# Distribution Grid Line Outage Identification with Unknown Pattern and Performance Guarantee

Chenhan Xiao, *Student Member, IEEE*, Yizheng Liao, *Member, IEEE*, Yang Weng, *Senior Member, IEEE*

*Abstract*—Line outage identification in distribution grids is essential for sustainable grid operation. In this work, we propose a practical yet robust detection approach that utilizes only readily available voltage magnitudes, eliminating the need for costly phase angles or power flow data. Given the sensor data, many existing detection methods based on change-point detection require prior knowledge of outage patterns, which are unknown for real-world outage scenarios. To remove this impractical requirement, we propose a data-driven method to learn the parameters of the post-outage distribution through gradient descent. However, directly using gradient descent presents feasibility issues. To address this, we modify our approach by adding a Bregman divergence constraint to control the trajectory of the parameter updates, which eliminates the feasibility problems. As timely operation is the key nowadays, we prove that the optimal parameters can be learned with convergence guarantees via leveraging the statistical and physical properties of voltage data. We evaluate our approach using many representative distribution grids and real load profiles with 17 outage configurations. The results show that we can detect and localize the outage in timely manner with only voltage magnitudes and without assuming the prior knowledge of outage patterns.

## I. INTRODUCTION

Distribution grid line outage occurrence detection and localization is essential for efficient system monitoring and sustainable system operation [1]. A timely identification of the line outage effectively reduces potential financial loss. According to the U.S. Energy Information Administration, customers had an average of 1.3 outages and went without power for four hours during 2016 [2]. The frequency and severity of line outages caused by extreme weather events and power supply shortages have also increased in recent years.

The traditional line outage identification in distribution grids relies on passive feedback from customer reporting [3] or the "last gasp" message from smart meters [4], which is a notification automatically transmitted to the utility when power to the meter is lost. However, the performance of these methods will degrade while the transmission of the "last gasp" signal is not assured [5]. For instance, as the growth of distributed energy resources (DERs) penetration in distribution grids, customer can still receive power from the rooftop solar panels, battery storage, and electrical vehicles when there is no power flow in the distribution circuit connecting to the customer. So the smart meter at the customer premises cannot report a power outage. Moreover, some secondary distribution grids are mesh networks in urban areas. In this scenario, a single line outage caused by circuit faults and human activities may not cause a power outage due to alternative paths for power supply. In this second case, we will also observe smart meters measuring power injections without sending the "last gasp" notification for reporting outages.

While alternative power sources make the "last gasp" notification fail to report outages, can we still find the line outage time and location? Answering this question, recent literature aimed at collecting additional information for smarter decisions. For example, power measurements, such as phasor angles from phasor measurement units (PMUs), were modeled in [6] as a Gaussian Markov random field to track the grid topology change. Other power measurements, like power flows and load estimates, were also utilized in a compressive system [7] and hypothesis-test-based detection method [8]. Non-power measurements were explored as well, such as human network information from social media [9] and the weather information from environment [10]. In distribution grids, obtaining measurements such as micro-PMUs and accurate power flow data can be challenging and costly, as they are not commonly deployed in households. To address this limitation, our earlier research [3] demonstrated that utilizing readily available voltage magnitudes could still yield accurate outage identification outcomes. However, an in-depth examination of the probability distribution of voltage data and a theoretical guarantee for learning this distribution were not included in our previous work. These aspects are crucial for understanding the outage identification procedure. Besides, the method in [3] has feasibility and accuracy issues when learning the probability distribution. In this work, we fill the above gaps via a novel approach with theoretical guarantees.

To utilize the aforementioned measurements, both deterministic and probabilistic methods were proposed. Deterministic methods usually set up a threshold and declared the outage when the change of data exceeds the threshold. Such methods are simple to apply but cannot accurately discern data change in complex or large-scale grids. Probabilistic methods analyzed the data spatially or temporally. For spatial analysis, [11] studied graph spectral to assess the grid topology for line outage detection. However, such methods required the grid topology as a prior. For temporal analysis, tracing the probability distribution change of the time-series measurements is a common approach [3]. This is usually studied in the change point detection framework, which aims to find the distribution change of measurements as quickly as possible under the constraint of false alarm tolerance [12]. Such framework has been used in line outage and fault detection in transmission grids [13], [14] and DC micro-grids [15]. Although the change point detection framework assures optimal performance [16], it typically necessitates knowledge of both distributions before and after the change. Nevertheless, in distribution grids, this requirement is not practical as the post-outage distribution is unpredictable due to the large number of possible outage patterns, whereas the pre-outage distribution can be learned from historical measurements.

For removing the impractical requirement discussed above, methods were proposed to provide approximation or simplification of the unknown post-change distribution in change point detection. For instance, an approximated maximum likelihood

estimation of unknown distribution parameters was proposed in [3]. A convexified estimation of the unknown distribution approach was introduced in [17]. [18], [19] bypassed the requirement in restricted distribution cases with partially unknown information (e.g., scalar Gaussian with unknown means and known variances). While these methods may mitigate the incompleteness of post-outage information, they have limitations on detection performance and parameter estimation.

In this paper, we propose a practical and straightforward method for utilities to identify line outages with unknown outage patterns. To address the challenge of limited data availability, our approach relies solely on voltage magnitudes obtained from smart meters. This is advantageous compared to expensive phase angle measurements and accurate power flow data, as voltage magnitudes are more readily accessible in typical distribution grids [20]. For the utilize of voltage magnitudes, we have made distinctive contributions. We demonstrate that the increment of voltage magnitudes before and after a line outage follows two distinct multivariate Gaussian distributions, where the distribution parameters are influenced by grid connectivity. Moreover, we provide theoretical guarantees for learning the unknown probability distribution parameters based on voltage magnitude data. By effectively utilizing voltage magnitudes and incorporating theoretical guarantees, we address the limitations posed by the absence of precise phase-angle data. Through the detection of changes in the learned Gaussian distributions, we can successfully identify line outages.

The second challenge is the unavailability of post-outage distribution parameters as analyzed earlier. To address this issue, we propose a data-driven method that directly learns these unknown parameters using Projected Gradient Descent (PGD). While Gradient Descent (GD) is susceptible to feasibility issues in parameter estimation, the iterative nature of GD allows us to control the parameter updating trajectory. Specifically, we formulate the distribution parameter learning problem as a projection optimization problem constrained by the Bergman divergence [21]. This not only resolves the feasibility issue but also leads to accurate parameter estimation with theoretical guarantees. By accurately learning the parameters, our approach can effectively detect and localize line outages, even in large grids.

In addition to accuracy, utilities are also concerned with timely operation. By utilizing the statistical and physical characteristics of voltage data, we can limit the search space of unknown parameters to a convex set, which allows for fast and accurate recovery of the post-outage distribution. We have demonstrated that PGD can achieve optimal parameter learning with a polynomial-time convergence guarantee. Furthermore, we have developed an efficient implementation of the PGD algorithm, which reduces computational time by 75% and makes it particularly well-suited for timely grid operations.

In summary, our proposed method offers several contributions. Firstly, it only requires simple data but have theoretical guarantees. Secondly, it does not require prior knowledge of the outage pattern. Thirdly, it enables timely operation. Furthermore, our approach comes with performance guarantees and does not rely on knowledge of the distribution grid's topology, nor does it require all households to have smart meters data. The method is validated using four distribution grids and real-world load profiles with 17 outage configurations. In the following, Section II models the problem of line outage identification. Section III discusses the voltage data and identification procedure. Section IV extends to identification with unknown outage pattern. Section V provides performance guarantees on timely operation. Section VI evaluates our method. Section VIII concludes the paper.

