Proposal Summary

EPCN: Collaborative Research: Learning and Optimizing Power Systems: A Geometric Approach, PI Yang Weng (ASU), Co-PI Baosen Zhang (UW)

The power distribution system is currently undertaking a dramatic transformation in both form and function. The large-scale development of technologies such as rooftop solar, electric vehicles, and smart home management systems has the potential to offer cheaper, cleaner and more controllable energy to the customers. Integrating these new resources of energy, however, has proven to be nontrivial, mainly because of their distributed nature and the inherent uncertainties of power variation. When compared to those of the transmission system, operators of distribution systems do not have access to the necessary information about the network to grasp an accurate picture. For one thing, the topologies of distribution systems are usually not known or are severely outdated, with some utilities still using paper-based maps. Second, the distribution system itself is, for the most part, not carefully monitored, which makes integrating distributed resources with large variabilities a particularly challenging task. Finally, lots of small outages are caused by failures or faults in the distribution system, and these have become increasingly common due to the aging infrastructure. Unless an operator adapts to accommodate these new trends and conditions, system performance and reliability will suffer severely in the long-term.

The question, therefore, is "How should operators in distribution systems adapt?" The answer lies in harnessing emerging tools to create a novel framework for distribution automation. New (heterogeneous) sources of data are being collected by smart meters and a range of sensor technologies. In addition, a renewed interest in the fundamentals of power flow has recently produced several new viewpoints on loadability. Also, when power outages occur, there are switches already in place that can be used for fast system recovery. This proposal will leverage both the technology and theory of the emerging new tools with insights in order to improve distribution system operations. Specifically, we seek to connect data, rigorous mathematical analysis, and engineering applications by facilitating novel algorithms with provable performance.

Intellectual Merit: This work significantly advances the current state of the art of topology estimation and a fundamental understanding of power flows in distribution networks. First, we use tools from probabilistic graphical models to propose a data-driven framework for topology estimation. The proposed framework is computationally efficient. It has provable performance guarantees and accommodates heterogeneous sources of data streams. Second, we present a geometric view of the feasibility of power flow solutions and use it to derive a new measure of the feasibility margin of an operating point. This approach fundamentally improves on current methods that are based on the power flow Jacobian matrix. We then use the results in a topology switching algorithm for system restoration in the event of power outages. This algorithm ensures that power can be securely restored through a set of switching actions that accounts for both uncertainties in the topology of the distribution system and variabilities in the distributed resources. Together, these tools enable seamless real-time operation of topology discovery, uncertainty management, and system recoveries during outages.

Broader Impact: Successful application of our proposed project will allow distribution system operators to answer various "what now" and "what if" questions deriving from those highly volatile grids with large amounts of distributed resources. It combines advanced techniques from data analytics, probability, power system operations, convex and non-convex optimization to realize more efficient and resilient system operations. Education curriculum developed from this project brings together students from different disciplines in applied mathematics, statistics, and electrical engineering. In particular, students will learn how data and advanced mathematics can be used to improve the efficiency of a basic infrastructure, namely, the power system, which is sometimes taken for granted. Consequently, this project enables the PIs to train a workforce for the future smart

cities. The PIs will also work closely with the industry (Centrica and Google) for implementation. Realistic data from existing NDA (Salt River Public Utility in Phoenix and Seattle City Light in Seattle) will also be used for validation.

PROJECT DESCRIPTION

1 Introduction

The transformations of the electrical grid present a plethora of challenges to system operators and utilities. They must adapt to manage a set of highly uncertain and distributed resources such as electric vehicles and wind farms, while at the same time operating a grid infrastructure that was designed decades ago [1–3]. These challenges are particularly acute in the distribution system, where the networks are traditionally not monitored closely and operators lack the essential information for them to grasp an accurate picture about the system. Also, the number of outages in the distribution system has started to increase because of the aging infrastructures and changes in the nature of the loads. The goal of this proposal is to overcome these challenges by *developing novel algorithms and new insights that increase the efficiency and resilience of the distribution systems.*

To achieve this goal, we focus on three thrusts: i) *system topology estimation* using the wealth of data made available by smart meters and other sensors, where the network may contain loops and the data may be highly heterogeneous; ii) *characterization of the feasibility of operating points* using a new geometric understanding of power flow that leads to provably efficient and optimal algorithms; and iii) *restoration of service right after outages through line switching* by using the results from the first two thrusts. These investigations bring in tools from power system analysis, optimization, and statistical learning to enable fundamental advances in the distribution system operations. In particular, these thrusts allow us to leverage recent advances in both technology and theory to develop timely and rigorous algorithms that solve some pressing engineering problems for the power grids. Below we discuss the practical significance of these thrusts and their respective technical challenges.

The first problem we consider is *identifying the topology of distribution systems*. A standard assumption in studies on optimization and control in power systems assumes that the topology and the line parameters of the system are given [4–7]. However, the topologies of distribution systems, especially the secondary distribution grid, are rarely known entirely. For example, many utilities use paper-based topology maps that may be badly outdated [8]. In addition, distribution networks undergo regular topology changes, ranging from once every four months [9], once every four weeks [10], to a few times a day if rooftop solar generation is coordinated through switching [11]. Lots of these changes are in fact not recorded [12]. Recently, the problem of topology identification has received intensive attention from the research community, see e.g., [13] and the references within. However, most of the existing algorithms have not been adopted in practice. Those algorithms either assume that a distribution system will be fully instrumented (which is rarely seen in the United States) or provide no guarantees on the results. The difficult combinatorial problems involved also make their adoption infeasible. Furthermore, many algorithms assume that the correct topology is radial, but distribution networks in dense downtown areas (e.g., Seattle and San Francisco) and microgrids often contain loops [14].

The main reason to learn the topology of a network is to understand and characterize the operation of the system. Specifically, one of the most fundamental problems in power system engineering is power flow feasibility: determining whether a particular set of power injections at the buses can be physically realized by a set of bus voltages, subject to operational constraints [15–17]. Because of the proliferation of distributed resources and their inherent uncertainties, the analysis in the stability of operating points has become increasingly useful in the distribution systems [18, 19]. Significant effort has been devoted to *quantifying the distance to the boundary of the feasible set of operating points*, because if a network is operating too close to the boundary, then sudden exogenous changes may lead to unstable behaviors [20–22]. Most of the existing algorithms fall into two categories. The first is to make the standard DC or decoupled power flow assumptions that lead to a system of linear equations [23]. However, these assumptions do not hold in the distribution system because of the high R/X ratios. The second is to adopt the full AC formulation. In this approach, a series of nonlinear algebraic equations need to be solved [24,25], which is computationally challenging and difficult to scale to large systems.

The two challenges above are fundamental and solving them would allow us to make contributions to a wide variety of problems. In this project, we apply them to a practical and pressing problem: *improving the resilience of distribution systems*. Recently, one of us (B. Zhang) was interviewed by a local TV station (KIRO 7, Seattle CBS affiliate) [26] to explain why Seattle had experienced numerous power outages in times of seemingly fair weather. In fact, the frustration of the citizens is backed up by data. Fig. 1 plots the total number of outages and the total duration of outages from 2001 to 2015 in Seattle, which shows an accelerated increase in the past few years. These outages are mostly caused by a single equipment failure (e.g., a transformer or a power line) and they tend to take many hours to restore. For example, in 2015 there were 872 outages with the average outage time of 8.5 hours[1]. Outages of this magnitude are not catastrophic, but they do significantly degrade the quality of life for the citizens and businesses.

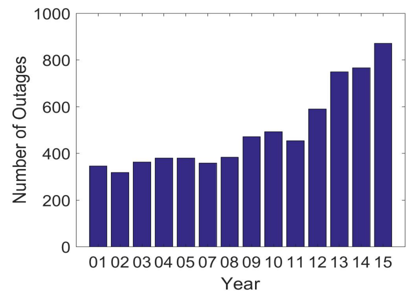


Figure 1: Number of outages in Seattle from 2001 to 2015. A clear increasing trend can be seen from the graph, going from 346 outages in 2001 to 872 outages in 2015.

One potential strategy to restore services faster is through *topology switching*, where lines are switched to create a new path for users disconnected by a fault [27]. Instead of waiting several *hours* for the repair crews to arrive, electricity can be restored in a matter of *seconds* [28]. However, to be practical, this method requires a fairly accurate estimation of the topology of the distributions system and a guaranty that reconnecting the load through another path would not adversely impact the robustness of the system [20, 29]. Such requirements are precisely the first two challenges considered in this project. Breakthroughs occuring in topology estimation and power flow feasibility then allow us to develop secure and efficient topology switching algorithms that can be directly applied in practice.

1.1 Research Directions

The central theme of this proposal is to develop rigorous algorithms that are both provably correct and efficient enough to be applied by practitioners in the field. It is based on developing new *geometric and graphical insights* for existing hard problems. Our research planned across three main thrusts:

Thrust 1: Topology Estimation. We describe a probabilistic framework to estimate the topology of a distribution system. This framework uses heterogeneous data from advanced metering infrastructure (AMI) [30], load side PMUs (micro-PMU) [31], and integrated sensing capabilities of some distributed resources (e.g., inverter and battery sensors) [32]. Specifically, we will build a graphical model of the network connectivity and formulate the topology estimation problem as minimizing a distance in the Kullback-Leibler (KL) divergence metric. We present an algorithm based on mutual information maximization that is computationally efficient and provides performance guarantees. Moreover, we show how this methodology can be applied to grids with loops. This topology estimation method serves as the basis for distribution grid optimization and control.

Thrust 2: Feasibility Margin. We will provide a fresh geometrical perspective on the classical power flow equations, which leads to conditions and metrics that quantify how close a power flow

[1] Statistics are calculated from logs provided by Seattle City Light.

solution is to the boundary of the feasible power flow region. Traditionally, the primary object of interest is the power flow Jacobian, which is the partial derivative of power with respect to voltage magnitude and phase angle. The determinant of the Jacobian provides local information around an operating point. If a point is on the boundary, the Jacobian is singular [33–35]. However, the converse is unnecessarily true, because a singular Jacobian does not necessarily imply that the corresponding point is on the boundary [36]. In this project, we go *beyond the Jacobian*, by explicitly looking at the geometry of the boundary points. We show that testing whether a solution is on the boundary can be considered as checking the feasibility of a *simple linear program*. Extending this intuition, we then propose a convex program to quantify the feasibility margin of a solution so that the system is robust to random fluctuations brought on by the newly integrated intermittent resources. In addition, the ideas presented here lead to a new algorithm for the AC power flow problem using fixed-point equations that have desirable contractive properties.

