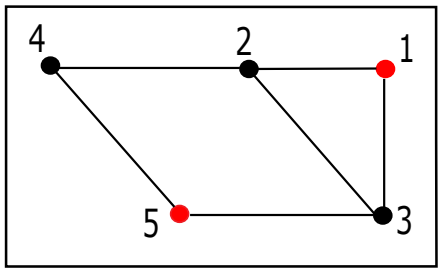


Figure 5.5: IEEE 5-node test feeder. Nodes with generators are depicted in red.

considered. Additionally, we compare the maximum mismatch in the nodal power injection. We can see that the maximum mismatch in the power injection of our solution is considerably lower than the mismatch in the solutions produced by [73] and [70]. Also, whereas the solutions given by the other algorithms violate the line flow constraints by small values, the FPP-SCA algorithm is capable of finding solutions that do not violate these constraints. In these other algorithms, we may need to use higher moments to reduce the mismatch and the violation which makes the computational problem much harder. The IPOPT solver is capable of finding solutions that are as accurate as the FPP-SCA solution; however, IPOPT may mistakenly indicate infeasibility of the OPF problem in cases where the problem is actually feasible.

As an illustrative example, the networks WB5 and *case9mod*[1] are utilized next to demonstrate the ability of the proposed algorithm to identify the constraints that render

Test Case	SDR Bound	FPP Cost	IPOPT Cost
WB5	1.1345×10^3	1.2647×10^3	–
case14Q	3.3016×10^3	3.3019×10^3	3.3018×10^3
case14L	9.3536×10^3	9.3875×10^3	9.3592×10^3
case39Q	1.0814×10^4	1.1225×10^4	1.1221×10^4
case39L	4.1889×10^4	4.1974×10^4	4.1896×10^4
case57Q	7.3472×10^3	7.3541×10^3	7.3518×10^3
case57L	4.3914×10^4	4.3998×10^4	4.3982×10^4
case118Q	8.1508×10^4	8.1521×10^4	8.1509×10^4
case118L	1.3391×10^5	1.3510×10^5	1.3490×10^5
case300	7.1957×10^5	7.2016×10^5	7.1973×10^5

Table 5.5: Test cases and results.

Case	Maximum Injection Mismatch (MVA)		
	MR [73]	LA [70]	FPP-SCA
WB5	7.72×10^{-9}	3.43	9.07×10^{-11}
case14Q	1.08×10^{-3}	1.20×10^{-5}	5.15×10^{-8}
case14L	5.67×10^{-2}	3.77×10^{-5}	2.57×10^{-8}
case39Q	1.36×10^{-1}	$--$	1.26×10^{-4}
case39L	4.60×10^{-3}	8.52×10^{-3}	2.83×10^{-5}
case57Q	6.49×10^{-3}	6.99×10^{-4}	2.45×10^{-7}
case57L	8.76×10^{-4}	4.42×10^{-4}	2.37×10^{-6}
case118Q	2.13×10^{-1}	2.98×10^{-3}	7.52×10^{-6}
case118L	4.42×10^{-1}	2.01×10^{-3}	1.02×10^{-4}
case300	5.14×10^{-2}	7.01×10^{-2}	7.74×10^{-3}

Table 5.6: Comparison between the power injection mismatch from [73] and [70] with our method.

the OPF infeasible. For WB5 network, the reactive demand at node 2 is increased from 20 MVAr to 70 MVAr. For this setting, the problem is infeasible. The value of the slack variables associated with the voltage magnitude constraints are illustrated in the upper panel of Fig. 5.6. Additionally, the slack variables associated with the loads are illustrated in the lower panel of Fig. 5.6, where s_k^P and s_k^Q are given by $\max\{s_{\underline{k}}^P, s_{\overline{k}}^P\}$ and $\max\{s_{\underline{k}}^Q, s_{\overline{k}}^Q\}$, respectively. The slack variables suggest that the upper limit of the voltage magnitude at node 1 is tight and the lower bound on the voltage magnitude of node 2 is tight. This suggests that the voltage difference between node 1 and node 2 should be larger in order to allow a higher flow of reactive power from the generator at node 1. Also, the slack variables that correspond to the load demand indicate that the demand at node 2 can not be satisfied under the existing network constraints.

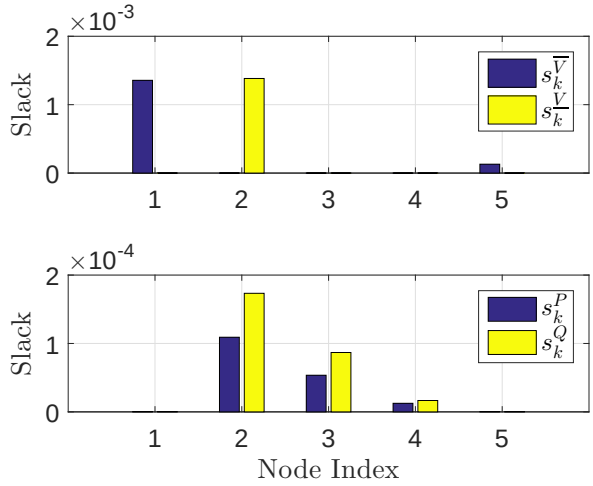


Figure 5.6: The values of slacks of the voltage and power demand constraints for infeasible WB5.

A modified version of a 9-bus network [15] is used next to further demonstrate the effectiveness of the FPP-SCA approach in identifying the problematic constraints. The voltage upper and lower limits were modified to be 1.05 and 0.95, respectively. In this scenario, the test case is infeasible. In Fig. 5.7, the lower panel shows the slacks associated with the active and reactive power demand constraints. The values of these slacks are very small ($\sim 10^{-6}$), indicating that these constraints are easily satisfied. In

[1]Available at `http://www.maths.ed.ac.uk/optenergy/LocalOpt/9busnetwork.html`

the top panel of Fig. 5.7, however, the slack associated with the lower limit constraint on the voltage magnitude at bus 9 is much higher, suggesting that this is the problematic constraint. Indeed, relaxing the lower limit of the voltage magnitude at bus 9 to 0.94 makes the problem feasible. These examples represent cases where the network operator can quickly discern the source of infeasibility from the results produced by the FPP method.

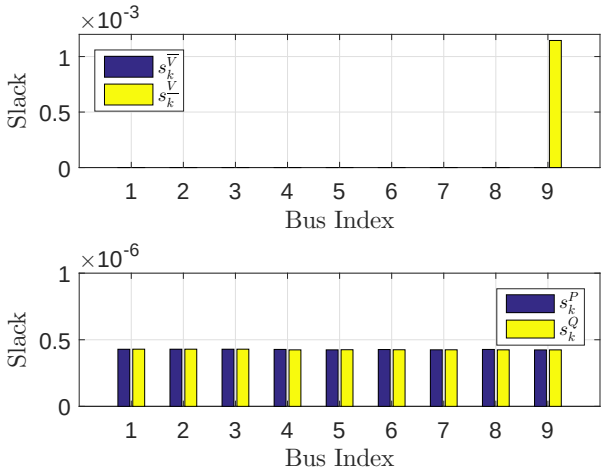


Figure 5.7: The values of the slacks of the voltage and power demand constraints for infeasible *case9mod*. The other bars on the top panel are not visible because their heights are $\sim 10^{-7}$.

5.4.2 Experimental Results for Radial Networks Featuring Delta-Connections

In this subsection, the efficacy of the proposed approach is demonstrated using two radial distribution feeders featuring both delta and wye connections. The proposed approach is shown to be able to provide a solution that minimizes the sum of the amount of power curtailed by the RESs and the power losses in the network while respecting the physical constraints.

In the first experiment, the IEEE 37-node test feeder shown in Fig. 5.1 is considered. Since this feeder features only delta-connected loads, wye connections are added as described in Table 5.9. The cost function $C_{loss}(p_{loss})$ is defined to be the square of p_{loss}.

Additionally, the cost function $C_k(\mathbf{p}_{\mathrm{Y},k}^{(c)}, \mathbf{p}_{\Delta,k}^{(c)})$ is defined to be the square of the amount of active power curtailed at bus k. The amount of power injected by every renewable source is shown in Table 5.7, along with the amount of power available from the RESs in Table 5.10. At the optimal solution using the FPP-SCA algorithm, the total power loss in the network is 58.6 KW, while the total amount of curtailed power at the renewables is 6.54 KW out of the 765.44 KW available active power at the RESs. Fig. 5.8 shows that the voltage profiles obtained using our proposed algorithm satisfy the magnitude constraints; i.e., the obtained operational point is indeed feasible (and optimal).