## II. System Model

For showing our probabilistic design for change point detection and localization, we define variables on a graph probabilistically. Specifically, we model the distribution grid as a graph $\mathcal{G} := \{1, 2, \cdots, M\}$ containing $M > 0$ buses connected by branches. Then, the voltage data from each bus $i \in \mathcal{G}$ is modeled as a random variable $V_i$. As a time-series, its realization at time $n$ is denoted as $v_i[n] = |v_i[n]| \exp(j\theta_i[n]) \in \mathcal{C}$, where $|v_i[n]| \in \mathcal{R}$ represents the voltage magnitude in per unit and $\theta_i[n] \in \mathcal{R}$ is the voltage phase angle in degrees. These steady-state measurements are sinusoidal signals at the same frequency. It's worth noting that unlike PMUs, smart meters typically do not measure phase angles. Therefore, we want to emphasize that even though voltage is represented in its phasor form, solely using the voltage magnitude can still effectively identify a line outage.

In the distribution grid $\mathcal{G}$, the collection of voltage variables is modeled as $\mathbf{V}_{\mathcal{G}} := [V_1, V_2, \cdots, V_M]^\top \in \mathcal{R}^M$. Moreover, since $\mathbf{V}_{\mathcal{G}}$ usually do not follow a regular distribution [3], we model the increment change of voltage data as $\Delta\mathbf{V}_{\mathcal{G}}$, whose realization at time $n$ is $\Delta\mathbf{v}[n] = \mathbf{v}[n] - \mathbf{v}[n-1]$. For the sake of simplicity, we also use the notation $\Delta\mathbf{v}^{1:N} = \{\Delta\mathbf{v}[1], \cdots, \Delta\mathbf{v}[N]\}$ to represent observations up to time $N$.

Based on the modeling, the problem of identifying the distribution grid line outage is formally defined as follows.

- **Given**: Voltage increments $\Delta\mathbf{v}^{1:N}$ from the smart meters.
- **Find**: The line outage time as soon as possible and the out-of-service branch as accurate as possible.

## III. Outage Identification via Voltage Magnitude

While the expensive phasor angles and accurate power flows are hard to obtain in distribution grids, [3] showed that the easier-to-acquire voltage magnitude could be utilized to identify the line outage. The authors found that although voltage data do not follow a regular distribution, the incremental change of voltage follows Gaussian distribution. However, two things were missing in [3]: a clear formula of the distribution and an elaborate analysis of how such distribution is affected by line outages. They are the key to understanding the procedure and performance of identifying the line outage, which will be discussed in detail in the following subsection.

### A. Gaussian Distribution of Voltage Increment

For answering the missing question, in this subsection, we elaborately prove that the increment of voltage data $\Delta\mathbf{V}_{\mathcal{G}}$ follows two multivariate Gaussian distributions before and after the line outage, and provide a clear formula of such

distribution. In doing so, we can identify the outage via tracing the change of the Gaussian distribution.

To study the distribution of $\Delta\mathbf{V}_\mathcal{G}$, we start from the Kirchhoff's Current Law: the relationship between voltages $\mathbf{V}_\mathcal{G} \in \mathcal{C}^M$ and currents $\mathbf{I}_\mathcal{G} \in \mathcal{C}^M$ in the grid is $\mathbf{Y}_\mathcal{G}\mathbf{V}_\mathcal{G} = \mathbf{I}_\mathcal{G}$, where the admittance matrix $\mathbf{Y}_\mathcal{G} \in \mathcal{C}^{M \times M}$ can be derived through the connectivity of the grid as [22]

$$\mathbf{Y}_\mathcal{G} = \mathbf{A}_{\mathcal{E},\mathcal{G}}^\top \mathbf{Y}_\mathcal{E} \mathbf{A}_{\mathcal{E},\mathcal{G}} + \mathbf{Y}_\mathcal{G}^s. \tag{1}$$

In the above equation, $\mathcal{E}$ denotes the set of branches in the grid $\mathcal{G}$. $\mathbf{A}_{\mathcal{E},\mathcal{G}} \in \mathcal{R}^{|\mathcal{E}| \times M}$ is the incidence matrix where each row represents a branch, and has exactly one entry of 1 and one entry of $-1$ to denote the two buses connected by this branch. We can swap the $-1$ and 1 since the grid network is undirectional. $\mathbf{Y}_\mathcal{E} \in \mathcal{C}^{|\mathcal{E}| \times |\mathcal{E}|}$ is a diagonal matrix with the series admittances of each branch, and $\mathbf{Y}_\mathcal{G}^s \in \mathcal{C}^{M \times M}$ is a diagonal matrix with the total shunt admittances at each bus.

By representing $\mathbf{Y}_\mathcal{G}$ in (1), we can discuss the invertibility of $\mathbf{Y}_\mathcal{G}$, which prepares us for the distribution analysis of $\Delta\mathbf{V}_\mathcal{G}$. To this end, we assume that the branches are not electromagnetically coupled and have non-zero admittance, i.e., $\mathbf{Y}_\mathcal{E}$ is full-rank. This assumption is common in distribution grids [22]. With full-rank $\mathbf{Y}_\mathcal{E}$, we show the invertibility of $\mathbf{Y}_\mathcal{G}$ in Lemma 1.

**Lemma 1.** *In a connected distribution grid $(\mathcal{G}, \mathcal{E})$, the admittance matrix $\mathbf{Y}_\mathcal{G} \in \mathcal{C}^{M \times M}$ is invertible after eliminating the slack-bus corresponding column and row.*

In the following, we consider the eliminated admittance matrix and keep the notation unchanged for convenience. Based on Lemma 1, the relationship between voltage increments $\Delta\mathbf{V}_\mathcal{G}$ and current increments $\Delta\mathbf{I}_\mathcal{G}$ can be expressed as

$$\Delta\mathbf{V}_\mathcal{G} = \mathbf{Z}_\mathcal{G}\Delta\mathbf{I}_\mathcal{G}, \quad \text{where} \quad \mathbf{Z}_\mathcal{G} = \mathbf{Y}_\mathcal{G}^{-1}. \tag{2}$$

To derive the distribution of $\Delta\mathbf{V}_\mathcal{G}$, we further introduce a common assumption regarding $\Delta\mathbf{I}_\mathcal{G}$. We consider that $\Delta I$ at each non-slack bus is independent and normally distributed:

$$\Delta I_i \perp \Delta I_k, \ i \neq k \quad \text{and} \quad \Delta I_k \sim \mathcal{N}(\mu_k, \sigma_k^2), \ k \in \mathcal{G}. \tag{3}$$

This statement is adopted and validated by real data in many works [23]–[25], where the authors computed the mutual information between current injections to justify the independence. The empirical histogram in Fig. 2 also suggests that $|\Delta I|$ roughly follows a Gaussian distribution. With the assumption in (3), we present the distribution analysis of $\Delta\mathbf{V}_\mathcal{G}$, which is key to identifying the line outage.
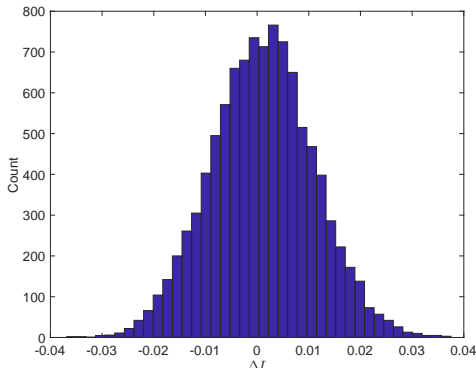


Fig. 2. The empirical histogram of $|\Delta I|$.