Thrust 3: Putting it Together: Topology Switching for Resilience. With an accurate topology estimation and an understanding of feasibility margin, we will develop a method to quickly and safely switch the topology of a distribution system for recovery from small to medium-sized outages. Line switching for power system recovery has been previously considered, for the most part, in the context of natural disasters [37], where many switches in the system must be considered, generating large-scale mixed-integer programs [38]. Here, we will take advantage of the fact that during a typical unplanned outage, most of the network is still connected and many switches do not need to be adjusted. We describe an optimization problem that addresses the challenge of guaranteeing that the power flow, after a switch, is within an acceptable feasibility margin with uncertainties from both topology estimation and distributed resources. We show how this problem can be solved efficiently using the tools developed from Thrusts 1 and 2, and the resulting algorithm can be easily implemented by utilities.

1.2 Validation and Implementation Plan

The two PIs bring distinct and complementary skill sets to this project. Y. Weng is an expert on state estimation and data analytics for distribution systems (see e.g., [8,13,39–48]) and B. Zhang is an expert on optimization and control of distribution systems (see e.g., [7,49–55]). The PIs already have a close working relationship from past collaborations [56–58]. Our team is unique in the sense that we have a *close relationship with utilities and companies in the Seattle and the Phoenix areas*, which ensures our ability to successfully translate the proposed algorithms to practice and implement the proposal ideas on real systems. Specifically, we take the following validation steps:

1. Both ASU and UW have access to real-time digital simulators (OPAL-RT, RTDS) that are capable of simulating large-scale power systems in detail. The team will leverage the equipment to perform realistic case studies of all of their algorithms, especially for research effort described in Thrust 2.
2. We have a strong relationship with commercial vendors and companies. For example, Centrica is an international energy and services company that operates across many regions in the world, including United States. Centrica is committed to working with us to deploy our algorithms, especially in Thrust 2 and Thrust 3, and provide feedback. Google is also a partner with us on the topology visualization in Thrust 1 on Google Map and other proposed tools based on Google platforms. Please see the attached letters of collaborations for more details.
3. We have existing NDAs on data sharing with local utilities in Phoenix, AZ (Salt River Project) and Seattle, WA (Seattle City Light). For Thrust 1, data from both utilities would be used to test our algorithms for topology estimation. For Thrust 3, we have detailed outage and repair logs from Seattle on each unexpected outage lasting more than a few minutes, and the logs provide a baseline comparison for the algorithms we develop in this project.

The rest of this proposal is organized as follows. Sections 2.1 to 2.3 describe each of the three thrusts in detail. Section 3 describes the educational and technological broader impact of the proposal. The attached collaboration plan provides the management framework for this proposal.

2 Research Plan

2.1 Topology Identification in Distribution Networks (Thrust 1)

An accurate estimation of a distribution system's topology is of fundamental importance to a myriad of applications. For example, almost all control and optimization algorithms developed in the recent years for distribution networks assume that either the full topology or at least the local connectivity around individual nodes is known [59, 60]. In Section 2.3 of this proposal, we consider the problem of topology switching to improve the resilience of distribution networks, but without an accurate estimation of the starting topology, the switching will be meaningless.

The exact topologies of distribution networks are often not known for lack of updated digitized maps (many substations only have paper-based maps) and frequent changes in topology that went unrecorded. Standard identification methods, such as the ones based on voltage estimation [61], are commonly developed for high-voltage transmission networks. These methods have been proven to be unsatisfactory in the distribution network because they are not supported by the limited sensor data or relatively frequent topological changes [62–65]. Dedicated methods for distribution networks based on state estimation [66, 67] or power flow [68] also assume the availability of admittance parameters and infrequent topology changes, which may not hold as distribution systems can switch fairly frequently. Lastly, most of the existing algorithms assume that the correct topology is radial, which is not always true in many urban networks and microgrids [14].

Fortunately, smart sensors are continuously being deployed across the distribution systems [69, 70], thanks to recent advances in communications, sensing, and targeted government investments. Some sensor examples include advanced metering infrastructure (AMI) and load side micro-PMUs (μ-PMUs) [31]. Additionally, private industries are integrating sensing capability into DERs for monitoring purposes, e.g., photovoltaic systems, commercial and residential charging systems, and in-home appliances such as thermostats.

Objective: In this project, we use these existing heterogeneous data streams as the basis of a topology estimation framework that: 1) does not need admittance information, 2) can tolerate for a large number of changes, and 3) includes loops in the topology.

In particular, we build a *probabilistic graphical model* of the power grid to characterize the interdependency between neighboring buses by thinking of the measurements as realizations of random variables. We use this model as a basis to solve the topology estimation problem.

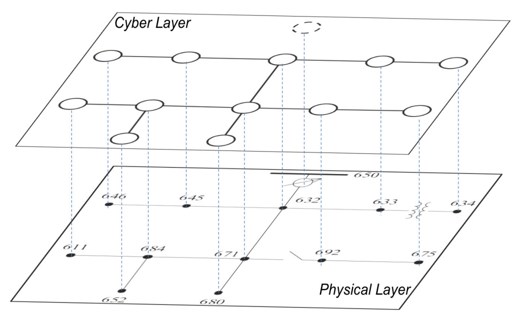


Figure 2: An example of a network with a physical layer and a cyber layer.

2.1.1 Framework: Probabilistic Graphical Modeling of Network Voltages

We use a graph to model the physical distribution network with n buses. We call this the *physical layer* to contrast with the *cyber layer* that models the data being collected in the network. Fig. 2 shows an example of this concept. We assume that the sensors in the network measure the voltages on each bus, and our goal is to construct the topology from a time series of these voltage measurements. We allow for both complex voltage (magnitude and angle) or only voltage magnitude measurements[2].

[2]In this project, we assume that bus voltage measurements are available. Currently, smart meters do measure bus currents, voltages, active and reactive power, sometimes only reports the active power. Therefore, voltage measure-

For each node in the system, we associate it with a random variable V_i (complex or real depending on the type of measurement) representing its voltage and a time series $v_i(t)$ representing the measured values. Let $\mathbf{V} = (V_1, \ldots, V_n)$ be the random vector of all bus voltages, and its joint probability distribution is $p(\mathbf{v}) = p(v_1, v_2, \ldots, v_{n-1}) = p(v_1)p(v_2|v_1)\ldots p(v_n|v_1, v_2, \ldots, v_{n-1})$. We assume that there is a reference bus in the network with fixed voltage at 1 per unit and angle $0°$. We label the reference bus as bus 0.

The true distribution $p(\mathbf{v})$ can be complex. The standard method is to discretize distribution, but the resulting high dimensional problem becomes intractable. Instead, our algorithm is based on approximating the true distribution $p(\mathbf{v})$ with another simplified distribution $p_a(\mathbf{v})$, which is much easier to learn. To measure the goodness of the approximation, we use the Kullback-Leiber (KL) divergence, which measures the difference between two probability distributions [71]: $D(p||p_a) = E_{p(\mathbf{v})} \log \frac{p(\mathbf{v})}{p_a(\mathbf{v})}$, where $p(\mathbf{v})$ comes from the measurement data and $p_a(\mathbf{v})$ is our candidate distribution. In fact, it is possible to show that minimizing the KL divergence between $p(\mathbf{v})$ and $p_a(\mathbf{v})$ is equivalent to maximizing the likelihood.

2.1.2 Proposed Approach: Maximizing Mutual Information

The approximate distributions $p_a(\mathbf{v})$ we consider are *probabilistic graphical models*, which are joint distributions with a correlation structure that follows a graph [72]. Let's start with considering a distribution system that operates with a tree topology. Then, we extend the algorithm to networks with loops. For a network with a tree topology, the natural class of graphical models is tree-based. This class of models is simple to work with because of its Markov property, which states that any two non-descendant nodes are independent, conditioning on their parents. Therefore, we can describe the joint probability as a product of pairwise conditional probability distribution: $p_a(\mathbf{v}) = \prod_{i=1}^{n} p(v_i|v_{\mathrm{pa}(i)})$, where $v_{\mathrm{pa}(i)}$ is the random variable designated as the direct predecessors or parent variable node of v_i in some orientation of the tree. Then, the problem of topology estimation becomes finding the *tree that minimizes the divergence between* $p_a(\mathbf{v})$ *and* $p(\mathbf{v})$.

To find this tree, we propose a *mutual information-based maximum weight spanning tree algorithm* shown in Algorithm 1. At an intuitive level, mutual information measures the mutual dependence between two variables. Mathematically, given two discrete random variables X and Y, their mutual information is given by: $I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right)$, where $p(x,y)$ is the joint distribution and $p(x)$ and $p(y)$ are the marginals. In our network, two nodes having a large empirical mutual information indicate strong correlation. It implies that the two nodes are likely to be connected. In essence, minimizing the KL-divergence of a tree-based graphical model to the measured data becomes finding a tree that maximizes the summed mutual information between connected nodes, leading to Algorithm 1:

Algorithm 1: A maximum mutual information-based greedy algorithm for constructing a spanning tree. This is the estimated topology given measurement data.

Data: Time series of voltage measurements at each bus.
Result: A spanning tree

Compute all pairwise empirical mutual information $I(V_i, V_j)$ from time series data $v(t)$.
Sort all pairs in decreasing order by $I(V_i, V_j)$.
Start with an empty tree as the maximum mutual information weighted spanning tree.
for *a pair* (i,j) *in the sorted list* **do**
    If adding the edge (i,j) to the tree does not result in a cycle, add (i,j) to the tree.
    Otherwise, do nothing

ments can be easily obtained via simple changes in the communication protocol.

This algorithm is computationally efficient since it only requires the computation of empirical mutual information, which can be calculated by looking at pairs of buses. In addition, this calculation can be extended to data sources with different rates by filtering techniques [73]. Fig. 3 shows an illustration of this procedure.