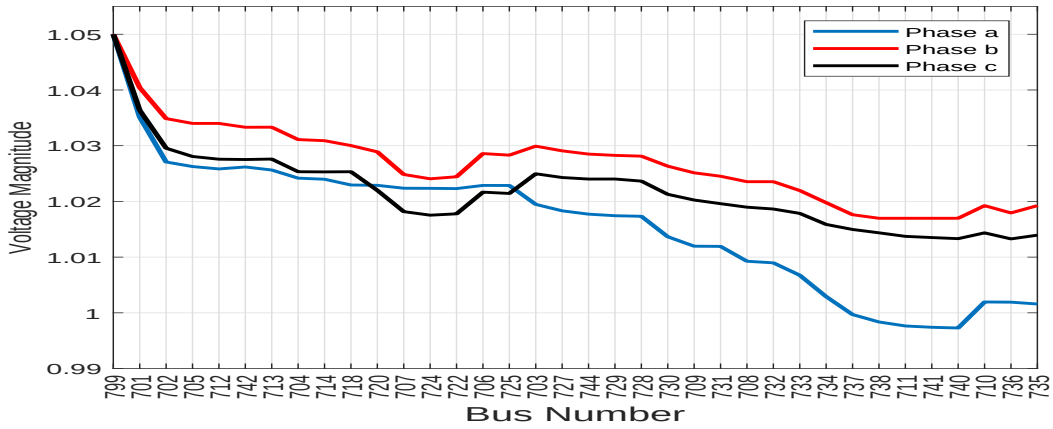


Figure 5.8: The optimal voltage profile (in pu) using FPP-SCA for IEEE 37-node test feeder.

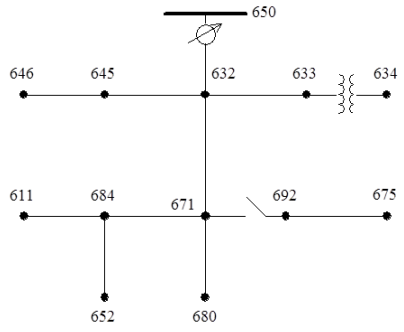


Figure 5.9: IEEE 13-node test feeder.

The IEEE 13-bus distribution network shown in Fig 5.9 is utilized in the second test. In this network, there are five shunt capacitor banks installed at different nodes that can

provide reactive power. These capacitors are modeled as a source that can provide only reactive power up to the capacitor's capacity. Details about the loads and the capacitor capacities are listed in Table 5.11. A cost function that minimizes the power losses in the network is used. The proposed approach was able to obtain a feasible solution that minimized the considered cost function. The injections from the capacitor along with the amount of of power drawn from the substation are summarized in Table 5.8. The total power loss in the network in the solution is 37.52 KW. The voltage magnitudes at all the buses are depicted in Fig. 5.10, where it is clear that all the magnitudes lie within the prescribed limits.

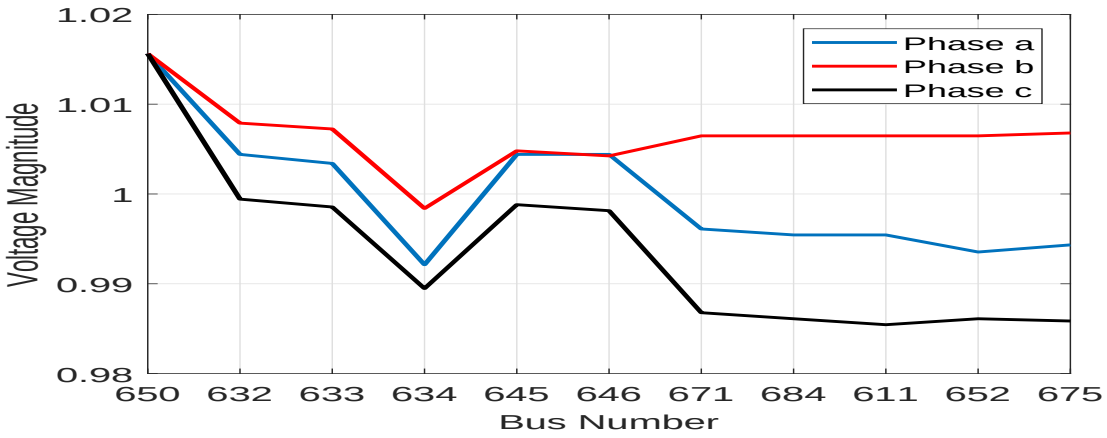


Figure 5.10: The optimal voltage profile (in pu) using FPP-SCA for IEEE 13-node test feeder.

For comparison purposes, it is worth emphasizing that the OpenDSS [27] software can provide flow solutions for the considered distribution networks without distributed energy sources, with maximum violation of the constraints in the order of 10^{-5}; in contrast, our algorithm is able to find feasible solutions even for cases where renewable energy sources are present, with accuracy in the order of 10^{-12}.

Table 5.7: Results of the proposed approach [kVA].

Node	Power Injected		
	Phase a	**Phase b**	**Phase c**
799	701+482i	604.5+322.9i	577.85+201.44i
713	0	0	33+16.5i
718	33+16.4i	0	0
724	0	66+33i	0
729	66+32.9i	0	0
730	0	0	66+25i
732	0	0	66+33i
734	0	0	33+16.5i
737	66+33i	0	0
738	66+23i	0	0
741	0	0	33
740	0	0	66+0.5i
736	0	66+33i	0
735	0	0	108.9+19.66i

Table 5.8: Results of the proposed approach in IEEE-13 feeder [kVA].

Node	Power Injected		
	Phase a	**Phase b**	**Phase c**
650	1228.4+584.15i	967.9+555.5i	1307.4+571.4i
611	0	0	100i
675	200i	0.8i	200i

Table 5.9: Load data for IEEE-37 distribution feeder [kVA].

Node	Loads					
	Delta-connected			Wye-connected		
	ab	bc	ac	Phase a	Phase b	Phase c
701	140+ 70i	140 + 70i	350 + 175i	0	0	0
702	0	0	0	0	0	0
705	0	0	0	0	0	0
712	0	0	85 + 40i	0	0	0
742	8 + 4i	85 + 40i	0	0	0	0
713	0	0	85 + 40i	0	0	0
704	0	0	0	0	0	0
714	17 + 8i	21+ 10i	0	0	0	0
718	85 + 40i	0	0	0	0	0
720	0	0	85 + 40i	0	0	0
707	0	0	0	0	0	0
724	0	42 + 21i	0	0	0	0
722	0	140 + 70i	21+ 10i	0	0	0
706	0	0	0	0	0	0
725	0	42 + 21i	0	0	0	0
703	0	0	0	0	0	0
727	0	0	42 + 21i	0	0	0
744	42 + 21i	0	0	0	0	0
729	42 + 21i	0	0	0	0	0
728	42 + 21i	42 + 21i	42 + 21i	65 + 30i	55 + 21i	42+ 21i
730	0	0	85 + 40i	0	0	0
709	0	0	0	0	0	0
731	0	85 + 40i	0	0	0	0
708	0	0	0	0	0	0
732	0	0	42 + 21i	0	0	0
733	85 + 40i	0	0	0	0	0
734	0	0	42 + 21i	0	0	0
737	140 + 70i	0	0	0	0	0
738	126 + 62i	0	0	0	0	0
711	0	0	0	0	0	0
741	0	0	42 + 21i	0	0	0
740	0	0	85 + 40i	0	0	0
710	0	0	0	0	0	0
736	0	42 + 21i	0	0	0	0
735	0	0	85 + 40i	0	0	0

Table 5.10: RES data for IEEE-37 distribution feeder [kVA].

Node	**Available power from RES**		
	Phase a	**Phase b**	**Phase c**
713	0	0	33
718	33	0	0
724	0	66	0
729	66	0	0
730	0	0	66
732	0	0	66
734	0	0	33
737	66	0	0
738	66	0	0
741	0	0	33
740	0	0	66
736	0	66	0
735	0	0	115.44

Table 5.11: Load data for IEEE-13 distribution feeder [kVA].