**Theorem 1.** *Provided that (3) hold, $\Delta\mathbf{V}_\mathcal{G}$ (excluding slack-bus) in a connected distribution grid $(\mathcal{G}, \mathcal{E})$*
*1) follows a multivariate Gaussian distribution,*
*2) still follows a multivariate Gaussian distribution (with different mean and covariance) after grid topology changes.*

*Proof.* With (2), $\Delta V_i$ of each bus $i \in \mathcal{G}$ (slack-bus is excluded from $\mathcal{G}$) can be expressed by $\Delta V_i = \sum_{k \in \mathcal{G}} Z_{ik}\Delta I_k$, where $Z_{ik}$ is the $(i,k)$ element of $\mathbf{Z}_\mathcal{G}$. Hence, any non-trivial linear combination of $\Delta V_i, i \in \mathcal{G}$ can also be represented by a linear combination of $\Delta I_k, k \in \mathcal{G}$, and is normally distributed. This implies that the joint distribution of $\Delta\mathbf{V}_\mathcal{G} = [\Delta V_1, \cdots, \Delta V_M]^\top$ is a multivariate Gaussian distribution as

$$\Delta\mathbf{V}_\mathcal{G} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{4}$$

where $\boldsymbol{\mu}_i = \sum_{k \in \mathcal{G}} Z_{ik}\mu_k$ and $\boldsymbol{\Sigma}_{ik} = \sum_{l \in \mathcal{G}} Z_{il}Z_{kl}\sigma_l^2$.

When the grid topology is changed (e.g., due to a line outage), the incidence matrix $\mathbf{A}_{\mathcal{E},\mathcal{G}}$ changes accordingly: if branch $l$ connecting bus $i$ and $k$ is out-of-service, $(\mathbf{A}_{\mathcal{E},\mathcal{G}})_{l,i}$ and $(\mathbf{A}_{\mathcal{E},\mathcal{G}})_{l,k}$ become zero. Denoting the new incidence matrix as $\widetilde{\mathbf{A}}_{\mathcal{E},\mathcal{G}}$, there are two scenarios:

- The grid network is still connected (which is our focus in this paper). In this case, the changed admittance matrix $\widetilde{\mathbf{Y}}_\mathcal{G} = \widetilde{\mathbf{A}}_{\mathcal{E},\mathcal{G}}^\top \mathbf{Y}_\mathcal{E} \widetilde{\mathbf{A}}_{\mathcal{E},\mathcal{G}} + \mathbf{Y}_\mathcal{G}^s$ is still invertible, which results in a varied $\widetilde{\mathbf{Z}}_\mathcal{G} = \widetilde{\mathbf{Y}}_\mathcal{G}^{-1}$. It implies that $\Delta\mathbf{V}_\mathcal{G}$ still follows a multivariate Gaussian distribution, only with different mean $\widetilde{\boldsymbol{\mu}}$ and different covariance $\widetilde{\boldsymbol{\Sigma}}$ calculated according to $\widetilde{\mathbf{Z}}_\mathcal{G}$.
- The grid network is disconnected. In this case, we view the network as disjoint islands where each part is a connected sub-network, e.g., $\mathcal{G} = \mathcal{G}_1 \cup \mathcal{G}_2, \mathcal{G}_1 \cap \mathcal{G}_2 = \emptyset$. By doing so, we can write the incidence matrix in block format, e.g., $\widetilde{\mathbf{A}}_{\mathcal{E},\mathcal{G}} = \begin{pmatrix} \widetilde{\mathbf{A}}_{\mathcal{E}_1,\mathcal{G}_1} & \mathbf{0} \\ \mathbf{0} & \widetilde{\mathbf{A}}_{\mathcal{E}_2,\mathcal{G}_2} \end{pmatrix}$. According to the first case, voltage increments in each sub-network follow a multivariate Gaussian distribution, and so does their joint distribution. In this scenario, since some houses lose power connection and will have zero voltages, the outage time and location can be more easily found via our approach. $\square$

Suppose the outage occurs at time $\lambda$, Theorem 1 allows us to write the sequence of voltage increments as

$$\begin{cases} \Delta\mathbf{v}[n] \overset{i.i.d}{\sim} g : \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0), & n = 1, 2, \cdots, \lambda-1, \\ \Delta\mathbf{v}[n] \overset{i.i.d}{\sim} f : \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), & n = \lambda, \lambda+1, \cdots, N, \end{cases} \tag{5}$$

where $g$ denotes the pre-outage Gaussian distribution and $f$ denotes the post-outage Gaussian distribution. The mean vectors $\boldsymbol{\mu}_0, \boldsymbol{\mu}_1$, along with covariance matrices $\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1$, are the parameters of these distributions. In our work, the pre-outage parameters $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ can be estimated using historical data during normal operation periods of the distribution grid [3]. The post-outage parameters $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_1$ are considered unknown to reflect real-world outage scenarios, since the outage pattern is unpredictable/unknown. To visualize the varying distribution of the sequence, we provide an illustration of $|\Delta v[N]| \in \mathcal{R}$ in Fig. 1(b).
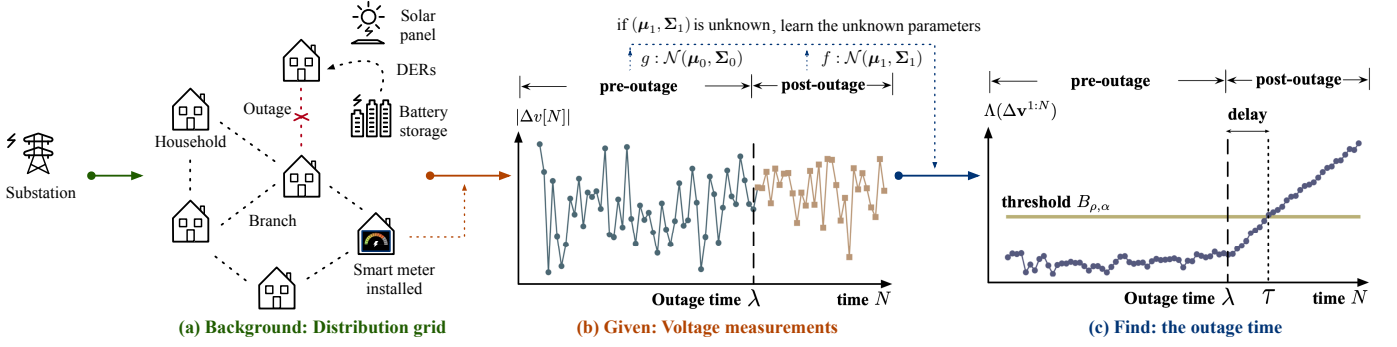
Fig. 1. An overview of the distribution grid line outage detection problem: we collect voltage magnitudes from smart meters installed at households and use the posterior probability ratio computed in (6) to detect the change in the underlying distribution of voltage increments.

## B. Outage Identification via Distribution Change

Before proposing our novel solution to unknown outage pattern, we present the commonly used framework to find the outage time $\lambda$ and outage branch, given voltage data in (5).

To identify the outage time $\lambda$, we conduct a sequential hypothesis test $\mathcal{H}_0 : \lambda > N$ and $\mathcal{H}_1 : \lambda \leq N$ at every time step $N$. As $N$ increases, the first time we reject the null hypothesis $\mathcal{H}_0$ determines the value of $\lambda$. To decide when to reject $\mathcal{H}_0$, we compute the posterior probability ratio at each time step $N$ as

$$
\begin{aligned}
\Lambda(\Delta \mathbf{v}^{1:N}) &= \frac{\mathbb{P}(\lambda \leq N | \Delta \mathbf{v}^{1:N})}{\mathbb{P}(\lambda > N | \Delta \mathbf{v}^{1:N})} \\
&= \frac{\sum_{k=1}^{N} \pi(k) \prod_{n=1}^{k-1} g(\Delta \mathbf{v}[n]) \prod_{n=k}^{N} f(\Delta \mathbf{v}[n])}{\sum_{k=N+1}^{\infty} \pi(k) \prod_{n=1}^{N} g(\Delta \mathbf{v}[n])},
\end{aligned}
\quad (6)
$$

where $\lambda \in \mathbb{N}$ is assumed to follow a prior distribution $\pi$. The posterior probability ratio in (6) compares the probabilities of "outage occurred $(\lambda \leq N)$" and "outage did not occur $(\lambda > N)$" based on the historical measurements $\Delta \mathbf{v}^{1:N}$. A larger posterior probability ratio indicates that "outage occurred" is more likely than "outage did not occur." Therefore, we declare the outage when the ratio (6) exceeds a predefined threshold. By the Shiryaev-Roberts-Pollaks procedure [12], [16], the following threshold in Theorem 2 optimally considers the trade-off between the false alarm and the detection delay.