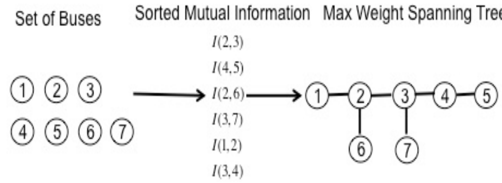


Figure 3: An example of applying Algorithm 1. We compute sorted (in descending order) mutual information, followed by constructing a spanning tree.

2.1.3 Planned Research Directions for Thrust 1

Research Task: Extend to Systems with Loops. In practice, some distribution systems have cycles to provide robustness [74–76]. However, this renders many algorithms that depend on the tree-topology assumption inappropriate. In our framework, the mutual information-based algorithm can be extended to handle the existence of loops. The key difference is to increase the number of parents that a node has. For example, if nodes can potentially have two parents, the probability distribution p_a can be written as $p_a(\mathbf{v}) = p(v_n|v_{\text{pa}(n),1}, v_{\text{pa}(n),2}) \prod_{i=2}^{n-1} p(v_i|v_{\text{pa}(i)})$, where $\{\text{pa}(n), 1\}$ and $\{\text{pa}(n), 2\}$ represent the two parent nodes of V_n. Then, instead of maximizing the mutual information to search for one parent, we maximize the *extended mutual information* between a node and a pair of possible parents. Algorithmically, for each extended mutual information $I(V_n; V_{\text{pa}(n),1}, V_{\text{pa}(n),2})$, we search for maximum weighted spanning tree and compute $\sum_{i=2}^{n-1} I(V_i; V_{\text{pa}(i)}) + I(V_n; V_{\text{pa}(n),1}, V_{\text{pa}(n),2})$. Subsequently, we compare and choose the topology with the largest value. Similarly, nodes with more parents can also be considered.

Research Task: Understand Computational Complexity. Since for each node we need to make n^2 calculations, the complexity of the Algorithm with loops is roughly on the order of n^3, where n is the total number of buses in the system. In general, the computational complexity would scale with the complexity of the network. If the network is potentially very complex and each node may have many parents, we need to compare a large number of mutual information values. However, most distribution networks only have a small possible number of loops. For example, many microgrids are designed with a central ring structure with trees hanging off it. In these instances, detecting the topology can be done with close to n^2 computations. Also, many buses in the network would never be part of a cycle, further reducing the computational complexity. In this project, we carefully study the conditions in distribution systems that would reduce the computational complexity of the algorithms while guaranteeing the performances.

Research Task: Show Provable Recovery. A key advantage of the algorithms proposed here is that they provide strong guarantees over successful topology identification. For example, under some mild technical condition, e.g., the one below, Algorithm 1 and its adaptation for loops always recover the correct topology if the estimated (empirical) mutual information is close to the true mutual information.

- Condition: Given a bus i and let j be a neighbor of i and k is not a neighbor of i, then $I(V_i, V_j) \geq I(V_i, V_k)$.

These results are similar in spirit to the Chow-Liu algorithm [77,78]. In our context, they state that maximizing the pairwise mutual information is equivalent to minimizing the KL divergence, and the minimum can be recovered. In this project, we extend these results further to include scenarios with missing and/or bad data, as well as to understand the behavior of the algorithms under frequently changing topologies. In these settings, the problem of mutual information estimation becomes interesting, and we seek to understand the data dependence of how quickly the empirical estimate approaches the true value. We have a well-founded suspicion that *using data from multiple heterogeneous sources would help* because of the diversity.

Research Task: Extend to Three-Phase Systems. In many distribution grids, the three phases are unbalanced. Even worse, some phase information is incorrect, e.g., a bus connected to phase B was labeled as on phase A. Therefore, we extend the graphical models to an unbalanced system by layering the phases on top of each other. With authorized access to RTDS and OPAL-RT equipment by both PIs, we leverage the real-time simulation capabilities at ASU and UW to conduct realistic studies using data provided by our partners and local utilities.

2.1.4 Implementation Plan

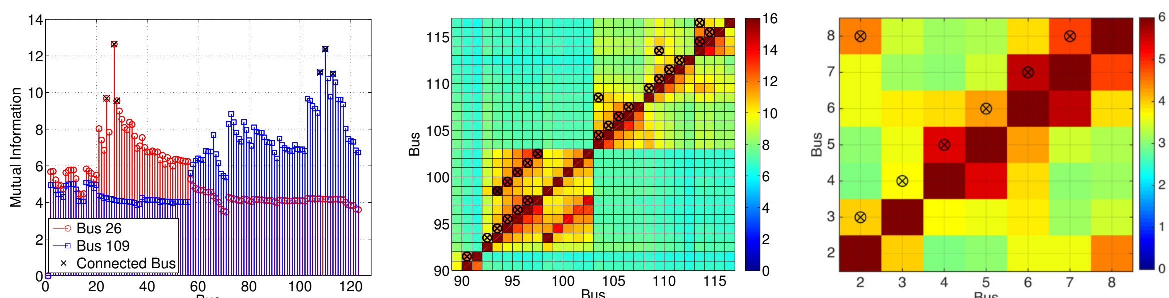
To implement and validate Algorithm 1 and its adaptation for loops, we execute the following plan:

1. Software simulations using standard test benchmarks [1] (see Section 2.1.5). GridLAB-D will be used for three phase simulation and validation.
2. Using existing RTDS and OPAL-RT at ASU and UW to collect realistic (noisy) measurements and to perform detailed simulations with hardware in the loop for testing our algorithms.
3. Using real data from Centrica and local utilities for validation. We have the nominal topologies (all switches are at their normal positions), voltage and load measurements from Phoenix and Seattle areas. The information can be programmed into the RTDS systems to simulate real-world implementations of our algorithms and serve as a convincing proof to field engineers.
4. The algorithms are implemented online with Google Map at the backend [79]. We will collaborate with Google to make this tool wide accessible (see the attached letter of collaboration).

2.1.5 Preliminary Results for Algorithm 1 and its Adaptation for Loops

Here, we show some preliminary results on the performance of Algorithm 1 and its adaptation for loops. We first validate the proposed topology identification method using the standard 123-bus distribution test feeder with PV [80,81]. In this simulation, we assume that 25 buses have unknown connections. We simulate the system to collect voltage measurements.

In Alg. 1, the first step is to calculate the mutual information between pairs of buses. Fig. 4a displays pairwise mutual information of bus 26 and bus 109. We see that the mutual information of a connected branch is much larger when compared with other branches. Fig. 4b shows a heat map of the mutual information matrix where buses 90 to 115 are included. Excluding the diagonal, the index with the highest mutual information on each row always corresponds to a physical branch. For a looped system, we add a loop to the standard IEEE 8-bus distribution network. The mutual information-based algorithm with loops achieves 100% accuracy in Fig. 4c.



(a) Pairwise mutual information for the 123-bus system

(b) Heat map for the 123-bus system

(c) Heat map for the 8-Bus System with a loop

Figure 4: Figure (a) shows the pairwise mutual information of buses with respect to bus 26 and bus 109 in the 123-bus system. Figure (b) shows the head map of the mutual information matrix for buses 90 to 115. Figure (c) shows the heat map of mutual information matrix for the 8-bus network with a loop. Black circle: true branches in the 123-bus network. Black crossing: detected branches. Algorithm 1 and its extended version for loops always reconstruct the correct topologies.

Finally, we compare detection error rate on the 123-bus system in comparison to two other state-of-the-art algorithms that we have implemented [44, 63] (these algorithms are not applicable to grids with loops). When there are 5 buses with unknown connections, the algorithm in [44] has an error rate (assigning an edge to buses that are not physically connected) of 67%, while the algorithm in [63] has an error rate of 40%. When the number of buses with unknown connections reaches 25, both algorithms in [44] and [63] essentially have 100% error rate, since some edges are always wrongly assigned. In contrast, our proposed Algorithm 1 achieves correct detection with an error rate of only 1%.

2.2 Voltage Stability of Power Flow Equations (Thrust 2)

The topology detection method serves as the basis for many applications in the distribution grid. In this section, we tackle a long-standing question in power systems: characterizing the power flow solutions. In particular, we focus on the static voltage stability of operating points[3].

Objectives: Taking a geometric view of the power flow problem to develop efficient and scalable algorithms to quantify how close a solution is to the boundary of the feasible region and solve the AC power flow problem.

2.2.1 Rectangular Representation of Power Flow Equations

The power flow equations have been the central object of interest in power system engineering for the past 60 plus years [83, 84]. These equations are well known to be nonlinear and nonconvex, making our understanding of them still incomplete despite many important breakthroughs. In this project, we use the rectangular coordinate-based power flow equations. Let the complex voltage at bus d be written as $v_{d,r} + jv_{d,i}$. Then, through some calculations, the power flow equations become:

$$p_d = t_{d,1} \cdot v_{d,r}^2 + t_{d,2} \cdot v_{d,r} + t_{d,1} \cdot v_{d,i}^2 + t_{d,3} \cdot v_{d,i}, \tag{1a}$$

$$q_d = t_{d,4} \cdot v_{d,r}^2 - t_{d,3} \cdot v_{d,r} + t_{d,4} \cdot v_{d,i}^2 + t_{d,2} \cdot v_{d,i}, \tag{1b}$$

where

$$t_{d,1} = -\sum_{k\in\mathcal{N}(d)} g_{kd},\ t_{d,2} = \sum_{k\in\mathcal{N}(d)} (v_{k,r}g_{kd} - v_{k,i}b_{kd}),\ t_{d,3} = \sum_{k\in\mathcal{N}(d)} (v_{k,r}b_{kd} + v_{k,i}g_{kd}),\ t_{d,4} = \sum_{k\in\mathcal{N}(d)} b_{kd},$$

where d is the bus index and $\mathcal{N}(d)$ denote the neighbors of bus d; n is the total number of buses; v_d is the complex phasor at bus d; p_d and q_d are the active and reactive power injections; g_{dk} and b_{dk} are the conductance and susceptance between bus d and bus k, and $y_{dk} = g_{dk} + j \cdot b_{dk}$. We adopt the convention that p_d (q_d) being positive means that bus d is consuming active (reactive) power and negative means bus d is supplying power.