Node	**Loads**						**Available reactive power from Caps**		
	Delta-connected			**Wye-connected**					
	ab	**bc**	**ac**	**Phase a**	**Phase b**	**Phase c**	**Phase a**	**Phase b**	**Phase c**
632	0	0	0	8.5+5i	33+19i	58.5+34i	0	0	0
633	0	0	0	0	0	0	0	0	0
634	0	0	0	160+110i	120+90i	120+90i	0	0	0
645	0	0	0	0	170+125i	0	0	0	0
646	0	230+132i	0	0	0	0	0	0	0
671	385+220i	385+220i	555+371i	8.5+5i	33+19i	58.5+34i	0	0	0
684	0	0	0	0	0	0	0	0	0
611	0	0	0	0	0	170+80i	0	0	100i
652	0	0	0	128+86i	0	0	0	0	0
675	0	0	0	485+190i	68+60i	290+212i	200i	200i	200i

Chapter 6

Conclusions and Future Research Directions

In this dissertation we presented machine learning and optimization approaches for optimal resource management and state estimation in multi-phase distribution networks. In this final chapter, we recap the main results and take-home points in the dissertation and point out possible directions for future research.

6.1 Thesis Summary

Benefiting from recent advances in machine learning and leveraging the abundance of historical data, a novel data-driven learning-based architecture was presented in Chapter 3 for distribution system state estimation. The proposed approach designs a neural network that can accommodate several types of measurements as well as pseudo-measurements. Historical load and energy generation data is used to train a neural network in order to produce an approximation of the network state. Then, this estimate is fed to a Gauss-Newton algorithm for refinement. Our realistic experiments suggest that the combination offers fast and reliable convergence to the optimal solution. The IEEE-37 test feeder was used to test the proposed approach in scenarios that include distributed energy sources. The proposed learning approach shows superior performance results in terms of the accuracy of the estimates as well as computation time.

The joint learning and optimization approach in Chapter 3 entails performing several Gauss-Newton iterations. The flip side is that these iterations might come at huge computational cost especially for large scale networks. Therefore, a novel learning model that facilitates real-time monitoring of distribution network operations was proposed in Chapter 4. The graph-pruned NN approach utilizes the approximate separability of the DSSE problem resulting from installing μPMUs at some buses. By pruning the unneeded NN connections, the resulting model prevents over-fitting behavior exploiting the available knowledge regarding the network physics. A greedy algorithm was proposed for installing μPMUs in order to minimize the diameter of the resulting partitioning of the distribution feeder. Simulation results corroborate the efficacy of the greedy algorithm for finding near-optimal placement solutions. Also, the proposed PAWNN approach shows superior performance in estimating the network state from few noisy real-time measurements and pseudo-measurements on the IEEE-37 distribution feeder. In addition, the proposed approach was shown to be robust against corrupted measurements.

Building on the so-called *Feasible Point Pursuit - Successive Convex Approximation* algorithm, an effective solver for AC OPF problem in multi-phase networks with renewables was presented in Chapter 5. First, the AC OPF problem was presented for multi-phase networks with potentially pervasive integration of renewable energy sources. The problem was formulated as a nonconvex QCQP, and solved using the FPP-SCA algorithm. The proposed algorithm was shown to be effective in solving the OPF problem in many settings, including single- and three-phase models for power networks with renewables. The FPP-SCA is able to identify optimal operating points that satisfy the network constraints even under high RES penetration setups.

Also, an extension of the AC OPF formulation was put forth for radial networks featuring wye- and delta-connected generation and load units. The resulting problem was also formulated as QCQP which is amenable to the FPP-SCA solver. The FPP-SCA was empirically shown to be very effective in solving the AC OPF when the power curtailed from RESs is to be minimized – a setup where competing solution methods based on convex relaxation approaches may fail in identifying even a feasible solution. Finally, the ability of the proposed algorithm to find more accurate solutions than the moment-based relaxation and the Laplacian-based approach was demonstrated using

several IEEE test cases with and without delta-connections. The algorithm was also shown to be able to identify constraints that render the OPF problem infeasible.

6.2 Future Research Directions

The results in this thesis open up interesting directions of future research topics including utilizing knowledge of physical model parameters in the learning approach for state estimation, using the physics-aware NN model for managing assets in the smart grid, and developing unit commitment solvers. Next, we outline some directions for future research.

6.2.1 Physics-Aware Learning Models

The results obtained from our data-driven approaches are truly intriguing – and hinge on the particular design of the learning models as well as the distinct training approaches utilized. This initial success prompts us to consider the following directions.

• **Physics-aware learning models** Whereas our particular design of the training procedure in the state estimator is a matter of engineering art, deliberate design of learning models that utilize knowledge of the network parameters in addition to the structure can boost the performance of the learning models. In addition, a verifiable learning approach can be devised by knowing the exact network parameters.

• **Theoretical analysis** What should the architectural complexity of the neural network 'preconditioner' be to ensure good performance? How many stages and how many nodes? We need theoretical complexity analysis to understand the trade-offs. By answering these questions, we can also solve the *dual* problem of optimal placement of measuring devices in order to achieve a monitoring accuracy employing a particular learning model.

• **Network-wide learning-based energy management** While our work focused on optimal control of energy assets, learning methods can make real-time control feasible [116]. The group behavior of multi-agent learning procedures is an area that requires more research. In addition, a model-free reinforcement learning framework can be developed that aims for maximal utilization of energy assets and system-wide stable operation.

6.2.2 Unit Commitment Problem

The Unit Commitment (UC) problem in power systems involves making multi-period generation schedules to be used to meet an anticipated demand over a future short-term period [107]. Due to increased penetration of RESs, solving the UC problem has become critically important in order to deal with the uncertainty introduced in the system by random fluctuations in the availability of wind and solar energy. In order to support integration of wind and solar energy resources, the UC problem has been studied for power systems with ES units [25]. Also, many variants of the UC problem have been proposed to determine the placement and operation of ES units in power networks [28, 105, 109]. Nevertheless, battery operation in distribution networks has not been well-studied in the literature, especially for multi-phase systems. Since many investor-owned ES units have been installed recently in distribution networks, solving the UC problem for ensuring a profitable turnover for these units is necessary. In standard form, the UC problem can be formulated as

$$\max_{\{\mathbf{v}(t),\mathbf{u}(t)\}_{0\le t\le T}} \sum_{0\le t\le T} f(\mathbf{u}(t)) \tag{6.1a}$$

$$\text{s.t.} \quad \left.\begin{aligned} \mathbf{h}(\mathbf{v}(t),\mathbf{u}(t)) &= \mathbf{0}, \\ \mathbf{g}(\mathbf{v}(t),\mathbf{u}(t)) &\le \mathbf{0}, \end{aligned}\right\} \quad 0 \le t \le T \tag{6.1b}$$

$$\boldsymbol{\Psi}(\mathbf{u}(t),\mathbf{u}(t-1)) \le \mathbf{0}, \quad 1 \le t \le T \tag{6.1c}$$

where $\mathbf{v}(t)$ denote the state variables, i.e, voltages, at time instance t, while $\mathbf{u}(t)$ collects the other variables at the same time step. The cost function aggregates the generation cost and utility-based terms. The constraint (6.1c) encapsulates the binding constraints on SoC and ramp constraints on all energy resources.

An accurate formulation that incorporates investor-installed ES units in distribution networks and maximizes profits accrued in battery operation is needed. Multi-phase unbalanced distribution systems offer a formidable challenge in obtaining accurate solutions that respect the network constraints while ensuring profitable returns for installed ES units. An SCA based approach in the vein of FPP-SCA can be adopted to tackle the problem [113], and a decentralized approach [53,110] can also be pursued that allows for distributed actions from the owners of the ES units. Furthermore, when the available

SoC and State-of-Health (SoH) estimates are inaccurate [81], a stochastic programming approach needs to be considered to handle the resulting uncertainty introduced in the formulation (6.1).

References

[1] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/

[2] O. Alsac, J. Bright, M. Prais, and B. Stott, "Further developments in lp-based optimal power flow," *IEEE Transactions on Power Systems*, vol. 5, no. 3, pp. 697–711, 1990.

[3] H. Ambriz-Perez, E. Acha, and C. Fuerte-Esquivel, "Advanced svc models for newton-raphson load flow and newton optimal power flow studies," *IEEE Transactions on power systems*, vol. 15, no. 1, pp. 129–136, 2000.

[4] S. Amin, V. Gerhart, E. Rodin, S. Amin, V. Gerhart, and E. Rodin, "System identification via artificial neural networks-applications to on-line aircraft parameter estimation," in *World Aviation Congress*, Anaheim, CA, 1997, p. 5612.