**Theorem 2.** *(Line outage detection). When $\lambda$ follows a geometric prior $Geo(\rho)$, we declare the outage at the first time when posterior probability ratio $\Lambda(\Delta \mathbf{v}^{1:N})$ surpasses the threshold $B_{\rho,\alpha} = (1 - \alpha)/(\rho \alpha)$ as*

$$
\tau = \inf\{N \in \mathbb{N} : \Lambda(\Delta \mathbf{v}^{1:N}) \geq B_{\rho,\alpha}\}, \quad (7)
$$

*where the false alarm rate $\mathbb{P}(\tau < \lambda)$ is upper bounded by maximal false alarm rate $\alpha$. As $\alpha \to 0$, $\tau$ in (7) is asymptotically optimal for minimizing the average detection delay $\mathbb{E}[\tau - \lambda | \tau \geq \lambda]$ as*

$$
\begin{aligned}
\mathbb{E}[\tau - \lambda | \tau \geq \lambda] &= \frac{|\log \alpha|}{-\log(1 - \rho) + D_{KL}(f \| g)} \\
&= \inf_{\mathbb{P}(\tau^* \leq \lambda) \leq \alpha} \mathbb{E}[\tau^* - \lambda | \tau^* \geq \lambda],
\end{aligned}
\quad (8)
$$

*where $D_{KL}(f \| g)$ is the KL divergence between $f$ and $g$.*

One notable feature of the detection procedure described above is its ability to function effectively without requiring knowledge of the grid topology. Additionally, it can handle non-Gaussian distributions for $f$ and $g$. As depicted in Fig. 1(c), we calculate the posterior probability ratio sequentially and identify the outage time when the ratio exceeds the threshold.

Once the line outage occurrence is detected, localizing the out-of-service branch is also critical for system recovery. In [3], the authors proposed an accurate outage localization method by proving that the voltage increments of two disconnected buses are conditionally independent. They computed the conditional correlation of every possible pair of buses in the grid and checked if the value changes from non-zero to zero. This approach differs from the utilization of nodal electric circuit matrices [26], [27] for estimating fault location, while our approach has also been effective (as shown in Section VI-D) and capitalizes on the learned covariance matrix in scenarios where the post-outage distribution is unknown.

To estimate the conditional correlation between bus $i$ and bus $k$, the covariance matrix $\boldsymbol{\Sigma}$ is utilized. Let set $\mathcal{I} := \{i, k\}$ and $\mathcal{K} := \mathcal{G} \backslash \{i, k\}$, the covariance matrix is decomposed as $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{\mathcal{II}} & \boldsymbol{\Sigma}_{\mathcal{IK}} \\ \boldsymbol{\Sigma}_{\mathcal{IK}}^{\top} & \boldsymbol{\Sigma}_{\mathcal{KK}} \end{bmatrix}$. Based on this, the conditional correlation $\rho_{ik}$ between bus $i$ and bus $k$ is

$$
\rho_{ik}(\boldsymbol{\Sigma}) = \frac{\boldsymbol{\Sigma}_{\mathcal{I}|\mathcal{K}}(1, 2)}{\sqrt{\boldsymbol{\Sigma}_{\mathcal{I}|\mathcal{K}}(1, 1) \boldsymbol{\Sigma}_{\mathcal{I}|\mathcal{K}}(2, 2)}}, \quad (9)
$$

where the conditional covariance is computed by the Schur complement [28] as $\boldsymbol{\Sigma}_{\mathcal{I}|\mathcal{K}} = \boldsymbol{\Sigma}_{\mathcal{II}} - \boldsymbol{\Sigma}_{\mathcal{IK}} \boldsymbol{\Sigma}_{\mathcal{KK}}^{-1} \boldsymbol{\Sigma}_{\mathcal{IK}}^{\top}$.

**Theorem 3.** *(Line outage localization). The conditional correlation is calculated based on (9) for every pair of $(i, k)$ as*

$$
\underbrace{\rho_{ik}^{-} = \rho_{ik}(\boldsymbol{\Sigma}_0)}_{\text{before outage}} \quad and \quad \underbrace{\rho_{ik}^{+} = \rho_{ik}(\widehat{\boldsymbol{\Sigma}}_1)}_{\text{after outage}}. \quad (10)
$$

*The branch between bus $i$ and $k$ is out-of-service if $|\rho_{ik}^{-}| > \delta_{\max}$ and $|\rho_{ik}^{+}| < \delta_{\min}$. The thresholds are set as $\delta_{\max} = 0.5$ and $\delta_{\min} = 0.1$ based on real-world outage data to check if the correlation changes from non-zero to near-zero value.*

According to Theorem 3, we track the change of covariance matrices to localize the out-of-service branch. Specifically, an out-of-service branch between bus $i$ and bus $k$ can be identified if both of the following conditions are met simultaneously: (1) $|\rho_{ik}^{-}| > \delta_{\max}$ indicating the presence of a branch between

buses $i$ and $k$ before the outage, and (2) $|\rho_{ik}^+| < \delta_{\min}$ indicating the absence of a branch between buses $i$ and $k$ after the outage. Notably, this process still does not need the grid topology as a prior.

## IV. OUTAGE IDENTIFICATION WITH UNKNOWN PATTERN

The detection and localization procedure in Section III requires knowing all the parameters of $g$ and $f$ in advance. However, this is impractical in real-world distribution grids. Specifically, although we know $g$ and $f$ are multivariate Gaussian distributions based on Theorem 1, the parameters (mean vectors and covariance matrices) of $f$ are usually hidden. While the pre-outage distribution parameters can be learned by historical measurements during the normal grid operation, the post-outage distribution parameters are often unavailable. In fact, since there are a large number of branches and therefore, substantial combinations of outage patterns, we can not predict the outage pattern or the post-outage distribution parameters. Hence, we need to estimate the unknown parameters before conducting the aforementioned methods to identify the outage.

To resolve such issue, we propose a data-driven framework to learn the post-outage distribution parameters $\boldsymbol{\theta} = (\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ jointly. Specifically, we want to find the parameter set that minimizes the negative likelihood function $L(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ as

$$(\widehat{\boldsymbol{\mu}}_1, \widehat{\boldsymbol{\Sigma}}_1) = \arg \min_{(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)} L(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \qquad (11)$$

where $L(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ is computed as

$$- \sum_{k=1}^{N} \pi(k) \prod_{n=1}^{k-1} g(\Delta \mathbf{v}[n]) \prod_{n=k}^{N} f(\Delta \mathbf{v}[n] | \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1). \qquad (12)$$

To address the non-convex nature of the likelihood expressed in equation (12), the authors in [3] proposed a convex approximation using Jensen's inequality and derived closed-form solutions for equation (11). However, the use of Jensen's inequality can introduce inaccuracies in the resulting closed-form solutions, particularly in determining the minimum point. Furthermore, the estimated covariance matrix may not always be feasible. Specifically, a feasible covariance matrix must be positive definite, i.e., $\boldsymbol{\Sigma}_1 \succ 0$, and if this condition is not met during the learning process, the computation of the probability density of $f$ can fail.

An alternative approach is using the Gradient Descent (GD) to find the solution to (11). While the vanilla GD also can not ensure the aforementioned feasibility of the parameters, the iterative learning nature in GD enables us to control the updating trajectory of parameters.