Rectangular coordinates have been used in the past for problems such as fast load flow because the power flow equations are a set of homogenous quadratic forms in real variables when expressed by rectangular coordinates [85–88]. In this project, by further exploiting analytic properties of rectangular coordinates, we show that many nonlinear problems can be made even more computationally efficient. In this section, we first study how to measure the distance between power flow solution (solution to (1)) and the feasibility boundary. Then, we show some intriguing possibility of using the representation in (1) to extend the notion of the PV-curve and develop a new power flow method based on fixed-point equations. Again, we do not make the assumption that distribution systems are necessarily trees and we consider the full AC power flow formulation.

2.2.2 Boundary of the Power Flow Region

Given a power system, we define $\mathcal{P} \subset \mathbb{R}^n$ as the set of all possible active power vectors $p_1, \dots, p_n$ that can be achieved with some complex voltages. This object is called the feasible injection region [89]. For distribution networks, we take the feeder to the natural slack bus, labeled as bus 0, and set its voltage to be 1 per unit with angle 0°. For now, suppose all other buses are loads and modeled as PQ buses with fixed active and reactive demands. Let $\mathcal{P}^+$ be the region where all components are positive, that is, all buses are consuming active power[4].

[3] We focus on voltage stability since frequency is not a major concern in the distribution system or microgrids [82].

[4] First, we focus on the active power since most customers are charged for their active power consumption (e.g., $/KWh), and we treat reactive power as constraints. Second, the positivity on active power can be relaxed.

The *boundary* of the feasibility region $\mathcal{P}^+$ is defined as its Pareto-front. That is, a point $\mathbf{p} \in \mathcal{P}^+$ is on the boundary if there does not exist another point $\tilde{\mathbf{p}} \in \mathcal{P}^+$, where $\tilde{p}_i \geq p_i$ for all i with strict inequality in at least one of the coordinates. This boundary is of significant interest since it represents the limits in the operation of a distribution system [80]. For example, if a point is operating too close to the boundary, then under increased variability of distributed resources, it may be pushed outside of the feasibility region, leading to unstable behavior such as voltage collapse [59].

The traditional method of determining whether a point is on the boundary is through the power flow Jacobian. More specifically, this Jacobian is a matrix where the elements correspond to partial derivatives of active (or reactive) power with respect to voltage magnitudes and angles. By the Implicit Function Theorem, if a point is on the boundary, then the Jacobian matrix becomes singular. Unfortunately, the converse is unnecessarily true: a singular Jacobian does not imply that a point is on the boundary. Therefore, a new condition that is both *necessary and sufficient* is needed to determine whether a point is on the boundary.

Example. Consider the 3-bus fully connected network (triangle network). For simplicity of illustration, we assume that the network is purely resistive, the voltages are real, and we consider only the active powers. The blue area in Fig. 5 shows the feasible region $\mathcal{P}^+$ and the red curve in Fig. 5 shows the points where the Jacobian is singular. Since the Jacobian can be singular in the strict interior of the region, it is by itself insufficient to reliably test if a point is on the boundary. Similar insights carry over to a network with complex admittances and voltages.

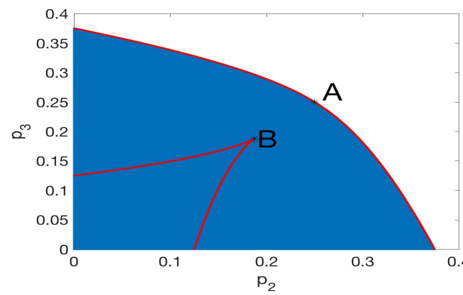


Figure 5: The feasible region and the power flow solutions where the Jacobian is singular for the 3-bus network. The red points denote operating points when the Jacobian is singular. They separate into two parts: the boundary of the region (e.g., point A) and points in the strict interior (e.g., point B).

Research Task: Determine the Boundary. Due to the insufficiency of using the singular Jacobian, a series of results have been developed to find and determine the boundary points [90–93]. However, these methods all require repeated trial and error computation of solving a set of nonlinear equations. In this proposal, we observe that by using (1), we can directly apply the geometric definition of the Pareto-point to obtain a condition that is equivalent to solving a set of linear inequalities. The key observation is that the partial derivatives become *linear functions* in the variables:

$$\frac{\partial p_d}{\partial v_{k,r}} = \begin{cases} 2t_{d,1}v_{d,r} + t_{d,2} & \text{if } k = d, \\ g_{kd}v_{d,r} + b_{kd}v_{d,i} & \text{if } k \neq d, \end{cases} \quad \frac{\partial p_d}{\partial v_{k,i}} = \begin{cases} 2t_{d,1}v_{d,i} + t_{d,3} & \text{if } k = d, \\ -b_{kd}v_{d,r} + g_{kd}v_{d,i} & \text{if } k \neq d, \end{cases} \tag{2}$$

$$\frac{\partial q_d}{\partial v_{k,r}} = \begin{cases} 2t_{d,4}v_{d,r} - t_{d,3} & \text{if } k = d, \\ -b_{kd}v_{d,r} + g_{kd}v_{d,i} & \text{if } k \neq d, \end{cases} \quad \frac{\partial q_d}{\partial v_{k,i}} = \begin{cases} 2t_{d,4}v_{d,i} + t_{d,2} & \text{if } k = d, \\ g_{kd}v_{d,r} - b_{kd}v_{d,i} & \text{if } k \neq d. \end{cases} \tag{3}$$

Note, $t_{d,1}$ and $t_{d,4}$ only depend on the line admittances (see (1)).

Let $\mathbf{h}_d$ be the gradient of p_d ($\mathbf{l}_d$ be the gradient of q_d) with respect to the real and imaginary parts of the voltages. From (2) and (3), these vectors are linear in the voltages. Then, checking whether a power flow solution is on the boundary is equivalent to checking if there is a direction such that moving the voltages in this direction would strictly increase at least one bus' active power without decreasing the others. Algorithmically, we solve the following feasibility problem:

Given a set of voltage vectors in rectangular coordinates, the corresponding power flow is on the boundary of the feasible injection region if and only if there does not exist a vector $\mathbf{x}$ *such that*

$$\mathbf{x}^T\mathbf{h}_d \geq 0,\ d = 1, \ldots, n, \ \text{ and } \ \mathbf{x}^T \sum_{d=1}^{n} \mathbf{h}_d = 1 \tag{4}$$

where $\mathbf{x}^T\mathbf{h}_d \geq 0$ is the condition that no active power can be decreased and $\mathbf{x}^T \sum_{d=1}^{n} \mathbf{h}_d = 1$ is satisfied as long as one of the inner products is strictly positive. The existence of a solution $\mathbf{x}$ via solving a linear program in (4) can be done easily with off-the-shelve algorithms [94].

2.2.3 Margin of Stability

The linear programming formulation in (4) presents an efficient method to check if a point is on the boundary, but characterizing the distance between a solution and the boundary is also of fundamental interest. In a system with uncertain distributed resources, this distance provides a safety *margin* of the operating point. If the margin is too small compared to the variability of the resources, the system runs the risk of instability.

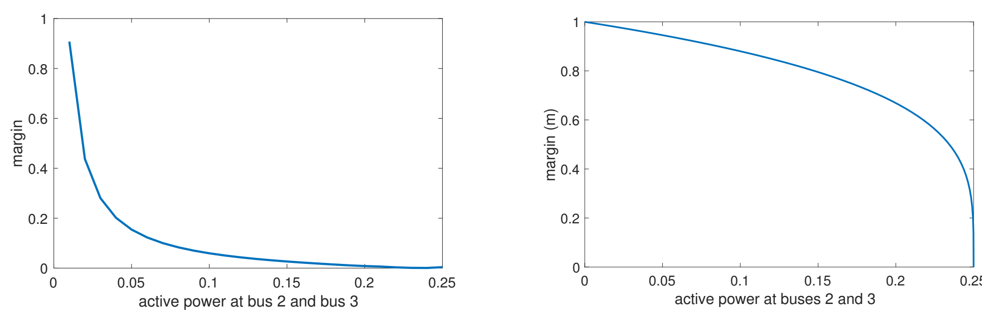
The traditional methods of gauging the margin either involve solving an optimal power flow problem or checking the determinant of the Jacobian. For a large system, resolving an optimal power flow problem may not be computationally desirable. The Jacobian-based method is typically easier to compute, and the condition number of a Jacobian can be used as a measure of the margin [36]. But from the example in Fig. 5, a set of interior points have singular Jacobians and looking at the condition number may misleadingly classify some interior regions as close to the boundary.

Research Task: Quantify the Stability Margin. In this proposal, we observe that the linear conditions in (4) can also be extended to measure how close a point is to the boundary. Essentially, for a point close to the boundary, it becomes harder to find a vector $\mathbf{x}$ that will satisfy (4). Then, instead of a feasibility problem, we consider the following optimization problem:

$$m = \underset{\mathbf{x}}{\text{maximize}}\ \mathbf{x}^T \sum_{d=1}^{n} \mathbf{h}_d \tag{5a}$$

$$\text{subject to } \mathbf{x}^T\mathbf{h}_d \geq 0\ d = 1, \ldots, n,\ ||\mathbf{x}|| \leq 1, \tag{5b}$$

where $\mathbf{h}_d$ are the gradients defined in (1) and m is the value of the optimization problem. This maximization problem essentially measures how "easy" it is to move the power injections in a positive direction under a norm constraint on $\mathbf{x}$. The further away from the boundary, the larger value m will be, which can be interpreted as providing a *margin* away from the boundary. Conversely, $1/m$ can be interpreted as the "risk" of the operating point, taking on a value of infinity for points on the boundary.



(a) Margin according to Thevenin method.

(b) Margin using the proposed method.

Figure 6: Margin comparison between the Thevenin equivalent method and our proposed method.