[5] X. Bai, H. Wei, K. Fujisawa, and Y. Wang, "Semidefinite programming for optimal power flow problems," *International Journal of Electrical Power and Energy Systems*, vol. 30, no. 67, pp. 383 – 392, July 2008.

[6] R. Baldick, B. H. Kim, C. Chase, and Y. Luo, "A fast distributed implementation of optimal power flow," *IEEE Tran. on Power Systems*, vol. 14, no. 3, pp. 858–864, Aug 1999.

[7] J. Bank and J. Hambrick, "Development of a high resolution, real time, distribution-level metering system and associated visualization, modeling, and

data analysis functions," National Renewable Energy Laboratory (NREL), Golden, CO., Tech. Rep., 2013.

[8] M. E. Baran and A. W. Kelley, "State estimation for real-time monitoring of distribution systems," *IEEE Trans. on Power Systems*, vol. 9, no. 3, pp. 1601–1609, Aug 1994.

[9] ——, "A branch-current-based state estimation method for distribution systems," *IEEE Trans. on Power Systems*, vol. 10, no. 1, pp. 483–491, Feb 1995.

[10] M. E. Baran and F. F. Wu, "Network reconfiguration in distribution systems for loss reduction and load balancing," *IEEE Transactions on Power delivery*, vol. 4, no. 2, pp. 1401–1407, 1989.

[11] P. N. P. Barbeiro, J. Krstulovic, H. Teixeira, J. Pereira, F. J. Soares, and J. P. Iria, "State estimation in distribution smart grids using autoencoders," in *2014 IEEE 8th International Power Engineering and Optimization Conference (PEOCO2014)*, March 2014, pp. 358–363.

[12] D. Bienstock and G. Munoz, "LP approximations to mixed-integer polynomial optimization problems," *arXiv preprint arXiv:1501.00288*, 2015.

[13] A. Bonfiglio, M. Brignone, F. Delfino, and R. Procopio, "Optimal control and operation of grid-connected photovoltaic production units for voltage support in medium-voltage networks," *IEEE Trans. on Sustainable Energy*, vol. 5, no. 1, pp. 254–263, Jan 2014.

[14] 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)*, July 2016, pp. 1–5.

[15] W. A. Bukhsh, A. Grothey, K. I. McKinnon, and P. A. Trodden, "Local solutions of the optimal power flow problem," *IEEE Trans. on Power Systems*, vol. 28, no. 4, pp. 4780–4788, Nov 2013.

[16] G. Cavraro and V. Kekatos, "Graph algorithms for topology identification using power grid probing," *IEEE Control Systems Letters*, vol. 2, pp. 689–694, 2018.

[17] G. Chartrand, *Introductory graph theory.* New York: Dover, 1985.

[18] C. Coffrin, H. L. Hijazi, and P. V. Hentenryck, "The QC Relaxation: Theoretical and Computational Results on Optimal Power Flow," *IEEE Trans. on Power Systems*, vol. 31, no. 4, pp. 3008–3018, July 2016.

[19] C. Coffrin and P. Van Hentenryck, "A linear-programming approximation of AC power flows," *INFORMS Journal on Computing*, vol. 26, no. 4, pp. 718–734, 2014.

[20] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of control, signals and systems*, vol. 2, no. 4, pp. 303–314, 1989.

[21] A. A. da Silva, A. L. da Silva, J. S. de Souza, and M. Do Coutto Filho, "State forecasting based on artificial neural networks," in *Proc. 11th PSCC*, 1993, pp. 461–467.

[22] E. Dall'Anese and G. B. Giannakis, "Convex distribution system reconfiguration using group sparsity," in *2013 IEEE Power Energy Society General Meeting*, July 2013, pp. 1–5.

[23] E. Dall'Anese, H. Zhu, and G. B. Giannakis, "Distributed optimal power flow for smart microgrids," *IEEE Trans. on Smart Grid*, vol. 4, no. 3, pp. 1464–1475, Sep 2013.

[24] E. DallAnese, S. V. Dhople, and G. B. Giannakis, "Optimal dispatch of photovoltaic inverters in residential distribution systems," *IEEE Trans. on Sustainable Energy*, vol. 5, no. 2, pp. 487–497, April 2014.

[25] Y. Degeilh and G. Gross, "Stochastic simulation of utility-scale storage resources in power systems with integrated renewable resources," *IEEE Transactions on Power Systems*, vol. 30, no. 3, pp. 1424–1434, 2015.

[26] D. Deka, M. Chertkov, and S. Backhaus, "Structure learning in power distribution networks," *IEEE Transactions on Control of Network Systems*, pp. 1–1, 2018.

[27] R. C. Dugan, "Reference guide: The open distribution system simulator (OpenDSS)," *Electric Power Research Institute, Inc*, vol. 7, 2012.

[28] Y. Dvorkin, R. Fernandez-Blanco, D. S. Kirschen, H. Pandžić, J.-P. Watson, and C. A. Silva-Monroy, "Ensuring profitability of energy storage," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 611–623, 2017.

[29] I. Dzafic, R. A. Jabr, and T. Hrnjic, "Hybrid state estimation in complex variables," *IEEE Transactions on Power Systems*, 2018, DOI:10.1109/TPWRS.2018.2794401.

[30] I. Džafić, R. A. Jabr, I. Huseinagić, and B. C. Pal, "Multi-phase state estimation featuring industrial-grade distribution network models," *IEEE Transactions on Smart Grid*, vol. 8, no. 2, pp. 609–618, Mar 2017.

[31] A. G. Expósito, A. Gomez-Exposito, A. J. Conejo, and C. Canizares, *Electric energy systems: analysis and operation*. CRC press, 2016.

[32] X. Fang, D. Yang, and G. Xue, "Online strategizing distributed renewable energy resource access in islanded microgrids," in *IEEE Global Telecommunications Conference (GLOBECOM)*. Houston, TX: IEEE, 2011, pp. 1–6.

[33] M. Farivar and S. Low, "Branch Flow Model: Relaxations and Convexification, Part I," *IEEE Trans. on Power Systems*, vol. 28, no. 3, pp. 2554–2564, Aug 2013.

[34] O. P. Ferreira, "Local convergence of newton's method in banach space from the viewpoint of the majorant principle," *IMA Journal of Numerical Analysis*, vol. 29, no. 3, pp. 746–759, 2009.

[35] M. B. D. C. Filho and J. C. S. de Souza, "Forecasting-aided state estimationpart i: Panorama," *IEEE Trans. on Power Systems*, vol. 24, no. 4, pp. 1667–1677, Nov 2009.

[36] L. Gan and S. H. Low, "Convex relaxations and linear approximation for optimal power flow in multiphase radial networks," in *Power Systems Computation Conference*, Wrocaw, Poland, Aug 2014.

[37] ——, "Convex relaxations and linear approximation for optimal power flow in multiphase radial networks," in *Power Systems Computation Conference*, Wrocaw, Poland, Aug 2014.

[38] A. Garces, "A linear three-phase load flow for power distribution systems," *IEEE Trans. on Power Systems*, vol. 31, no. 1, pp. 827–828, Jan 2016.

[39] A. K. Ghosh, D. L. Lubkeman, and R. H. Jones, "Load modeling for distribution circuit state estimation," *IEEE Trans. on Power Del.*, vol. 12, no. 2, pp. 999–1005, Apr 1997.

[40] G. B. Giannakis, V. Kekatos, N. Gatsis, S. J. Kim, H. Zhu, and B. F. Wollenberg, "Monitoring and optimization for power grids: A signal processing perspective," *IEEE Signal Processing Magazine*, vol. 30, no. 5, pp. 107–128, Sept 2013.

[41] F. V. Gomes, S. Carneiro, J. L. R. Pereira, M. P. Vinagre, P. A. N. Garcia, and L. R. D. Araujo, "A new distribution system reconfiguration approach using optimum power flow and sensitivity analysis for loss reduction," *IEEE Transactions on Power Systems*, vol. 21, no. 4, pp. 1616–1623, Nov 2006.

[42] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio, "Maxout networks," *arXiv preprint arXiv:1302.4389*, 2013.

[43] S. S. Guggilam, E. Dall'Anese, Y. C. Chen, S. V. Dhople, and G. B. Giannakis, "Scalable optimization methods for distribution networks with high pv integration," *IEEE Trans. on Smart Grid*, vol. 7, no. 4, pp. 2061–2070, July 2016.