### A. Unknown Parameters Estimation via Projected Gradient Descent with Bregman Divergence Constraint

To guarantee that the estimation of parameters $\boldsymbol{\theta} = \{\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1\}$ are always feasible, we introduce the Bregman divergence [21] to constrain the estimate in each iteration of GD and arrive at a series of optimization problems as

$$\boldsymbol{\theta}_i^{(e+1)} = \arg \min_{\boldsymbol{\theta}_i} \underbrace{\Delta_\Phi(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{-i}^{(e)})}_{\text{Bregman divergence}} + \eta L(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{-i}^{(e)}). \qquad (13)$$

In the above equation, $\boldsymbol{\theta}_i^{(e)}$ is the update of the $i^{th}$ parameter at $e^{th}$ iteration, $\boldsymbol{\theta}_{-i}^{(e)} = \boldsymbol{\theta}^{(e)} \setminus \boldsymbol{\theta}_i^{(e)}$ is a complement set, and $\eta$ is the trade-off learning rate. The Bregman divergence

$$\Delta_\Phi(\boldsymbol{\theta}_i, \boldsymbol{\theta}_i^{(e)}) := \Phi(\boldsymbol{\theta}_i) - \Phi(\boldsymbol{\theta}_i^{(e)}) - \text{tr}\left((\boldsymbol{\theta}_i - \boldsymbol{\theta}_i^{(e)})\Phi(\boldsymbol{\theta}_i^{(e)})^\top\right)$$

provides a distance measurement between two variables $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_i^{(e)}$, where $\Phi$ is a strictly convex differentiable function. Intuitively, it restricts the new estimate $\boldsymbol{\theta}_i^{(e+1)}$ to stay relatively close to the previous estimate $\boldsymbol{\theta}_i^{(e)}$. Therefore, if the initial guess $\boldsymbol{\theta}_i^{(0)}$ is within the feasible domain, we can expect that the parameter $\boldsymbol{\theta}_i$ following (13) will update towards the direction of minimizing the negative likelihood and meanwhile satisfy a similar property (e.g., positive definiteness of covariance matrix) since the update step is restricted.

Finding the solution to (13) relies on one characteristic of Bregman divergence: its gradient with respect to $\boldsymbol{\theta}_i$ has a simple form $\nabla_{\boldsymbol{\theta}_i} \Delta_\Phi(\boldsymbol{\theta}_i, \boldsymbol{\theta}_i^{(e)}) = \dot{\Phi}(\boldsymbol{\theta}_i) - \dot{\Phi}(\boldsymbol{\theta}_i^{(e)})$, where $\dot{\Phi}$ is the differential of $\Phi$. Based on this, we can eliminate the argmin in (13) by setting the gradient (with respect to $\boldsymbol{\theta}_i$) of the objective function to zero, and derive the following Projected Gradient Descent solution [29].

**Lemma 2.** *The optimization problem in* (13) *is solved as*

$$\boldsymbol{\theta}_i^{(e+1)} = \dot{\Phi}^{-1}\left(\dot{\Phi}(\boldsymbol{\theta}_i^{(e)}) + \eta \nabla_{\boldsymbol{\theta}_i} L(\boldsymbol{\theta}^{(e)})\right). \qquad (14)$$

With Lemma 2, we propose the important result of our paper: the learning scheme of unknown parameters with feasibility guarantee. We will further show that this learning scheme is accurate and has convergence guarantees.

**Theorem 4.** *(Projected Gradient Descent of learning $\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1$). With careful customization of Bregman divergence (i.e., choosing the appropriate function $\Phi$), the Projected Gradient Descent learning in* (14) *becomes feasible.*

- $\underline{\boldsymbol{\mu}_1 \in \mathcal{R}^M}$: $\Phi(\boldsymbol{\mu}_1) = \frac{1}{2}\|\boldsymbol{\mu}_1\|_2^2$ *and* $\dot{\Phi}^{-1}(\boldsymbol{\mu}_1) = \boldsymbol{\mu}_1$. *The learning scheme is*

$$\boldsymbol{\mu}_1^{(e+1)} = \boldsymbol{\mu}_1^{(e)} - \eta \nabla_{\boldsymbol{\mu}_1} L(\boldsymbol{\mu}_1^{(e)}, \boldsymbol{\Sigma}_1^{(e)}). \qquad (15)$$

- $\underline{\boldsymbol{\Sigma}_1 \succ 0}$: $\Phi(\boldsymbol{\Sigma}_1) = \text{tr}(\boldsymbol{\Sigma}_1 \log \boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_1)$ *and* $\dot{\Phi}^{-1}(\boldsymbol{\Sigma}_1) = \exp \boldsymbol{\Sigma}_1$. *The learning scheme is*

$$\boldsymbol{\Sigma}_1^{(e+1)} = \exp\left(\log \boldsymbol{\Sigma}_1^{(e)} - \eta \nabla_{\boldsymbol{\Sigma}_1} L(\boldsymbol{\mu}_1^{(e)}, \boldsymbol{\Sigma}_1^{(e)})\right). \qquad (16)$$

The learning process of $\boldsymbol{\Sigma}_1$ is shown in Fig. 3. Since matrix exponential maps any symmetric matrix to a positive definite matrix, the learning scheme (16) maintains the property of positive definiteness, i.e., $\boldsymbol{\Sigma}_1^{(e+1)} \succ 0$ if $\boldsymbol{\Sigma}_1^{(e)} \succ 0$.
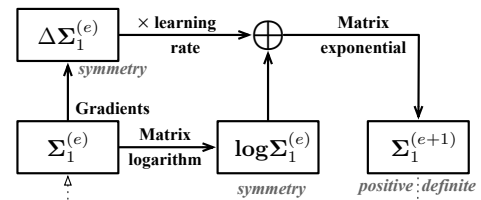


Fig. 3. The Projected Gradient Descent update of $\boldsymbol{\Sigma}_1$.

Besides the statistical properties (e.g., the covariance matrix is positive definite), smart meter data have physical properties as well due to grid operation. For example, because the

standard range of voltage magnitude is between $0$ $p.u.$ and $1.1$ $p.u.$, the mean value of voltage increment should be between $-1.1$ $p.u.$ and $1.1$ $p.u.$. Our learning scheme in (14) can also satisfy this requirement by defining

$$\Phi(\boldsymbol{\mu}_1) = \sum_{i=1}^{M} [(\mu_i + 1.1)\log(\mu_i + 1.1)$$
$$+ (1.1 - \mu_i)\log(1.1 - \mu_i) + \mu_i], \qquad (17)$$

where $\mu_i$ is the $i^{th}$ element of the mean vector $\boldsymbol{\mu}_1$.

To conclude, when post-outage distribution parameters $\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1$ are unknown, we use (17) and (16) to accurately learn them with feasibility and convergence guarantees. The pre-outage parameters $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ can be estimated using historical data during normal operation periods of the distribution grid [3]. By obtaining these parameters of the Gaussian density functions $g$ and $f$ in (6), we can explicitly calculate $g$ and $f$. It enables us to implement Theorem 2 for detecting the outage time and use Theorem 3 for localizing the outage branch. This framework is summarized into Algorithm 1.

---

**Algorithm 1:** Line outage identification with unknown post-outage distribution parameters

---

**Input:** New observation $\Delta\mathbf{v}[N]$
**Output:** Outage time $\tau$ and outage location
Set $\boldsymbol{\mu}_1^{(0)}, \boldsymbol{\Sigma}_1^{(0)}$ from $(N-1)^{th}$ step     // warm start
**for** $e = 0, 1, \ldots$ **do**
  $\boldsymbol{\mu}_1^{(e+1)} \leftarrow \dot{\Phi}^{-1}\left(\dot{\Phi}(\boldsymbol{\mu}_1^{(e)}) - \eta\nabla_{\boldsymbol{\mu}_1}L\right)$     // use $\Phi$ in (17)
  $\boldsymbol{\Sigma}_1^{(e+1)} \leftarrow \exp\left(\log\boldsymbol{\Sigma}_1^{(e)} - \eta\nabla_{\boldsymbol{\Sigma}_1}L\right)$
  **if** $|L(\boldsymbol{\mu}_1^{(e)}, \boldsymbol{\Sigma}_1^{(e)}) - L(\boldsymbol{\mu}_1^{(e+1)}, \boldsymbol{\Sigma}_1^{(e+1)})| \leq 10^{-3}$
  **then**
    **return** $\boldsymbol{\mu}_1^{best}, \boldsymbol{\Sigma}_1^{best}$     // return best update result, defined in Theorem 5
**end**
**if** $\Lambda(\Delta\mathbf{v}^{1:N}) \geq B_{\rho,\alpha}$ **then**
  **for** $i, k \in \mathcal{G}$ **do**
    **if** $|\rho_{ik}^-| > \delta_{\max}$ *and* $|\rho_{ik}^+| < \delta_{\min}$ **then**
      **return** $\tau = N$, *report the out-of-service branch between bus $i$ and $k$*
  **end**

---

## V. TIMELY OUTAGE IDENTIFICATION WITH PERFORMANCE GUARANTEE

In addition to the feasibility issue that has already been addressed in Theorem 4, the accuracy and computation time of the proposed learning scheme are two other concerns when implementing such a method for real-world outage identification. In this section, we demonstrate that our proposed method can achieve the optimal parameter solution with a guaranteed convergence. Furthermore, we present an efficient implementation for timely operation.