We compare the proposed method with the popular Thevenin equivalent method [95] for margin calculations. Using a resistive network, we plot the margin of stability as load increases in buses 1 and 2 in Fig. 6. Both margins approach 0 at the maximum possible load, but the Thevenin equivalent method is much more conservative than ours. For example, when the load is half of the maximum load, the Thevenin equivalent method has a relatively small margin, which may lead an operator to conclude that the load cannot be increased much more and operate conservatively. This would result in inefficiencies in operations, especially in an aging grid that is facing more complex loading environments [96]. In contrast, our method provides a much larger margin when the load is far away from the maximum, and the margin decreases rapidly once the load approaches the maximum. This allows operators to better gauge system states for more efficient grid operations.

Research Task: Add Constraints. There are various types of constraints in power system operations, from voltage magnitudes to active and reactive power injections, and any practical method need to incorporate these constraints. In this project, we show how the framework presented in this section would accommodate these constraints.

Box and linear constraints on the *active power* can be easily added to (4) and (5) by restricting the direction of movements. Given an operating point, for the constraints on active power that are tight, we can add constraints into (4) such that $\mathbf{y}$ must move the active powers to stay within the constraint. Similarly, *reactive power limits* can be included by adding more linear constraints. Incorporating *current* and *voltage* limits is straightforward. For example, in the problem in this subsection, we are interested in checking whether a voltage operating point is on the boundary of the feasible region. To include voltage constraints, we can simply add in these as bounds in the voltage space. Similarly, since a current is linear in voltage, current limits can be presented as constraints in the voltage space. After checking these constraints, we can apply the same arguments in this section to obtain a linear program.

2.2.4 A New Fixed Point Power Flow Algorithm

The rectangular coordinates in (1) present an additional geometric insight that leads to a new power flow calculation method on meshed AC grids that only require a sequence of trivial algebraic calculations. With increasing number of distribution systems and microgrids adding loops and the high R/X ratios, this method offers a faster and more robust solver for real-time power flow analysis.

The key observation is to note that for fixed constants $t_{d,1}, t_{d,2}, t_{d,3}, t_{d,4}$, (1a) and (1b) describe two circles in the $v_{d,r}$ and $v_{d,i}$ space. The centers and radii of these power flow "circles" are: the center of the active power flow circle is $(-t_{d,2}/2t_{d,1}, -t_{d,3}/2t_{d,1})$ for bus d and its radius decreases when p_d increases; the center of the reactive power flow circle is $(t_{d,3}/2t_{d,4}, -t_{d,2}/2t_{d,4})$ for bus d, and its radius decreases when q_d increases. Examples of these circles are shown in Figs. 7a and 7b. The intersection of the active and reactive powers are the possible power flow solutions [58].[5]

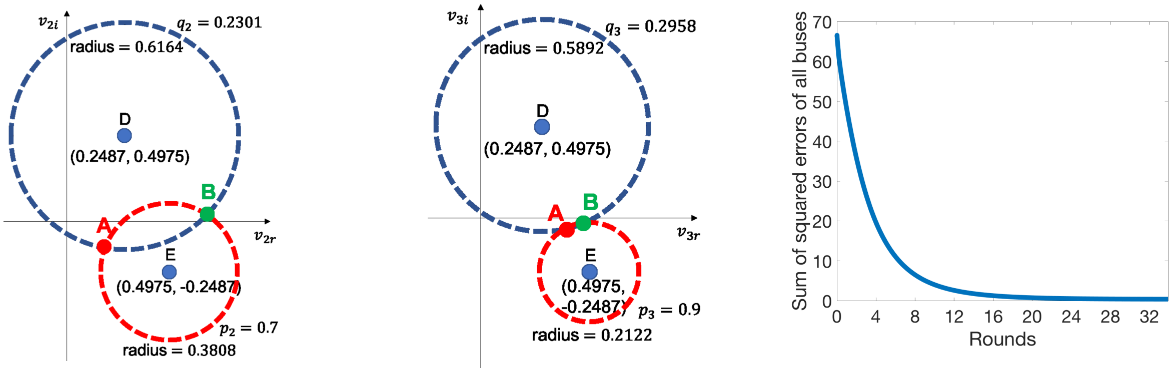
Thinking of power flow solutions as intersections of circles offers a surprising simple calculation strategy. For a network, suppose bus d is a PQ bus with given active power and reactive power. At a specific iteration step, the active and reactive circles can be determined from the current voltage values. Then we can update the complex voltages at bus d by finding the intersection points of the active and reactive circles. If there are more than one intersections, we take the one with the higher voltage magnitude. Then the neighbors of d can be updated in a similar fashion.

This update procedure is essentially solving a fixed point equation, where the algorithm terminates if none of the buses update their complex voltages in a round. Note each step in this algorithm only requires finding the intersection points of two circles (regardless of the size of the network), which can be done in an extremely small amount of time since the points can be found analytically. Fig. 7c shows the error to the true power flow solution of this algorithm on the 118-bus case, which converges in less than 20 rounds and less than a millisecond on a laptop.

Research Task: Guaranteeing Convergence In our current tests, the fixed point algorithm converges for almost all cases,[6] resulting in extremely fast calculations of the power flow problem. In this project, we seek to quantify the condition of when convergence is guaranteed. By existing hardness results, it cannot always converge in all cases [97], but current experiments suggest it

[5]Here we think of the buses as PQ buses, but PV buses can be visualized and analyzed very similarly.

[6]It always converges if some random restarts are allowed



(a) P and Q at bus 2.

(b) P and Q at bus 3.

(c) Convergence of the fixed-point algorithm.

Figure 7: Active and reactive circles for 118-bus system and the convergence rate of the algorithm.

is much more robust than the regular Newton method [98, 99] and the forward/backward sweep methods [100]. We conjecture that the fixed point equation is a contraction for a wide range of operating conditions.

2.2.5 Implementation Plan

For the results in this section, we adopt the following implementation plan:

1. We again leverage our access to the RTDS equipment to validate the results presented in this section. In the above discussions, preliminary results suggest our approach is more efficient and robust than current methods. However, for larger systems closer to the stability margin, they become harder to simulate in software due to numerical issues of existing solvers. Using RTDS with hardware in the loop, we can implement and assess the performance of our proposed methods under a variety of possible scenarios.
2. The use of RTDS specifically allows us to model different types of constraints that may exist in the distribution system and how they impact the stability margin of operating points. We will conducted detailed studies to understand the role that different constraints play on voltage stability, especially since buses are transforming from PQ loads to PV buses (distributed generation) and mixed/power electronic-based buses (e.g., electrical vehicles and photovoltaics.)
3. Both Phoenix and Seattle have experienced high growth in rooftop solar and electrical vehicles, although it is an open question how they have impacted the voltage stability of the grid. With data from local utilities and the topology detection result from Thrust 1, we will conduct studies that inform if the current grid can tolerate advancement of these new resources.

2.3 Putting it Together: Topology Switching for Resilience (Thrust 3)

The North American Electric Reliability Corporation (NERC) [101] enforces reliability standards that all electrical utilities must follow. However, despite its success, this *preventative approach* is starting to be insufficient in some instances. As shown in Fig. 1 in the introduction and other studies [76, 102], for cities in the United States, the number and duration of outages are starting to increase in recent years because of aging equipment and changes in load. In addition to replacing or building new infrastructures that may be prohibitively expensive, *leveraging existing switches in the network can dramatically shorten the duration of outages.*

Here, we focus on the *resilience* of the grid, not its reliability [103, 104]. We measure the resilience of a system by how quickly the electricity can be restored after an outage. This problem has been considered in the past by several studies [38, 105–108]. Most of these works focus on the *connectivity* of the network, where the topology of the network after a fault *is assumed to be known*, and switches are opened and closed to restore the network to a connected tree.

This proposal has argued that: 1) in practice, the topology is often not well known and 2) the loads are often uncertain and intermittent. This section builds on prior studies and our results from the first two thrusts are used to study the power restoration problem under uncertainties about the topology and the variations in the loads. This work bridges a gap between the algorithms developed in previous works and the current capability and operation paradigm of the utilities. Therefore, after obtaining an accurate data-driven topology estimation in Thrust 1 and topology-based power flow analysis in Thrust 2, an operational planning on switching can be used in operations.

2.3.1 Optimization Problem

We model faults as disconnecting existing lines in a network [105, 108]. An outage occurs if the resulting network is a disconnected graph or a part of the network becomes overloaded and is disconnected by protection equipment [109]. An example of this is shown in Fig. 8. Let $\mathbf{s}$ be an indicator vector of the status of the switches in the network. We assume that the locations of these switches and their status are estimated by the topology detection algorithm.

In Algorithm 1 (Section 2.1.2), we used mutual information to identify the best spanning tree explaining the collected time series data. However, we can easily expand this algorithm to provide *several* spanning trees that are the most likely candidates for the true topology of the system. For example, this can be done by considering two possible connections both with high mutual information. Therefore, to increase the robustness of switching decisions, we assume that *a set of candidates topologies* known right before the fault occurred. During the recovery switching process, we require that the switching actions result in a connected tree for every candidate topology. Note that once a topology is decided, the line admittance parameters can be easily learned especially when both voltage and power data are available [110, 111]. In addition, these parameters change slowly in the distribution system.

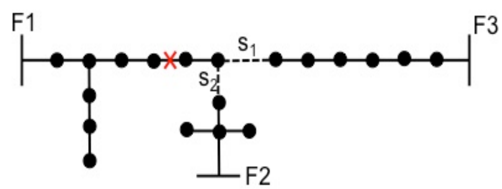


Figure 8: An example of a distribution with multiple feeders [108]. Dotted lines are normally open switches. The red cross shows the location of a fault. In this case, either s_1 or s_2 (but not both) can be closed to restore service.