[44] G. Hug-Glanzmann and G. Andersson, "Decentralized optimal power flow control for overlapping areas in power systems," *IEEE Trans. on Power Systems*, vol. 24, no. 1, pp. 327–336, Feb 2009.

[45] M. S. Ibrahim, A. S. Zamzam, X. Fu, and N. D. Sidiropoulos, "Learning-based antenna selection for multicasting," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2018, pp. 1–5.

[46] V. N. Ioannidis, A. G. Marques, and G. B. Giannakis, "A recurrent graph neural network for multi-relational data," *arXiv preprint arXiv:1811.02061*, 2018.

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

[48] V. Kekatos and G. B. Giannakis, "Distributed robust power system state estimation," *IEEE Trans. on Power Systems*, vol. 28, no. 2, pp. 1617–1626, May 2013.

[49] V. Kekatos, G. Wang, H. Zhu, and G. B. Giannakis, "PSSE redux: Convex relaxation, decentralized, robust, and dynamic approaches," *CoRR*, vol. abs/1708.03981, 2017. [Online]. Available: http://arxiv.org/abs/1708.03981

[50] B. H. Kim and R. Baldick, "Coarse-grained distributed optimal power flow," *IEEE Trans. on Power Systems*, vol. 12, no. 2, pp. 932–939, May 1997.

[51] S.-J. Kim, G. Wang, and G. B. Giannakis, "Online semidefinite programming for power system state estimation," in *Proc. IEEE Conf. on Acoustics, Speech and Signal Process.*, Florence, Italy, May 2014, pp. 6024–6027.

[52] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[53] A. Konar, A. S. Zamzam, and N. D. Sidiropoulos, "Decentralized power system state estimation via non-convex multi-agent optimization," in *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Nov 2017, pp. 1060–1064.

[54] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[55] A. Y. Lam, B. Zhang, and D. N. C. Tse, "Distributed algorithms for optimal power flow problem," in *IEEE 51st Annual Conference on Decision and Control (CDC)*, 2012, pp. 430–437.

[56] J. Lavaei and S. Low, "Zero Duality Gap in Optimal Power Flow Problem," *IEEE Trans. on Power Systems*, vol. 27, no. 1, pp. 92–107, Feb 2012.

[57] J. Lavaei, D. Tse, and B. Zhang, "Geometry of power flows in tree networks," in *IEEE Power and Energy Society General Meeting*, San Diego, CA, July 2012.

[58] K. Lehmann, A. Grastien, and P. V. Hentenryck, "AC-Feasibility on Tree Networks is NP-Hard," *IEEE Trans. on Power Systems*, vol. 31, no. 1, pp. 798–801, Jan 2016.

[59] B. Lesieutre, D. Molzahn, A. Borden, and C. DeMarco, "Examining the limits of the application of semidefinite programming to power flow problems," in *49th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Sep 2011.

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

[61] Y. Lin and A. Abur, "A highly efficient bad data identification approach for very large scale power systems," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 5979–5989, 2018.

[62] J. Löfberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," in *IEEE International Symposium on Computer Aided Control Systems Design*, 2004.

[63] R. Louca, P. Seiler, and E. Bitar, "A rank minimization algorithm to enhance semidefinite relaxations of Optimal Power Flow," in *51st Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct 2013.

[64] S. H. Low, "Convex Relaxation of Optimal Power Flow, Part I: Formulations and Equivalence," *IEEE Trans. on Control of Network Systems*, vol. 1, no. 1, pp. 15–27, Mar 2014.

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

[66] E. Manitsas, R. Singh, B. C. Pal, and G. Strbac, "Distribution system state estimation using an artificial neural network approach for pseudo measurement modeling," *IEEE Trans. on Power Systems*, vol. 27, no. 4, pp. 1888–1896, Nov 2012.

[67] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Oper. Res.*, vol. 26, no. 4, pp. 681–683, Aug. 1978.

[68] O. Mehanna, K. Huang, B. Gopalakrishnan, A. Konar, and N. Sidiropoulos, "Feasible point pursuit and successive approximation of non-convex QCQPs," *IEEE Signal Processing Letters*, vol. 22, no. 7, pp. 804–808, July 2015.

[69] H. N. Mhaskar, "Neural networks for optimal approximation of smooth and analytic functions," *Neural computation*, vol. 8, no. 1, pp. 164–177, 1996.

[70] D. K. Molzahn, C. Josz, I. A. Hiskens, and P. Panciatici, "A Laplacian-Based Approach for Finding Near Globally Optimal Solutions to OPF Problems," *IEEE Trans. on Power Systems*, vol. 32, no. 1, pp. 305–315, Jan 2017.

[71] D. K. Molzahn, J. T. Holzer, B. C. Lesieutre, and C. L. DeMarco, "Implementation of a large-scale optimal power flow solver based on semidefinite programming," *IEEE Trans. on Power Systems*, vol. 28, no. 4, pp. 3987–3998, 2013.

[72] D. Molzahn and I. Hiskens, "Moment-based relaxation of the optimal power flow problem," in *Power Systems Computation Conference*, Wrocaw, Poland, Aug 2014.

[73] ——, "Sparsity-Exploiting Moment-Based Relaxations of the Optimal Power Flow Problem," *IEEE Trans. on Power Systems*, vol. 30, no. 6, pp. 3168–3180, Nov 2015.

[74] J. A. Momoh, M. El-Hawary, and R. Adapa, "A review of selected optimal power flow literature to 1993. ii. newton, linear programming and interior point methods," *IEEE Transactions on Power Systems*, vol. 14, no. 1, pp. 105–111, 1999.

[75] A. Monticelli, *State estimation in electric power systems: a generalized approach*. Springer Science & Business Media, 1999, vol. 507.

[76] A. L. Morelato and A. J. Monticelli, "Heuristic search approach to distribution system restoration," *IEEE Transactions on Power Delivery*, vol. 4, no. 4, pp. 2235–2241, Oct 1989.

[77] F. J. Nogales, F. J. Prieto, and A. J. Conejo, "A decomposition methodology applied to the multi-area optimal power flow problem," *Annals of operations research*, vol. 120, no. 1-4, pp. 99–116, 2003.

[78] R. Nuqui and A. G. Phadke, "Hybrid linear state estimation utilizing synchronized phasor measurements," in *IEEE Power Tech.* Lausanne, Switzerland: IEEE, 2007, pp. 1665–1669.

[79] D. O'Neill, M. Levorato, A. Goldsmith, and U. Mitra, "Residential demand response using reinforcement learning," in *First IEEE International Conference on Smart Grid Communications (SmartGridComm).* Gaithersburg, MD: IEEE, 2010, pp. 409–414.

[80] A. G. Phadke, J. S. Thorp, and K. Karimi, "State estimation with phasor measurements," *IEEE Trans. on Power Systems*, vol. 1, no. 1, pp. 233–238, Feb 1986.

[81] H. Rahimi-Eichi, U. Ojha, F. Baronti, and M.-Y. Chow, "Battery management system: An overview of its application in the smart grid and electric vehicles," *IEEE Industrial Electronics Magazine*, vol. 7, no. 2, pp. 4–16, 2013.

[82] M. Razaviyayn, "Successive convex approximation: Analysis and applications," Ph.D. dissertation, University of Minnesota, 2014.

[83] B. A. Robbins, H. Zhu, and A. D. Domínguez-Garca, "Optimal tap setting of voltage regulation transformers in unbalanced distribution systems," *IEEE Trans. on Power Systems*, vol. 31, no. 1, pp. 256–267, Jan 2016.

[84] P. Rousseaux, D. Mallieu, T. Van Cutsem, and M. Ribbens-Pavella, "Dynamic state prediction and hierarchical filtering for power system state estimation," *Automatica*, vol. 24, no. 5, pp. 595–618, 1988.

[85] W. Rudin, *Principles of mathematical analysis.* McGraw-hill New York, 1964, vol. 3.

[86] K. P. Schneider, B. A. Mather, B. C. Pal, C. . Ten, G. J. Shirek, H. Zhu, J. C. Fuller, J. L. R. Pereira, L. F. Ochoa, L. R. de Araujo, R. C. Dugan, S. Matthias, S. Paudyal, T. E. McDermott, and W. Kersting, "Analytic considerations and

design basis for the ieee distribution test feeders," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 3181–3188, May 2018.