### A. Restricted Convexity for Convergence Guarantee

While the non-convexity of likelihood $L(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ in (12) hinders us from deriving a convergence analysis directly, we note that $L(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ is constrained convex. Specifically, we

notice that the unknown parameters $\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1$ of $f$ are not supposed to be far away from the known parameters $\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0$ of $g$, thus freeing us from searching the entire parameter space. This is because the alternative power supply makes the impact of line outage less severe. In fact, if $\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1$ are significantly far from $\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0$, distinguishing between $f$ and $g$ would become trivial: imaging a large leap from pre-outage data to post-outage data in Fig. 1(b), one can detect this change very easily. Hence, a reasonable assumption is that $\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1$ are relatively close to $\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0$, which actually results in a much harder detection problem. With this assumption, we can restrict our search for parameters in a constrained set where $L(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ has good properties. To formally present this, we introduce the restricted convexity [30].

**Definition 1.** *A continuously differentiable function $H$ : $\mathcal{R}^M \to \mathcal{R}$ is restricted convex over a possibly non-convex region $\mathcal{D} \subseteq \mathcal{R}^M$ if every $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{D}$ we have $H(\boldsymbol{y}) \geq H(\boldsymbol{x}) + \langle\nabla H(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x}\rangle$.*

Then, we show that $L(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ satisfies the restricted convexity in Definition 1. Specifically, $L(\boldsymbol{\mu}_1)$ is restricted convex on constrained set $\left\{\boldsymbol{\mu}_1 \Big| \boldsymbol{\Sigma}_1 \succeq \frac{\boldsymbol{v}_k(\boldsymbol{\mu}_1)\boldsymbol{v}_k(\boldsymbol{\mu}_1)^T}{(N-k+1)}, \forall k \leq N\right\}$ where $\boldsymbol{v}_k(\boldsymbol{\mu}_1) = \sum_{n=k}^{N}(\Delta\mathbf{v}[n] - \boldsymbol{\mu}_1)$. Similarly, $L(\boldsymbol{\Sigma}_1)$ is restricted convex on constrained set $\{\boldsymbol{\Sigma}_1 \succ 0 | \nabla^2 L(\text{vec}(\boldsymbol{\Sigma}_1)) \succeq 0\}$. Based on this property, we derive in Theorem 5 the convergence of updating $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_1$.

**Theorem 5.** *Using PGD to iteratively update $\boldsymbol{\theta}_i$, the best update $\boldsymbol{\theta}_i^{best} := \arg\min_{e \in [E]} L(\boldsymbol{\theta}_i^{(e)})$ and the averaged update $\boldsymbol{\theta}_i^{avg} := \frac{1}{E}\sum_{e=1}^{E}\boldsymbol{\theta}_i^{(e)}$ will converge to the optimal value $\boldsymbol{\theta}_i^* = \arg\min_{\boldsymbol{\theta}_i} L(\boldsymbol{\theta}_i)$ with a step size $\eta = \frac{1}{\sqrt{E}}$:*

$$L(\boldsymbol{\theta}_i^{best}) \leq L(\boldsymbol{\theta}_i^*) + \varepsilon \quad and \quad L(\boldsymbol{\theta}_i^{avg}) \leq L(\boldsymbol{\theta}_i^*) + \varepsilon,$$

*for any $\varepsilon > 0$, after at most $E = \mathcal{O}(\frac{1}{\varepsilon^2})$ iterations.*

The proof is in Appendix B. Moreover, since the best update converges faster as shown in Section VI, we choose it as the output of the learning scheme in Algorithm 1. To better visualize how the parameters are updated in the restricted convex area via Projected Gradient Descent (PGD), we provide Fig. 4. As we see, although the likelihood function $L$ is not convex in the entire parameter space, the restricted area is convex, opening the door for learning accurate and feasible parameter solutions to (11).
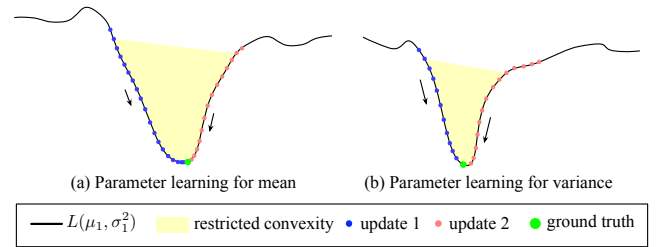


(a) Parameter learning for mean     (b) Parameter learning for variance

— $L(\mu_1, \sigma_1^2)$     restricted convexity     • update 1     • update 2     • ground truth

Fig. 4. Visualization of learning parameters in $g \sim \mathcal{N}(0, \frac{1}{2})$, $f \sim \mathcal{N}(1, \frac{1}{5})$.

### B. Acceleration for Timely Operation

Theorem 5 shows that our proposed method can find the optimal parameters with polynomial-time complexity, which enables quick operation. In this subsection, we provide an

efficient implementation of the learning scheme to further accelerate the algorithm for timely outage identification. To achieve so, we notice that while the matrix exponential and logarithm operations in (16) provide good properties of covariance estimation, it is very time-consuming when calculating them. The calculation is time-consuming because it is often based on their infinite Taylor series. To accelerate these operations, we propose to use finite terms of their Taylor series to approximate the operations.

The matrix exponential is given by the power series in (18), and can be approximated by its first $K_{\text{exp}}$ terms since $\frac{1}{k!}$ decreases drastically when $k$ becomes large. However, once we replace the original matrix exponential operation with the approximated operation $\widehat{\exp}$, we need to verify that $\widehat{\exp}$ also maps any symmetric matrix to a positive definite matrix to satisfy the conclusion in Theorem 4. Otherwise, we will arrive at covariance estimates outside the feasible domain. With this motivation, we show in Lemma 3 that if we choose an appropriate value of $K_{\text{exp}}$ for approximation, the operation $\widehat{\exp}$ has similar properties as $\exp$.

**Lemma 3.** *The matrix exponential can be approximated as*

$$\exp(\boldsymbol{X}) = \sum_{k=0}^{\infty} \frac{1}{k!} \boldsymbol{X}^k \approx \sum_{k=0}^{K_{exp}} \frac{1}{k!} \boldsymbol{X}^k := \widehat{\exp}(\boldsymbol{X}), \quad (18)$$

*for any real matrix $\boldsymbol{X}$. Also, $\widehat{\exp}(\boldsymbol{X}) \succ 0$ for any symmetric $\boldsymbol{X}$ if $K_{exp}$ is even and $K_{exp} > \max\{0, -a_{\min}\}$, where $a_{\min}$ is the smallest eigenvalue of $\boldsymbol{X}$.*

Similarly, the matrix logarithm is given by the power series in (19), and can be approximated by the first $K_{\text{log}}$ terms.