Let k denote the set of feeders in the system. The connections between the buses and feeders are functions of the estimated topology and the status of the switches. For notational simplicity, let $f_t(\mathbf{s})$ be the system topology under the t'th estimated topology. Let $y_{i,k}$ be an indicator variable denoting if bus i is connected to feeder k and is a function of $f_t(\mathbf{s})$. Collecting these into a matrix, we write $\mathbf{Y}_t = f_t(\mathbf{s})$ as the indication matrix of bus connections to feeders under topology t and switch status $\mathbf{s}$. We assume that the complex power demands and voltages at all buses are known. For the buses that experienced an outage, we assume that this information is transmitted via a method like the last gasp mechanism of smart meters [112, 113] or via pseudo-measurements [114]. After an outage, we seek to restore services to the branch such that a weighted maximum feasibility margin is achieved. Notionally, let $m(\mathbf{Y}; \mathbf{v})$ denote the power flow feasibility margin (see (5)) under the topology $\mathbf{Y}$ and the complex voltage vector $\mathbf{v}$.

$$\text{maximize} \sum_t \alpha_t m(\mathbf{Y}_t; \mathbf{v}) \tag{6a}$$

$$\text{subject to} \sum_k y_{i,k,t} = 1 \ \forall i \ \forall t, \ y_{i,k,t} = 0, 1 \ \forall i, k, Y_t = f_t(\mathbf{s}), \tag{6b}$$

where the constraints in $\sum_k y_{i,k,t} = 1$ and $y_{i,k,t} = 0, 1$ enforce that a bus is connected to only one feeder under all of the estimated topologies. The objective function in (6a) maximizes the weighted sum of the feasibility margins, where the weight α_t is the probability of the i'th candidate topology.

2.3.2 Solution Methodology

From Section 2.2.3 and (5), the feasibility margin of a solution is given via a convex (quadratically constrained) linear program. Then, the optimization problem in (6) is a mixed-integer convex program with a linear objective function. It has somewhat of a stochastic flavor since multiple candidate topologies are considered. These types of problems can be solved via off-the-shelf optimization solvers [115, 116]. However, a direct implementation may not be efficient because of the large number of variables.

In our setting, we use the fact that the system is largely connected and only a small number of faults have occurred. Therefore, a large number of switch positions need not be changed and the number of solutions to be considered is not prohibitively large. This naturally leads to a branch-and-bound algorithm, where a large part of the solution spaces can be efficiently eliminated. For the initialization step of the branch-and-bound algorithm, we use the most likely estimated topology as a starting point to find a feasible solution. Then, we proceed by changing the positions of the switches through a depth-first search.

2.3.3 Planned Research Directions for Thrust 3

Research Task: Generalize the Result to Multiple Outages. The formulation in (6) requires that a set of switch positions satisfy the connectivity condition by $\sum_k y_{i,k,t} = 1 \ \forall i \ \forall t$ for a set of possible topologies. If there are many outages, this constraint may become too stringent. The question becomes how do we balance feasibility of the problem and maintain a high degree of confidence of the solution under topology uncertainty? In this project, we will investigate a risk-limiting [53, 117] or chance constraint formulation of the problem.

Research Task: Extend to Unbalanced Systems. The distribution system may not operate as a three-phase balanced system [1]. Therefore, it is important to consider the balance between different phases in the switching procedure. We believe this can be addressed by viewing the network as a multigraph where an edge is identified by its phase. Then, the goal is to find a switching solution that achieves a large feasibility margin in all three phases.

Research Task: Include Distributed Generation and Microgrids. We can extend this framework to a system with distributed generation and microgrids. For example, instead of feeders, we can imagine reconnecting a disconnected branch to a distributed generator to form a microgrid. Then, how should economics (cost of generation) be considered? How do one balance the tradeoff between heavily loading an existing feeder vs. using more distributed generation?

Research Task: Calculate Required Data Length. Right before an outage, historical data may give different topology candidates. The resolution of metered data also has an impact on the optimization (6). So, it is important to understand the best data resolution by using the topology information in Thrust 1 and 2. This raises an important setup for outage restoration: data quality.

2.3.4 Implementation Plan

To implement the switching detection algorithm (6) and evaluate the results from the research tasks, we adopt the following implementation plan:

1. Software simulations using standard test benchmarks [1] (see Section 2.1.5 for preliminary results.) and the new test cases at [118] for outages. Apply our algorithm for validation.
2. Using existing RTDS and OPAL-RT and ASU and UW to perform realistic simulations with hardware in the loop for outage scenarios. Apply our algorithm to validate the performance.
3. Using real data from utilities based on existing NDAs. Utilize detailed outage and repair logs from Seattle on each unexpected outage lasting more than a few minutes, and the logs provide a baseline comparison for the algorithms we develop in this section. Validate our algorithms.

3 Broader Impacts

Build on Existing Critical Infrastructure. Distribution systems are designed to be robust and simple enough such that they never need to be managed. This proposal allows utilities to adapt to high penetration of demand-side resources and increased variability by better using existing resources and leveraging the increase in data collected from various components. The algorithms presented in this proposal would lead to more efficient and reliable power systems without the need to first build new lines or installing dedicated sensors, directly benefiting the customers.

Integration of Education and Research. There is an urgent need to replenish the workforce in the area of power systems and train experts at the intersection of data analytics and energy. Currently, there are very few classes that integrate data analytics and power systems. This project offers us a unique opportunity to address the educational challenge by integrating rigorous probability, data analytics, and power systems. We plan to engage the public by exposing them to the results and approaches developed in the project. For high-school level, PI Weng is mentoring Singhania Dev from Hamilton High School in Chandler, Arizona. PI Weng also serves as IEEE PES Region 6 Scholarship Plus Committee to provide scholarship for undergraduates. At the graduate level, PI Weng co-chairs the IEEE PES Webinar Series on Big Data & Analytics for Power Systems.

PI Zhang actively participates in the engineering "redshirt" program, where students from under-privileged backgrounds (e.g., first generation students from rural communities in Eastern Washington) are given an extra year before they start studying at UW. PI Zhang has worked with Waynetta Dennison, a female Native American engineering student who conducted research on solar cell installation for her reservation in New Mexico. PI Zhang is also engaged with the broader community through the Seattle Science Center. Fig. 9 was taken during an exhibition for K-6 students, where the PI and his graduate student Chase Dowling (pictured) demonstrated the principle of max-flow min-cut using water flowing through pipes. We plan to continue these types of demonstrations, to connect science and engineering using fun and easy examples.

Figure 9: Demo.

The research in this proposal is especially suited for teaching and learning for students since the reliability of electrical grids can be conveyed using relatively simple examples. Also, power outages are especially keenly felt by students from underprivileged backgrounds. Students can easily participate in data collection and analysis, giving them a firm understanding of engineering. Also, much of the result developed in this proposal can be applied to rural areas, for example, the rural communities in South Eastern Arizona and Eastern Washington.

Open Source Algorithms. The PIs are also committed to sharing the resulting algorithms and data (after anonymization) with communities. Most of our algorithms will be open source and freely accessible online. The followings are some related tools: 1) demo with mutual information calculation [79], 2) MapD big data platform [120], and 3) data sources [121].

4 Prior Results from NSF Support

B. Zhang has had several NSF grants; his most closely related grant is (b) "CPS: Breakthrough: Collaborative Research: The Interweaving of Humans and Physical Systems: A Perspective from Power Systems" (a) (CNS 1544160, $250,000, 10/1/2015-9/30/2018, CO-PI: Ramesh Johari, Stanford University). (c) **Intellectual Merit:** This award is supporting a range of projects related to developing the theory of understanding the interaction between human users and physical systems, see (d) [57, 122–124]. **Broader Impact:** This grant is directly funding the work of an underrepresented Ph.D. student at the University of Washington. e) n/a (f) n/a.

Y. Weng has no prior NSF support.

References

[1] W. H. Kersting, "Distribution system modeling and analysis," in *Electric Power Generation, Transmission, and Distribution, Third Edition*. CRC press, 2012, pp. 1–58.

[2] Y. Tang, "Power distribution system planning with reliability modeling and optimization," *IEEE Transactions on Power Systems*, vol. 11, no. 1, pp. 181–189, 1996.

[3] P. Fox-Penner, *Smart power anniversary edition: Climate change, the smart grid, and the future of electric utilities*. Island Press, 2014.

[4] S. H. Low, "Convex relaxation of optimal power flowpart i: Formulations and equivalence," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 1, pp. 15–27, 2014.

[5] V. A. Venikov, V. A. Stroev, V. I. Idelchick, and V. I. Tarasov, "Estimation of electrical power system steady-state stability in load flow calculations," *IEEE Transactions on Power Apparatus and Systems*, vol. 94, no. 3, p. 1034, May 1975.

[6] S. H. Low, "Convex relaxation of optimal power flowpart ii: Exactness," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 2, pp. 177–189, 2014.

[7] B. Zhang and D. Tse, "Geometry of injection regions of power networks," *Power Systems, IEEE Transactions on*, vol. 28, no. 2, pp. 788–797, 2013.

[8] Y. Weng, R. Negi, and M. Ilic, "Graphical model for state estimation in electric power systems," *IEEE SmartGridComm Symposium*, Oct. 2013.

[9] E. Bueno, C. Lyra, and C. Cavellucci, "Distribution network reconfiguration for loss reduction with variable demands," in *Transmission and Distribution Conference and Exposition: Latin America, 2004 IEEE/PES*. IEEE, 2004, pp. 384–389.

[10] O. F. Fajardo and A. Vargas, "Reconfiguration of mv distribution networks with multicost and multipoint alternative supply, part ii: Reconfiguration plan," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 1401–1407, 2008.

[11] R. A. Jabr, "Minimum loss operation of distribution networks with photovoltaic generation," *IET Renewable Power Generation*, vol. 8, no. 1, pp. 33–44, 2014.

[12] A. Abur and A. G. Exposito, "Power system state estimation: Theory and implementation," *CRC Press*, Mar. 2004.

[13] Y. Weng, Y. Liao, and R. Rajagopal, "Distributed energy resources connectivity identification via graphical modeling," *IEEE Transactions on Power Systems*, no. 99, p. 1, Nov. 2016.

[14] R. E. Brown, "Impact of smart grid on distribution system design," in *2008 IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, July 2008, pp. 1–4.

[15] J. D. Glover, M. S. Sarma, and T. Overbye, *Power System Analysis & Design, SI Version*. Cengage Learning, 2012.

[16] C. S. Cheng and D. Shirmohammadi, "A three-phase power flow method for real-time distribution system analysis," *IEEE Transactions on Power Systems*, vol. 10, no. 2, pp. 671–679, 1995.