[87] K. Schneider and E. Stuart, "Necpuc distribution systems & planning training: Data and Analytics for Distribution," Sep. 2017. [Online]. Available: http://eta-publications.lbl.gov/sites/default/files/9a._data_and_analytics_ems_core-ls.pdf

[88] R. Singh, B. Pal, and R. Jabr, "Choice of estimator for distribution system state estimation," *IET Generation, Transmission & Distribution*, vol. 3, no. 7, pp. 666–678, July 2009.

[89] SmartCitiesWorld News Team, "Germany predicted to be smart grid investment hot spot," Sep 2016. [Online]. Available: https://www.smartcitiesworld.net/news/news/germany-predicted-to-be-smart-grid-investment-hot-spot-974

[90] S. Sojoudi and J. Lavaei, "Physics of power networks makes hard optimization problems easy to solve," in *IEEE Power and Energy Society General Meeting*, San Diego, CA, July 2012.

[91] L. Sorber, M. V. Barel, and L. D. Lathauwer, "Unconstrained optimization of real functions in complex variables," *SIAM Journal on Optimization*, vol. 22, no. 3, pp. 879–898, 2012.

[92] B. Stott and O. Alsac, "Fast decoupled load flow," *IEEE transactions on power apparatus and systems*, no. 3, pp. 859–869, 1974.

[93] B. Stott, J. Jardim, and O. Alsaç, "Dc power flow revisited," *IEEE Transactions on Power Systems*, vol. 24, no. 3, pp. 1290–1300, 2009.

[94] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization methods and software*, vol. 11, no. 1-4, pp. 625–653, Jan 1999.

[95] X. Su, M. A. S. Masoum, and P. J. Wolfs, "Optimal PV inverter reactive power control and real power curtailment to improve performance of unbalanced four-wire LV distribution networks," *IEEE Trans. on Sustainable Energy*, vol. 5, no. 3, pp. 967–977, July 2014.

[96] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu, and N. D. Sidiropoulos, "Learning to optimize: Training deep neural networks for wireless resource management," *CoRR*, vol. abs/1705.09412, 2017. [Online]. Available: http://arxiv.org/abs/1705.09412

[97] L. Wan, M. Zeiler, S. Zhang, Y. Le Cun, and R. Fergus, "Regularization of neural networks using dropconnect," in *International Conference on Machine Learning*, 2013, pp. 1058–1066.

[98] G. Wang, V. Kekatos, A. J. Conejo, and G. B. Giannakis, "Ergodic energy management leveraging resource variability in distribution grids," *IEEE Trans. on Power Systems*, vol. PP, no. 99, pp. 1–11, Feb 2016.

[99] G. Wang, A. S. Zamzam, G. B. Giannakis, and N. D. Sidiropoulos, "Power system state estimation via feasible point pursuit: Algorithms and Cramér-Rao bound," *IEEE Transactions on Signal Processing*, vol. 66, no. 6, pp. 1649–1658, Mar 2018.

[100] G. Wang, H. Zhu, G. B. Giannakis, and J. Sun, "Robust power system state estimation from rank-one measurements," *IEEE Trans. Control Netw. Syst.*, 2019.

[101] G. Wang, G. B. Giannakis, and J. Chen, "Robust and scalable power system state estimation via composite optimization," *arXiv preprint arXiv:1708.06013*, 2017.

[102] G. Wang, G. B. Giannakis, J. Chen, and J. Sun, "Distribution system state estimation: An overview of recent developments," *Front. Inf. Technol. Electron. Eng.*, vol. 20, no. 1, pp. 1–14, Jan. 2019.

[103] H. Wang and N. N. Schulz, "A revised branch current-based distribution system state estimation algorithm and meter placement impact," *IEEE Trans. on Power Systems*, vol. 19, no. 1, pp. 207–213, Feb 2004.

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

[105] Y. Wen, C. Guo, H. Pandžić, and D. S. Kirschen, "Enhanced security-constrained unit commitment with emerging utility-scale energy storage," *IEEE Transactions on power Systems*, vol. 31, no. 1, pp. 652–662, 2016.

[106] Y. Weng, Y. Liao, and R. Rajagopal, "Distributed energy resources topology identification via graphical modeling," *IEEE Transactions on Power Systems*, vol. 32, pp. 2682–2694, 2017.

[107] A. J. Wood, B. F. Wollenberg, and G. B. Sheblé, *Power generation, operation, and control*. John Wiley & Sons, 2013.

[108] W. A. Wulf, "Great achievements and grand challenges," *The Bridge*, vol. 30, no. 3, p. 4, 2000.

[109] A. S. Zamzam, E. Dall'Anese, and N. D. Sidiropoulos, "A dantzig-wolfe decomposition-based approach for distributed storage management," in *Power Systems Computation Conference*, Dublin, Ireland, Jun 2018.

[110] A. S. Zamzam, X. Fu, E. Dall'Anese, and N. D. Sidiropoulos, "Distributed optimal power flow using feasible point pursuit," in *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Curacao, Dutch Islands, Dec 2017.

[111] A. S. Zamzam, N. D. Sidiropoulos, and E. Dall'Anese, "Beyond relaxation and newton-raphson: Solving AC OPF for Multi-phase Systems with Renewables," *IEEE Trans. on Smart Grid*, vol. 9, no. 5, pp. 3966 – 3975, Sept. 2018.

[112] A. S. Zamzam, C. Zhao, E. Dall'Anese, and N. D. Sidiropoulos, "A QCQP Approach for OPF in Multiphase Radial Networks with Wye and Delta Connections," in *IREP Symposium Bulk Power System Dynamics and Control*, Espinho, Portugal, Aug 2017.

[113] A. S. Zamzam, E. Dall'Anese, C. Zhao, J. A. Taylor, and N. D. Sidiropoulos, "Optimal water-power flow problem: Formulation and distributed optimal solution," *IEEE Trans. Control Netw. Syst.*, vol. 6, no. 1, Mar 2019.

[114] A. S. Zamzam, X. Fu, and N. D. Sidiropoulos, "Data-driven learning-based optimization for distribution system state estimation," *IEEE Trans. on Power Systems*, 2019. (Early Access).

[115] A. S. Zamzam and N. D. Sidiropoulos, "Physics-aware neural networks for distribution system state estimation," *arXiv preprint arXiv:1903.09669*, 2019.

[116] A. S. Zamzam, B. Yang, and N. D. Sidiropoulos, "Energy storage management via deep q-networks," *arXiv preprint arXiv:1903.11107*, 2019.

[117] B. Zhang, A. S. Lam, A. Dominguez-Garcia, and D. Tse, "An optimal and distributed method for voltage regulation in power distribution systems," *IEEE Trans. on Power Systems*, vol. 30, no. 4, pp. 1714–1726, July 2015.

[118] H. Zhang, V. Vittal, G. T. Heydt, and J. Quintero, "A relaxed AC optimal power flow model based on a Taylor series," in *Innovative Smart Grid Technologies-Asia (ISGT Asia)*, 2013.

[119] L. Zhang, G. Wang, and G. B. Giannakis, "Real-time power system state estimation and forecasting via deep neural networks," *arXiv preprint arXiv:1811.06146*, 2018.

[120] Q. Zhang, Y. Chakhchoukh, V. Vittal, G. T. Heydt, N. Logic, and S. Sturgill, "Impact of pmu measurement buffer length on state estimation and its optimization," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1657–1665, May 2013.

[121] Y. Zhang, R. Madani, and J. Lavaei, "Conic relaxations for power system state estimation with line measurements," *IEEE Trans. Control Netw. Syst.*, 2017 (to appear).

[122] ——, "Conic relaxations for power system state estimation with line measurements," *IEEE Transactions on Control of Network Systems*, to be published.

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

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

[125] R. Zivanovic and C. Cairns, "Implementation of PMU technology in state estimation: an overview," in *IEEE AFRICON*, Stellenbosch, South Africa, 1996, pp. 1006–1011.

Appendix A

Proof of Proposition 1

To prove the proposition, we first invoke the following lemma:

Lemma 1 ([20, Theorem 2])**.** *Let $\sigma(\cdot)$ be any continuous sigmoidal function. Then, given any function $f(\cdot)$ that is continuous on the d-dimensional unit cube $\boldsymbol{I}_d = [0,1]^d$, and $\epsilon > 0$, there is a sum, $g(\cdot) : \mathbb{R}^d \to \mathbb{R}$, of the form*

$$g(\mathbf{z}) = \sum_{t=1}^{T} \alpha_t \sigma(\mathbf{w}_t^T \mathbf{z} + \beta_t) \tag{A.1}$$

for which,

$$|g(\mathbf{z}) - f(\mathbf{z})| < \epsilon \qquad \forall \mathbf{z} \in \boldsymbol{I}_d.$$

Proof of Proposition 1. Note that the vector-valued function $\mathbf{F}(\cdot)$ can be represented as K separate scalar-valued functions. In order to prove the proposition, we start by considering approximating a scalar-valued function $f_k(\mathbf{z})$ that represents the mapping between $\mathbf{z}$ and the k-th element of $\mathbf{F}(\mathbf{z})$.