**Lemma 4.** *The matrix logarithm can be approximated as*

$$\log(\boldsymbol{X}) \approx \sum_{k=1}^{K_{log}} \frac{1}{k}(-1)^{k+1}(\boldsymbol{X} - \boldsymbol{I})^k := \widehat{\log}(\boldsymbol{X}). \quad (19)$$

In order to make Theorem 4 still hold true when we use the approximated operation $\widehat{\log}$, we only need the symmetry of $\widehat{\log}$, which can be easily verified as $\widehat{\log}(\boldsymbol{X}^\top) = \sum_{k=1}^{K_{log}} \frac{1}{k}(-1)^{k+1}\left((\boldsymbol{X}-\boldsymbol{I})^k\right)^\top = \left(\widehat{\log}(\boldsymbol{X})\right)^\top$.

In summary, the proposed two approximation operations $\widehat{\exp}$ and $\widehat{\log}$ can still preserve the feasibility of the covariance matrix estimation. The choice of $K_{\text{exp}}$ and $K_{\text{log}}$ is a trade-off between execution time and approximation accuracy: when $K_{\text{exp}}$ is small, the operation $\widehat{\exp}$ is very fast but provides a poor approximation of $\exp$ and vice versa. In Section VI, we will demonstrate how we choose an appropriate approximation level which results in almost zero errors and over 75% reduction in execution time.

## VI. VALIDATE ON EXTENSIVE OUTAGE SCENARIOS WITH REAL-WORLD DATA

This section shows how our proposed method performs in various distribution grids with real-world data. To evaluate our method in systems with different sizes and environments, we design extensive experiments on IEEE 8-bus, IEEE 123-bus networks [31], as well as two European representative distribution systems: medium voltage (MV) network in the urban area and low voltage (LV) network in the suburban area [32]. In each network, bus 1 is selected as the slack bus.

To account for more complex outage scenarios in real-world distribution grids, we examine situations where alternative power sources are available after a line outage. In these scenarios, the "last gasp" notification is ineffective, making it more difficult to detect the line outage. We simulated the following two representative scenarios to replicate this complex scenario. It should also be noted that if certain buses are disconnected from the main grid and experience a voltage magnitude of zero following an outage, our method can accurately and quickly identify the out-of-service line. This is a simpler case compared to the ones we simulated below.

- Mesh networks where most buses have non-zero voltages after the outage since they can receive power from alternative branches. Mesh network often depicts the outage scenario in urban areas. For simulating mesh networks, we add loops in each aforementioned network to ensure it is still connected after line outages, following the study in [3] and [25].
- Radial networks with DERs where some buses still receive power from DERs though they are isolated from the main grid after an outage (see Fig. 1(a)). This outage scenario is typical in residential areas. To simulate DERs, we select multiple buses to have solar power panels with batteries as the storage. For the solar panel, we use the power generation profile computed by PVWatts Calculator [33].

For simulating more realistic data, we use the real residential power profile from Duquesne Light Company (DLC) in Pittsburgh, PA, USA. This profile contains anonymized and secure hourly (and 15-minute) smart meter readings of active power over more than 5,000 houses in the year 2016. The basic statistics of this dataset are summarized in Table I.

TABLE I
STATISTICAL ANALYSIS OF DLC POWER DATASET.

| Statistics | Value |
|---|---|
| Minimum Value | $-2.6040$ |
| Maximum Value | $26.6860$ |
| Mean | $0.8473$ |
| Standard Deviation | $0.6387$ |
| Skewness | $1.7441$ |

The time-series voltage data are simulated by the MATLAB Power System Simulation Package (MATPOWER) in MATLAB R2022b. In every distribution network, we assign active power $p_i[n]$ from the above DLC power profile to each bus $i$ at time $n$. The reactive power $q_i[n]$ is computed according to a randomly generated power factor $pf_i[n]$, which follows a uniform distribution, e.g. $pf_i[n] \sim \text{Unif}(0.9, 1)$. Based on the active and reactive power, we use MATPOWER to solve power flow equations and obtain voltage measurements. Moreover, we can simulate an outage scenario by setting the admittance of a branch or several branches to zero. Hence, we can generate the voltage data during both normal operation and outage scenarios. After we obtain voltage data from MATLAB, all the remaining calculations in outage detection in Algorithm 1 are implemented with Python 3.8 on a personal computer with a Windows 10 operating system, an Intel Core i7 processor clocked at 2.2 GHz, and 16 GB of RAM.

Due to the limited deployment of PMU in reality, the voltage phase angles are hard to obtain. Hence, as mentioned in Section II, we only use the voltage magnitude in the following

experiments even though we model the voltage data in its phasor form. Another concern of data is the high dimensionality in large-scale grids. To resolve this computational issue, we apply the whitening transformation to our data as $\Delta \mathbf{v}^{1:N} \rightarrow \boldsymbol{W} \Delta \mathbf{v}^{1:N}$ based on the PCA whitening matrix satisfying $\boldsymbol{W}^\top \boldsymbol{W} = \boldsymbol{\Sigma}_0^{-1}$. Since the whitening transformation does not change the KL divergence between $g$ and $f$, it has no impact on the outage detection performance.

In the subsequent experiments, we compare our proposed method with various baselines. When full knowledge of post-outage distribution $f$ is known, we refer to the optimal Bayesian procedure as $f$ **known**. When the parameters of $f$ are unknown, our method is referred to as **PGD**. For baseline methods specifically designed for outage detection with unknown post-outage distribution, we consider an approximated maximum likelihood estimation (**MLE**) proposed to learn the unknown parameters [3], [34], a generalized likelihood ratio test (**GLRT**) that only considers finite possibilities [35] of post-outage distributions $f$, a **Shewhart** test [36] that utilizes meanshift and covariance changes in the data to detect outages. For methods that are developed for unknown post-change distribution in the change point detection, we consider a nonparametric binned generalized statistic (**BGS**) proposed to approximate the original ratio test in classic CPD [37], a non-parametric uncertain likelihood ratio (**ULR**) proposed to replace the original ratio [38], a distributed approach (**DIS**) [1], and a deep Q-network approach (**DCQ**) [39].

For more robust evaluation, each experiment will be conducted by the Monte Carlo simulation with over 1000 replications. In every replication, we randomly simulate outage time $\lambda$ through geometric distribution $\mathrm{Geo}(\rho)$. This geometric prior is based on our belief that outages can occur independently at any time step, with an equal probability of $\rho$. We choose $\rho = 0.04$ in our experiments, which is derived from historical outage data, indicating that each time step has a 4% chance of experiencing a line outage.

### A. Parameters Estimation with Accuracy and Convergence

Prior to demonstrating the accurate identification of outages with unknown post-outage distribution parameters, we must first verify that our method can learn the optimal parameters with a guaranteed convergence. Throughout the parameter learning iterations, we plot the Euclidean distance between the best update and the ground truth in Fig. 5. The plot indicates that our learning process converges to the ground truth, thereby verifying the convergence conclusion stated in Theorem 5.
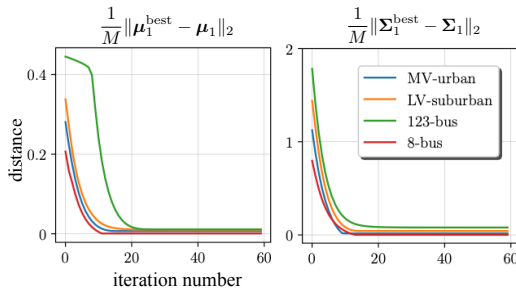


Fig. 5. Distance between best update and ground truth against iterations.

### B. Outage Detection with Small Delay and Rare False Alarm

After evaluating the effectiveness of using PGD to learn the unknown parameters, we then verify the performance of outage detection using such learned parameters.