[17] R. D. Zimmerman and H.-D. Chiang, "Fast decoupled power flow for unbalanced radial distribution systems," *IEEE Transactions on Power Systems*, vol. 10, no. 4, pp. 2045–2052, 1995.

[18] D. K. Molzahn, B. C. Lesieutre, and C. L. DeMarco, "A sufficient condition for power flow insolvability with applications to voltage stability margins," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2592–2601, 2013.

[19] S. Grijalva and P. W. Sauer, "A necessary condition for power flow jacobian singularity based on branch complex flows," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, no. 7, pp. 1406–1413, 2005.

[20] J. W. Simpson-Porco, F. Dörfler, and F. Bullo, "Voltage collapse in complex power grids," *Nature communications*, vol. 7, 2016.

[21] M. Cespedes and J. Sun, "Renewable energy systems instability involving grid-parallel inverters," in *Applied Power Electronics Conference and Exposition, 2009. APEC 2009. Twenty-Fourth Annual IEEE*. IEEE, 2009, pp. 1971–1977.

[22] R. Al Abri, E. F. El-Saadany, and Y. M. Atwa, "Optimal placement and sizing method to improve the voltage stability margin in a distribution system using distributed generation," *IEEE transactions on power systems*, vol. 28, no. 1, pp. 326–334, 2013.

[23] K. Purchala, L. Meeus, D. V. Dommelen, and R. Belmans, "Usefulness of dc power flow for active power flow analysis," *IEEE Power & Energy Society General Meeting*, pp. 454–459, Jun. 2005.

[24] Z. C. Zeng, F. D. Galiana, B. T. Ooi, and N. Yorino, "Simplified approach to estimate maximum loading conditions in the load flow problem," *IEEE Transactions on Power Systems*, vol. 8, no. 2, p. 646, May 1993.

[25] C. Gomez-Quiles, A. Gomez-Exposito, and W. Vargas, "Computation of maximum loading points via the factored load flow," *IEEE Transactions on Power Systems*, 2015.

[26] KIRO7 News, "Seattle city light still doesnt know why power failed," News Report, 2017.

[27] Y. Xu, C.-C. Liu, K. P. Schneider, and D. T. Ton, "Toward a resilient distribution system," in *Power & Energy Society General Meeting, 2015 IEEE*. IEEE, 2015, pp. 1–5.

[28] H. Gao, Y. Chen, Y. Xu, and C.-C. Liu, "Resilience-oriented critical load restoration using microgrids in distribution systems," *IEEE Transactions on Smart Grid*, vol. 7, no. 6, pp. 2837–2848, 2016.

[29] E. Yamangil, R. Bent, and S. Backhaus, "Designing resilient electrical distribution grids," *arXiv preprint arXiv:1409.4477*, 2014.

[30] L. Wenpeng, "Advanced metering infrastructure," *Southern Power System Technology*, vol. 3, no. 2, pp. 6–10, 2009.

[31] A. Von Meier, D. Culler, A. McEachern, and R. Arghandeh, "Micro-synchrophasors for distribution systems," in *Innovative Smart Grid Technologies Conference (ISGT), 2014 IEEE PES*. IEEE, 2014, pp. 1–5.

[32] Solar City, "Poweguide app," http://www.solarcity.com/residential/energymonitoring- system, 2015.

[33] W. Kaplan, *Advanced calculus.* Addison-Wesley Reading, MA, 1952, vol. 4.

[34] I. A. Hiskens and R. J. Davy, "Exploring the power flow solution space boundary," *IEEE Transactions on Power Systems*, vol. 16, no. 3, pp. 389–395, 2001.

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

[36] P. Sauer and M. Pai, "Power system steady-state stability and the load-flow jacobian," *IEEE Transactions on power systems*, vol. 5, no. 4, pp. 1374–1383, 1990.

[37] Y. Wang, C. Chen, J. Wang, and R. Baldick, "Research on resilience of power systems under natural disastersa review," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 1604–1613, 2016.

[38] C. Chen, J. Wang, F. Qiu, and D. Zhao, "Resilient distribution system by microgrids formation after natural disasters," *IEEE Transactions on smart grid*, vol. 7, no. 2, pp. 958–966, 2016.

[39] Y. Weng, R. Negi, Q. Liu, and M. D. Ilić, "Robust state-estimation procedure using a least trimmed squares pre-processor," *IEEE Innovative Smart Grid Technologies*, pp. 1–6, Jan. 2011.

[40] Y. Weng, Q. Li, R. Negi, and M. D. Ilić, "Semidefinite programming for power system state estimation," *IEEE Power and Energy Society General Meeting*, Jul. 2012.

[41] Y. Weng, R. Negi, and M. Ilic, "A search method for obtaining initial guesses for smart grid state estimation," *IEEE SmartGridComm Symposium*, Dec. 2012.

[42] Y. Weng, M. D. Ilić, Q. Li, and R. Negi, "Convexification of bad data and topology error detection and identification problems in AC electric power systems," *IET Generation, Transmission & Distribution*, 2015.

[43] J. Yu, Y. Weng, and R. Rajagopal, "Probabilistic estimation of the potentials of intervention-based demand side energy management," *IEEE SmartGridComm Symposium*, Nov. 2015.

[44] Y. Liao, Y. Weng, M. Wu, and R. Rajagopal, "Distribution grid topology reconstruction: An information theoretic approach," *North American Power Symposium*, Oct. 2015.

[45] Y. Liao, Y. Weng, and R. Rajagopal, "Urban distribution grid topology reconstruction via lasso," *IEEE Power and Energy Society General Meeting*, Jul. 2016.

[46] Y. Weng, M. D. Ilic, Q. Li, and R. Negi, "Distributed algorithms for convexified bad data and topology error detection and identification problems," *International Journal of Electrical Power and Energy Systems*, vol. 83, p. 241, Feb. 2016.

[47] Y. Weng, R. Negi, C. Faloutsos, and M. D. Ilic, "Robust data-driven state estimation for smart grid," *IEEE Transactions on Smart Grids*, no. 99, p. 1, Jan. 2016.

[48] Y. Liao, Y. Weng, C. Tan, and R. Rajagopal, "Urban distribution grid line outage identification," *International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, Oct. 2016.

[49] G. Cezar, R. Rajagopal, and B. Zhang, "Stability of interconnected DC converters," in *Proc. 54th IEEE Conf. Decision and Control (CDC)*, Dec. 2015, pp. 9–14.

[50] Y. Kanoria, A. Montanari, D. Tse, and B. Zhang, "Distributed storage for intermittent energy sources: Control design and performance limits," in *Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on.* IEEE, 2011, pp. 1310–1317.

[51] B. Zhang and D. Tse, "Geometry of feasible injection region of power networks," in *Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on.* IEEE, 2011, pp. 1508–1515.

[52] B. Zhang, A. D. Dominguez-Garcia, and D. Tse, "A local control approach to voltage regulation in distribution networks," in *Proc. North American Power Symp. (NAPS)*, Sep. 2013, pp. 1–6.

[53] B. Zhang, R. Rajagopal, and D. Tse, "Network risk limiting dispatch: Optimal control and price of uncertainty," *IEEE Transactions on Automatic Control*, vol. 59, no. 9, pp. 2442–2456, Sep. 2014.

[54] B. Zhang, A. Lam, A. D. Domínguez-García, and D. Tse, "An optimal and distributed method for voltage regulation in power distribution systems," *Power Systems, IEEE Transactions on*, vol. 30, no. 4, pp. 1714–1726, 2015.

[55] B. Zhang, R. Johari, and R. Rajagopal, "Competition and coalition formation of renewable power producers," *IEEE Transactions on Power Systems*, vol. 30, no. 3, pp. 1624–1632, May 2015.

[56] P. Li, B. Zhang, Y. Weng, and R. Rajagopal, "Autoregressive model for individual consumption data-lasso selection and significance test," *arXiv preprint arXiv:1511.01853*, 2015.

[57] ——, "Autoregressive model for individual consumption data - sparsity recovery and significance test," in *Proc. IEEE Power and Energy Society General Meeting (PESGM)*, Jul. 2016, pp. 1–5.

[58] Y. Weng, R. Rajagopal, and B. Zhang, "Geometric understanding of the stability in power flow solutions," *IEEE Transaction on Power Systems*, 2017, in revision.

[59] A. Wood and B. Wollenberg, *Power generation, operation, and control*, 2nd ed., 1996.

[60] A. G. Exposito, A. Abur, A. V. Jaen, and C. G. Quiles, "A multilevel state estimation paradigm for smart grids," *Proceedings of the IEEE*, p. 952, Jun. 2011.

[61] R. Lugtu, D. Hackett, K. Liu, and D. Might, "Power system state estimation: Detection of topological errors," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-99, no. 6, pp. 2406–2411, Nov. 1980.

[62] M. R. Dorostkar-Ghamsari, M. Fotuhi-Firuzabad, M. Lehtonen, and A. Safdarian, "Value of distribution network reconfiguration in presence of renewable energy resources," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 1879–1888, 2016.

[63] S. Bolognani, N. Bof, D. Michelotti, R. Muraro, and L. Schenato, "Identification of power distribution network topology via voltage correlation analysis," in *IEEE 52nd Annual Conference on Decision and Control*, 2013, pp. 1659–1664.

[64] S. Xu, R. C. de Lamare, and H. V. Poor, "Dynamic topology adaptation for distributed estimation in smart grids," in *IEEE 5th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, 2013, pp. 420–423.

[65] J. Huang, V. Gupta, and Y.-F. Huang, "Electric grid state estimators for distribution systems with microgrids," in *46th Annual Conference on Information Sciences and Systems*, 2012, pp. 1–6.

[66] G. N. Korres and N. M. Manousakis, "A state estimation algorithm for monitoring topology changes in distribution systems," in *2012 IEEE Power and Energy Society General Meeting*. IEEE, 2012, pp. 1–8.

[67] M. Baran, J. Jung, and T. McDermott, "Topology error identification using branch current state estimation for distribution systems," in *IEEE Transmission & Distribution Conference & Exposition: Asia and Pacific*, 2009, pp. 1–4.