Since $\mathbf{z}^j$'s are finite with length L, finite maximum and minimum along each dimension can be obtained. Let the vectors that collect the maximum and minimum values be denoted by $\overline{\mathbf{z}}$ and $\underline{\mathbf{z}}$, respectively. Then, each training sample $\mathbf{z}^j$ is replaced by $\tilde{\mathbf{z}}^j = \mathbf{D}_{\overline{\mathbf{z}}-\underline{\mathbf{z}}}(\mathbf{z}^j - \underline{\mathbf{z}})$, where $\mathbf{D}_{\overline{\mathbf{z}}-\underline{\mathbf{z}}}$ is a diagonal matrix that has the values of $\overline{\mathbf{z}} - \underline{\mathbf{z}}$ on the diagonal. Therefore, the vectors $\tilde{\mathbf{z}}^j$ are inside the L-dimensional cube $\boldsymbol{I}_L$. According to

Lemma 1, there exists a sum $\tilde{g}_k(\tilde{\mathbf{z}})$ in the form of

$$\tilde{g}_k(\tilde{\mathbf{z}}) = \sum_{t=1}^{T_k} \tilde{\alpha}_{t,k}\sigma(\tilde{\mathbf{w}}_{t,k}^T\mathbf{z} + \tilde{\beta}_{t,k}) \tag{A.2}$$

that satisfies

$$|f_k(\tilde{\mathbf{z}}^j) - \tilde{g}_k(\tilde{\mathbf{z}}^j)| < \epsilon_1 \qquad \forall\ \tilde{\mathbf{z}}^j \tag{A.3}$$

for $\epsilon_1 > 0$. Then, let $g_k(\mathbf{z})$ be a mapping in the form of (A.2) where the parameters are given by

$$\alpha_{t,k} = \tilde{\alpha}_{t,k}, \quad \beta_{t,k} = \tilde{\beta}_{t,k} - \tilde{\mathbf{w}}_{t,k}^T\mathbf{D}_{\overline{\mathbf{z}}-\underline{\mathbf{z}}}\ \underline{\mathbf{z}}, \quad \mathbf{w}_{t,k} = \mathbf{D}_{\overline{\mathbf{z}}-\underline{\mathbf{z}}}\tilde{\mathbf{w}}_{t,k}.$$

Then, for all $\mathbf{z}^j$ we have

$$|f_k(\mathbf{z}^j) - g_k(\mathbf{z}^j)| < \epsilon_1. \tag{A.4}$$

This result holds for each of the K scalar elements of $\mathbf{F}(\mathbf{z})$. Therefore, by parallel concatenation of the K neural networks used to approximate the K scalar-valued functions, we obtain a shallow neural network that has K outputs and $(\sum_i T_i)$ neurons at the hidden layer. Setting $\epsilon = \sqrt{K}\ \epsilon_1$, we deduce that there exists a sum $\mathbf{g}_T(\mathbf{z})$ in the form of (3.1) that satisfies

$$\|\mathbf{F}(\mathbf{z}^j) - \mathbf{g}_T(\mathbf{z}^j)\|_2 < \epsilon \qquad \forall\ \mathbf{z}^j. \tag{A.5}$$

It is clear now that the parameters of this function $\mathbf{g}_T(\mathbf{z})$, i.e., $\boldsymbol{\alpha}_t$, $\mathbf{w}_t$, and β_t, achieve a zero cost function solving Problem (3.3), and hence is optimal in solving (3.3).

In addition, since an approximation can be realized using any sigmoid functions, the main result in [69] specifies that the minimum number of neurons required to achieve accuracy at least ϵ_1 for a scalar-valued function is given by

$$T = \mathcal{O}(\epsilon_1^{-\frac{L}{r}}) \tag{A.6}$$

where r denotes the number of continuous derivatives of the approximated function $f(\mathbf{z})$, and L represents the number of parameters of the function. In order to achieve

ϵ accuracy for approximating $\mathbf{F}(\mathbf{z})$, at least one of the real-valued functions that construct $\mathbf{F}(\mathbf{z})$ has to achieve $\frac{\epsilon}{\sqrt{K}}$. Hence, the complexity of shallow neural networks that optimally solve (3.3) for $\epsilon > 0$ is at least

$$T = \mathcal{O}\bigg(\Big(\frac{\epsilon}{\sqrt{K}}\Big)^{-\frac{L}{r}}\bigg).$$

□

Appendix B

Continuity of State Estimation Mapping: An Example Network

The neural networks are known as universal functions approximators. Nevertheless, the theoretical results on the ability to approximate function are usually limited to continuous functions. Hence, for continuous functions, the neural networks are expected to be able to achieve high approximation accuracy. Unfortunately, checking the continuity of the state estimation solution, which is an inverse mapping of a highly nonlinear function, is not simple to be checked.

In this appendix, a 3-bus balanced lossless network is presented, in Fig. B.1, in order to inspect the continuity of the state estimation mapping. We assume that the simple network has 3 buses and that the magnitude of the voltages are measured at all buses. In addition, the active and reactive power flows are measured at all lines. Since, the phase at Bus 1 can be taken as a reference for the other buses, the state estimation problem amounts to estimating the lines phase differences, or equivalently, the phases at Bus 2 and Bus 3.

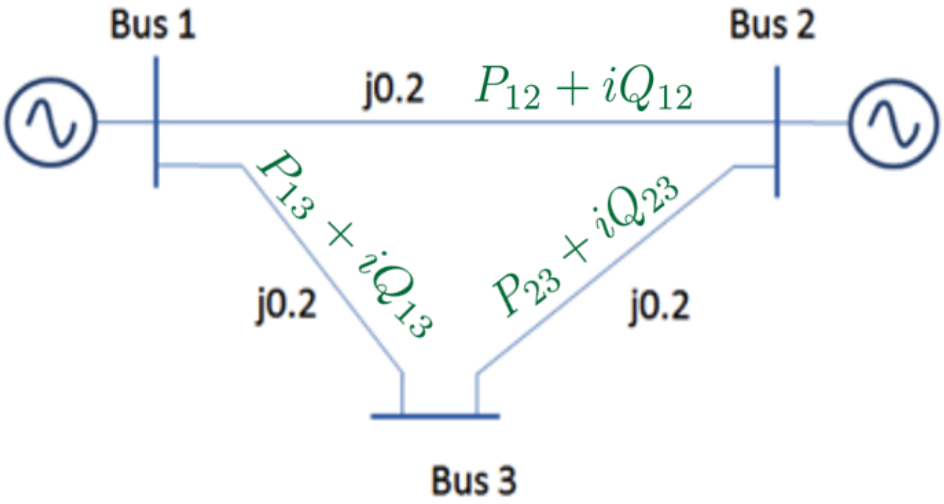


Figure B.1: Example 3-bus network

The power flow equations can be expressed as follows

$$P_{12} = B_{12}|v_1||v_2|\sin(\theta_{12}), \tag{B.1}$$

$$Q_{12} = |v_1|^2 - B_{12}|v_1||v_2|\cos(\theta_{12}), \tag{B.2}$$

$$P_{13} = B_{12}|v_1||v_3|\sin(\theta_{13}), \tag{B.3}$$

$$Q_{13} = |v_1|^2 - B_{12}|v_1||v_3|\cos(\theta_{13}) \tag{B.4}$$

where B_{ij} is the susceptance of the line between Bus i and Bus j, $|v_i|$ is the voltage magnitude at the i-th Bus, and θ_{ij} is the angle difference on the line (i, j). Assuming that the collected measurements are noiseless, the solution of the state estimation problem can be written in closed-form as

$$\theta_{12} = \sin^{-1}\left(\frac{P_{12}}{B_{12}|v_1||v_2|}\right), \tag{B.5}$$

$$\theta_{13} = \sin^{-1}\left(\frac{P_{13}}{B_{13}|v_1||v_3|}\right). \tag{B.6}$$

Claim 4. *The mapping between the measurements $P_{12}, P_{13}, |v_1|,$ and $|v_2|$ and the state*

of the network, i.e., θ_{12} and θ_{13}, is continuous if

$$B_{12}|v_1||v_2| \geq \epsilon, \text{ and} \qquad B_{13}|v_1||v_3| \geq \epsilon \tag{B.7}$$

for any $\epsilon > 0$.

Proof. The proof is straightforward and build upon basic results from real functions analysis. First, the function

$$f_1(P_{12}, |v_1|, |v_2|) = \frac{P_{12}}{B_{12}|v_1||v_2|} \tag{B.8}$$

is continuous on $B_{12}|v_1||v_2| \in [\epsilon, \infty]$ for any $\epsilon > 0$. Then, the mapping functions in (B.5) is composite function of f_1 and $\sin^{-1}(\cdot)$ which is a continuous function. Therefore, the mapping in (B.5) is continuous on $B_{12}|v_1||v_2| \in [\epsilon, \infty]$ for any $\epsilon > 0$ [85]. The same follows for θ_{13}. □

Appendix C

Construction of Quadratic Matrices for QCQP Formulation of ACOPF

The quadratic equality constraints (5.19) and (5.20) involve complex voltages $\mathbf{x}_v$, and currents $\mathbf{x}_i$ and $\mathbf{x}_\Delta$. As mentioned in Section III-B, the complex voltage $\tilde{\mathbf{x}}$ is constructed by concatenating the three vectors. In order to write the constraints (5.19) and (5.20) in the general form quadratic constraint $\tilde{\mathbf{x}}$, we need to construct quadratic matrices that represent the constraints. To this end, sort the lines in the network according the index of the receiving end of each line by a nondecreasing order. Note that, due to the tree structure assumed on the network, each bus except the substation appears only once as a receiving end. Let us define some transformation matrices as follows.

$$\mathbf{T}_{v_k} = [\mathbf{0}_{3\times 3k} \quad \mathbf{I}_{3\times 3} \quad \mathbf{0}_{3\times 3(n-k)} \quad \mathbf{0}_{3\times 3(n+1)} \quad \mathbf{0}_{3\times 3n}]^T$$
$$\mathbf{T}_{\Delta,k} = [\mathbf{0}_{3\times 3(n+1)} \quad \mathbf{0}_{3\times 3k} \quad \mathbf{I}_{3\times 3} \quad \mathbf{0}_{3\times 3(n-k)} \quad \mathbf{0}_{3\times 3n}]^T$$
$$\mathbf{T}_{i_{kj}} = [\mathbf{0}_{3\times 6(n+1)} \quad \mathbf{0}_{3\times 3(j-1)} \quad \mathbf{I}_{3\times 3} \quad \mathbf{0}_{3\times 3(n-j)}]^T$$
$$\mathbf{U}_{\Delta,k} = [\mathbf{T}_{v_k} \quad \mathbf{T}_{\Delta_k}]$$
$$\mathbf{U}_{kj} = [\mathbf{T}_{v_k} \quad \mathbf{T}_{i_{kj}}]$$

Hence, the vectors $\mathbf{x}_{\Delta,k}$ and $\mathbf{x}_{ki}$ are given by $\mathbf{U}_{\Delta,k}^T\tilde{\mathbf{x}}$ and $\mathbf{U}_{kj}^T\tilde{\mathbf{x}}$, respectively. Accordingly, the equality (5.19) can be re-written as

$$\begin{bmatrix} s_{\Delta,k}^{ab} \\ s_{\Delta,k}^{bc} \\ s_{\Delta,k}^{ca} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{x}}^H\mathbf{U}_{\Delta,k}\left(E_{41}-E_{42}\right)\mathbf{U}_{\Delta,k}^T\tilde{\mathbf{x}} \\ \tilde{\mathbf{x}}^H\mathbf{U}_{\Delta,k}\left(E_{52}-E_{53}\right)\mathbf{U}_{\Delta,k}^T\tilde{\mathbf{x}} \\ \tilde{\mathbf{x}}^H\mathbf{U}_{\Delta,k}\left(E_{63}-E_{61}\right)\mathbf{U}_{\Delta,k}^T\tilde{\mathbf{x}} \end{bmatrix}, \quad \forall k \in \mathcal{N}. \tag{C.1}$$

These three complex equalities can be written as six real quadratic equations involving the complex vector $\tilde{\mathbf{x}}$. Introduce the Hermitian matrices $\mathbf{Y}_{\Delta,k}^{\phi}$ for all $\phi \in \Phi_\Delta$ which are as follows.

$$\begin{aligned}
\mathbf{Y}_{\Delta,k}^{ab} &= \frac{1}{2}\mathbf{U}_{\Delta,k}((E_{41}-E_{42})+(E_{41}-E_{42})^T)\mathbf{U}_{\Delta,k}^T \\
\tilde{\mathbf{Y}}_{\Delta,k}^{ab} &= \frac{1}{2i}\mathbf{U}_{\Delta,k}((E_{41}-E_{42})-(E_{41}-E_{42})^T)\mathbf{U}_{\Delta,k}^T \\
\mathbf{Y}_{\Delta,k}^{bc} &= \frac{1}{2}\mathbf{U}_{\Delta,k}((E_{52}-E_{53})+(E_{52}-E_{53})^T)\mathbf{U}_{\Delta,k}^T \\
\tilde{\mathbf{Y}}_{\Delta,k}^{bc} &= \frac{1}{2i}\mathbf{U}_{\Delta,k}((E_{52}-E_{53})-(E_{52}-E_{53})^T)\mathbf{U}_{\Delta,k}^T \\
\mathbf{Y}_{\Delta,k}^{ca} &= \frac{1}{2}\mathbf{U}_{\Delta,k}((E_{63}-E_{61})+(E_{63}-E_{61})^T)\mathbf{U}_{\Delta,k}^T \\
\tilde{\mathbf{Y}}_{\Delta,k}^{ca} &= \frac{1}{2i}\mathbf{U}_{\Delta,k}((E_{63}-E_{61})-(E_{63}-E_{61})^T)\mathbf{U}_{\Delta,k}^T
\end{aligned}$$

Then, the symmetric matrix $\mathbf{B}_{\Delta,k}^{\phi}$ for $\phi \in \Phi_\Delta$ can defined as

$$\mathbf{B}_{\Delta,k}^{\phi} = \begin{bmatrix} \Re\{\mathbf{Y}_{\Delta,k}^{\phi}\} & -\Im\{\mathbf{Y}_{\Delta,k}^{\phi}\} \\ \Im\{\mathbf{Y}_{\Delta,k}^{\phi}\} & \Re\{\mathbf{Y}_{\Delta,k}^{\phi}\} \end{bmatrix}. \tag{C.2}$$

Similarly, we construct $\tilde{\mathbf{B}}_{\text{Delta},k}^{\phi}$ as function $\tilde{\mathbf{Y}}_{\Delta,k}^{\phi}$. Using the matrices $\mathbf{U}_{kj}$ and $\mathbf{T}_{v_k}$, one can follow the same steps to construct the symmetric matrices $\mathbf{B}_{\text{Y},k}^{\phi}$ and $\tilde{\mathbf{B}}_{\text{Y},k}^{\phi}$ for all $k \in \mathcal{N}$ and $\phi \in \Phi$.

Appendix D

Acronyms

Care has been taken in this dissertation to minimize the use of jargon and acronyms, but this cannot always be achieved. This appendix contains a table of acronyms and their meaning.

Table D.1: Acronyms

Acronym	Meaning
AC	Alternating Current
DC	Direct Current
DFS	Depth First Search
DNN	Deep Neural Network
DSSE	Distribution System State Estimation
ES	Energy Storage
FPP	Feasible Point Pursuit
MIPS	Matlab Interior Point Method
NN	Neural Network
OPF	Optimal Power Flow
PDIPM	Primal Dual Interior Point Method
PMU	Phasor Measurement Unit
PV	Photovoltaic

Continued on next page

Table D.1 – Continued from previous page

Acronym	Meaning
QC	Quadratic Convex
QCQP	Quadratically Constrained Quadratic Program
RES	Renewable Energy Source
SCA	Successive Convex Approximation
SCADA	Supervisory Control and Data Acquisition
SDR	Semidefinite Relaxation
SE	State Estimation
SOCP	Second Order Cone Program
TRALM	Trusted Region Augmented Lagrangian Multipliers
WLS	Weighted Least-Squares