[68] R. Arghandeh, M. Gahr, A. von Meier, G. Cavraro, M. Ruh, and G. Andersson, "Topology detection in microgrids with micro-synchrophasors," *arXiv preprint arXiv:1502.06938*, 2015.

[69] Y. Weng and R. Rajagopal, "Probabilistic baseline estimation via gaussian process," *IEEE Power and Energy Society General Meeting (PESGM)*, Jul. 2015.

[70] J. Yu, Y. Weng, and R. Rajagopal, "Probabilistic estimation of the potentials of intervention-based demand side energy management," *IEEE SmartGridComm Symposium (SGC)*, Nov. 2015.

[71] T. M. Cover and J. A. Thomas, *Elements of information theory.* John Wiley & Sons, 2012.

[72] J. Whittaker, *Graphical models in applied multivariate statistics.* Wiley Publishing, 2009.

[73] R. H. Shumway and D. S. Stoffer, *Time series analysis and its applications: with R examples.* Springer Science & Business Media, 2010.

[74] D. Shirmohammadi and H. W. Hong, "Reconfiguration of electric distribution networks for resistive line losses reduction," *IEEE Transactions on Power Delivery*, vol. 4, no. 2, pp. 1492–1498, 1989.

[75] G. Celli, F. Pilo, G. Pisano, V. Allegranza, R. Cicoria, and A. Iaria, "Meshed vs. radial mv distribution network in presence of large amount of dg," in *Power Systems Conference and Exposition, 2004. IEEE PES.* IEEE, 2004, pp. 709–714.

[76] C. Rudin, S. Ertekin, R. Passonneau, A. Radeva, A. Tomar, B. Xie, S. Lewis, M. Riddle, D. Pangsrivinij, J. Shipman, and T. McCormick, "Analytics for power grid distribution reliability in new york city," *Interfaces*, vol. 44, no. 4, pp. 364–383, 2014.

[77] C. Chow and C. Liu, "Approximating discrete probability distributions with dependence trees," *IEEE Transactions on Information Theory*, vol. 14, no. 3, pp. 462–467, 1968.

[78] V. Y. Tan, A. Anandkumar, L. Tong, and A. S. Willsky, "A large-deviation analysis of the maximum-likelihood learning of markov tree structures," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1714–1735, 2011.

[79] Y. Weng, "Topology identification on google map," *http://www.public.asu.edu/ yweng2/software_top.html*, 2017.

[80] W. H. Kersting, "Radial distribution test feeders," *IEEE Power Engineering Society Winter Meeting*, vol. 2, pp. 908–912, 2001.

[81] A. P. Dobos, "PVWatts version 5 manual," *National Renewable Energy Laboratory, September*, 2014.

[82] H. Laaksonen, P. Saari, and R. Komulainen, "Voltage and frequency control of inverter based weak lv network microgrid," in *IEEE International Conference on Future Power Systems*, 2005, pp. 1–6.

[83] S. Frank, I. Steponavice, and S. Rebennack, "Optimal power flow: a bibliographic survey i," *Energy Systems*, vol. 3, no. 3, pp. 221–258, 2012.

[84] M. B. Cain, R. P. Oneill, and A. Castillo, "History of optimal power flow and formulations," *Federal Energy Regulatory Commission*, pp. 1–36, 2012.

[85] J. K. R. Y. H. Wei, H. H. Sasaki, "An interior point nonlinear programming for optimal power flow problems with a novel data structure," *IEEE Transactions on Power Systems*, vol. 13, no. 3, pp. 870–877, 1998.

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

[87] S. Iwamoto and Y. Tamura, "A fast load flow method retaining nonlinearity," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-97, no. 5, pp. 1586–1599, Sep. 1978.

[88] M. D. Schaffer and D. J. Tylavsky, "A nondiverging polar-form newton-based power flow," *IEEE Transactions on Industry Applications*, vol. 24, no. 5, pp. 870–877, Sep. 1988.

[89] B. Zhang and D. Tse, "Geometry of feasible injection region of power networks," *2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 1508–1515, Sep. 2011.

[90] T. Van Cutsem, "An approach to corrective control of voltage instability using simulation and sensitivity," *IEEE Transactions on Power Systems*, vol. 10, no. 2, pp. 616–622, 1995.

[91] F. Capitanescu and T. V. Cutsem, "Unified sensitivity analysis of unstable or low voltages caused by load increases or contingencies," *IEEE Transactions on Power Systems*, vol. 20, no. 1, pp. 321–329, Feb 2005.

[92] C. Hamon, M. Perninge, and L. Salder, "A stochastic optimal power flow problem with stability constraints part i: Approximating the stability boundary," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1839–1848, May 2013.

[93] M. Perninge and L. Soder, "On the validity of local approximations of the power system loadability surface," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2143–2153, Nov 2011.

[94] G. Dantzig, *Linear programming and extensions.* Princeton university press, 2016.

[95] K. Vu, M. M. Begovic, D. Novosel, and M. M. Saha, "Use of local measurements to estimate voltage-stability margin," *IEEE Transactions on Power Systems*, vol. 14, no. 3, pp. 1029–1035, Aug 1999.

[96] E. L. M. Vaiman, S. Maslennikov and X. Luo, "Calculation and visualization of power system stability margin based on pmu measurements," in *IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2010, pp. 31–36.

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

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

[99] F. Zhang and C. S. Cheng, "A modified newton method for radial distribution system power flow analysis," *IEEE Transactions on Power Systems*, vol. 12, no. 1, pp. 389–397, 1997.

[100] U. Eminoglu and M. H. Hocaoglu, "Distribution systems forward/backward sweep-based power flow algorithms: a review and comparison study," *Electric Power Components and Systems*, vol. 37, no. 1, pp. 91–110, 2008.

[101] V. Sulzberger and T. Gallagher, "Reliability and security: the nerc new standards development process," *IEEE Power and Energy Magazine*, vol. 2, no. 2, pp. 56–61, 2004.

[102] C. Rudin, D. Waltz, R. N. Anderson, A. Boulanger, A. Salleb-Aouissi, M. Chow, H. Dutta, P. Gross, B. Huang, S. Ierome, D. Isaac, A. Kressner, R. J. Passonneau, A. Radeva, and L. Wu, "Machine learning for the new york city power grid," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 2, pp. 328–345, 2012.

[103] D. Kirschen and F. Bouffard, "Keeping the lights on and the information flowing," *IEEE Power and Energy magazine*, vol. 7, no. 1, 2009.

[104] M. Panteli and P. Mancarella, "The grid: Stronger, bigger, smarter?: Presenting a conceptual framework of power system resilience," *IEEE Power and Energy Magazine*, vol. 13, no. 3, pp. 58–66, 2015.

[105] L. Yutian, F. Rui, and V. Terzija, "Power system restoration: a literature review from 2006 to 2016," *Journal of Modern Power Systems and Clean Energy*, vol. 4, no. 3, pp. 332–341, 2016.

[106] Z. Wang, J. Wang, and C. Chen, "A three-phase microgrid restoration model considering unbalanced operation of distributed generation," *IEEE Transactions on Smart Grid*, 2016.

[107] R. Sarkar, A. Gusrialdi, and Z. Qu, "A restorative strategy for resilient unbalanced power distribution networks," in *North American Power Symposium (NAPS), 2016*. IEEE, 2016, pp. 1–6.

[108] J. Li, X.-Y. Ma, C.-C. Liu, and K. P. Schneider, "Distribution system restoration with microgrids using spanning tree search," *IEEE Transactions on Power Systems*, vol. 29, no. 6, pp. 3021–3029, 2014.

[109] H. J. A. Ferrer and E. O. Schweitzer, *Modern Solutions for Protection, Control, and Monitoring of Electric Power Systems*. Schweitzer Engineering Laboratories, 2010.

[110] G. Andersson, "Modelling and analysis of electric power systems," *ETH Zurich, september*, 2008.

[111] Y. Weng, R. Negi, C. Faloutsos, and M. D. Ilić, "Robust data-driven state estimation for smart grid," *IEEE Transactions on Smart Grid*, 2016.

[112] H. Tram, "Technical and operation considerations in using smart metering for outage management," in *Transmission and Distribution Conference and Exposition, 2008. T&D. IEEE/PES.* IEEE, 2008, pp. 1–3.

[113] A. Singh, J. Bapat, and D. Das, "Two tier communication architecture for smart meter," in *Communication Systems and Networks (COMSNETS), 2013 Fifth International Conference on.* IEEE, 2013, pp. 1–2.

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

[115] D. Bienstock, "Computational study of a family of mixed-integer quadratic programming problems," *Mathematical programming*, vol. 74, no. 2, pp. 121–140, 1996.

[116] M. Ponnavaikko, K. S. P. Rao, and S. S. Venkata, "Distribution system planning through a quadratic mixed integer programming approach," *IEEE Power Engineering Review*, vol. PER-7, no. 10, p. 55, Oct. 1987.

[117] P. P. Varaiya, F. F. Wu, and J. W. Bialek, "Smart operation of smart grid: Risk-limiting dispatch," *Proceedings of the IEEE*, vol. 99, no. 1, pp. 40–57, 2011.

[118] Better Grids, "Better grids' test cases," *http://bettergrids.org*, 2017.

[119] Arizona State University, "American indian student support services," Online, 2017.

[120] MapD, "Data analytics and parallel processing platform," *10.206.203.195:9092, Due to hosting restrictions, it requires a registration process, and the interested readers can contact the authors for free access. We are in the process of obtaining a public static IP address that would be openly available (without registration), and we anticipate it would be ready before the end of 2017.*, 2017.

[121] Y. Weng, "Data sources," *http://www.public.asu.edu/ yweng2/source.html*, 2017.

[122] P. Li and B. Zhang, "Estimating treatment effects in demand response," in *Proc. IEEE Statistical Signal Processing Workshop (SSP)*, Jun. 2016, pp. 1–5.

[123] C. Riquelme, R. Johari, and B. Zhang, "Online active learning via thresholding," in *Proceedings of AAAI Conference on Artificial Intelligence, Acceptance Rate 25%*, 2017.

[124] P. Li and B. Zhang, "An optimal treatment assignment strategy to evaluate demand response effect," in *Proceedings of the Allerton Conference on Communication, Control and Computing*, 2017.