The first criterion to evaluate our detection procedure is the average detection delay. To validate the asymptotic optimality of the detection delay in Theorem 2, in Fig. 6, we plot the average delay $\mathbb{E}(\tau - \lambda | \tau \geq \lambda)$ divided by $|\log \alpha|$ and the theoretical lower bound $-\log(1 - \rho) + D_{KL}(f||g)$. We observe that the average detection delay of the case when $f$ is known and that of the PGD both achieve the optimal lower bound asymptotically, while the delay of PGD is slightly higher. Moreover, using PGD and accelerated PGD to learn the unknown post-outage distribution statistics enables quicker line outage detection compared to the method of MLE.
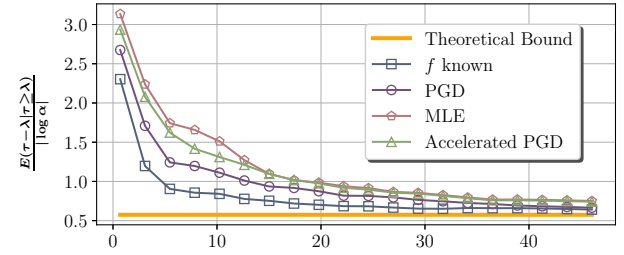


Fig. 6. Plots of the slope $\frac{1}{|\log \alpha|} \mathbb{E}(\tau - \lambda | \tau \geq \lambda)$ against $|\log \alpha|$ for outage detection in loopy 8-bus system (outage branch 4-7).

The detection rule in Theorem 2 can also restrict the false alarm rate below maximum tolerance $\alpha$. To verify this, we calculate the empirical false alarm rate $\mathbb{P}(\tau < \lambda)$ and compare it against the upper bound $\alpha$, as shown in Fig. 7. Our proposed method has similar performance compared to the case when $f$ is known since the empirical false alarm is mainly below the upper bound $\alpha$ (especially when $\alpha \rightarrow 0$). This observation demonstrates that our proposed algorithm could quickly detect line outages with a low false alarm rate, even when the post-outage distribution statistics are unknown.
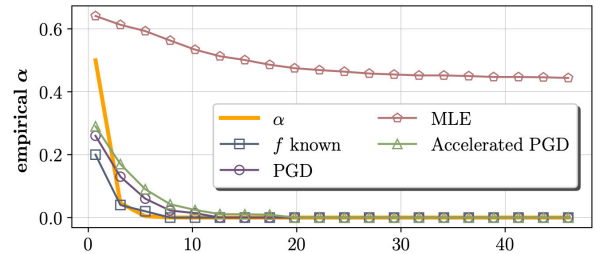


Fig. 7. Plots of the empirical false alarm rate against the theoretical probability of false alarm $\alpha$ in loopy 8-bus system (outage branch 4-7).

In Table II, we present a summary of our proposed method's performance in various grid systems under different outage configurations. Our method demonstrates the ability to handle diverse outage scenarios in both mesh and radial networks with DERs penetration. Specifically, when $f$ is unknown, our method exhibits a lower detection delay and significantly lower false alarm rate in comparison to MLE. Even when $f$ is given, our proposed method only experiences a slight degradation

TABLE II
PERFORMANCE COMPARISON ON VARIOUS SYSTEMS, $\alpha = 1\%$.

| System | | # Branches | # DERs | Outage | Average Detection Delay (1 unit) | | | | Empirical False Alarm Rate (%) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Optimal | $f$ known | PGD | MLE | $f$ known | PGD | MLE |
| | 8-bus | 9 | 0 | 4-7 | 2.70 | 3.13 | 4.28 | 5.21 | 0.9 | 3.1 | 44.2 |
| | 8-bus | 9 | 8 | 4-7 | 3.12 | 3.83 | 5.02 | 5.81 | 1.4 | 3.3 | 46.9 |
| Mesh network | 123-bus | 124 | 0 | 73-74 | 0.76 | 0.91 | 1.43 | 1.51 | 0.5 | 0.7 | 52.8 |
| | 123-bus | 124 | 0 | 73-74,14-15 | 0.73 | 0.87 | 1.35 | 1.48 | 0.4 | 0.7 | 51.3 |
| | 123-bus | 124 | 0 | 5 branches | 0.65 | 0.74 | 0.99 | 1.26 | 0.6 | 1.1 | 47.8 |
| (added loops) | LV suburban | 129 | 0 | 26-95 | 2.99 | 3.77 | 4.65 | 4.91 | 1.5 | 3.7 | 53.5 |
| | LV suburban | 129 | 30 | 26-95 | 3.24 | 3.92 | 4.85 | 5.14 | 1.3 | 2.4 | 48.0 |
| | MV urban | 48 | 0 | 34-35 | 0.56 | 0.78 | 1.34 | 1.31 | 0.3 | 0.6 | 46.1 |
| | MV urban | 48 | 7 | 34-35 | 1.01 | 1.58 | 2.04 | 2.33 | 0.7 | 1.1 | 45.9 |
| | 8-bus | 7 | 8 | 4-7 | 2.96 | 3.47 | 4.51 | 5.68 | 0.8 | 3.3 | 42.2 |
| | 8-bus | 7 | 8 | 2-6 | 3.02 | 3.60 | 4.73 | 5.81 | 0.8 | 2.5 | 45.4 |
| Radial network | 123-bus | 122 | 12 | 73-74 | 0.98 | 1.32 | 1.87 | 2.32 | 0.5 | 1.1 | 46.5 |
| | 123-bus | 122 | 122 | 73-74 | 6.23 | 6.90 | 7.73 | 9.24 | 2.7 | 4.9 | 69.3 |
| (with DERs) | LV suburban | 114 | 30 | 26-95 | 3.15 | 3.34 | 4.12 | 4.35 | 1.3 | 3.2 | 44.7 |
| | LV suburban | 114 | 113 | 26-95 | 8.62 | 9.33 | 10.46 | 12.70 | 3.8 | 5.4 | 65.8 |
| | MV urban | 38 | 7 | 34-35 | 0.88 | 1.45 | 1.97 | 2.12 | 0.6 | 1.1 | 49.3 |
| | MV urban | 38 | 7 | 23-35 | 1.02 | 1.73 | 2.24 | 2.49 | 1.1 | 1.7 | 43.0 |

compared to the benchmark. Furthermore, Table II reveals two additional phenomena. First, when multiple branches are out-of-service simultaneously, the average detection delay is shorter than in a single-line outage scenario due to the larger KL distance between distributions $g$ and $f$ when multiple lines are disconnected. Second, in the radial network with more simulated DERs, it takes more time to detect the line outage as the KL distance between $g$ and $f$ is smaller in this case.

To compare with more relevant methods in the literature, we provide in Table III the detection performance of our proposed method and other methods. The comparison of average detection delay and false alarm rate shows that our method is only slightly degraded from the benchmark even though we have incomplete information, and outperforms other methods that also has incomplete information. The reason for this is our performance guarantee, as stated in Theorem 5, which ensures the accurate estimation of unknown post-outage distribution parameters. Furthermore, upon comparing our approach (PGD) with the machine-learning-based method (DCQ), we notice that the latter displays a greater variance in the average detection delay and false alarm rate. This can be attributed to the fact that the neural network's parameters are randomly initialized during training, leading to a more varied estimation of the unknown post-outage distribution parameters.

TABLE III
COMPARISON WITH BENCHMARK ($f$ KNOWN) AND OTHER BASELINE METHODS DEALING WITH UNKNOWN $f$, $\alpha = 1\%$.

| Method | Average Detection Delay | False Alarm Rate(%) |
|---|---|---|
| $f$ known (benchmark) | $3.67 \pm 0.65$ | $0.05 \pm 0.03$ |
| PGD | $\mathbf{4.12} \pm 0.83$ | $\mathbf{1.06} \pm 0.37$ |
| MLE | $4.56 \pm 0.77$ | $38.2 \pm 7.39$ |
| GLRT | $5.03 \pm 1.17$ | $7.52 \pm 1.01$ |
| Shewhart | $4.78 \pm 0.84$ | $3.84 \pm 0.95$ |
| BGS | $4.85 \pm 0.81$ | $4.95 \pm 0.81$ |
| ULR | $4.69 \pm 1.01$ | $3.16 \pm 0.74$ |
| DIS | $5.03 \pm 0.95$ | $4.28 \pm 0.91$ |
| DCQ | $4.77 \pm 1.22$ | $5.16 \pm 1.08$ |

### C. Analysis of Execution Time for Timely Operation

In addition to detecting delay and false alarm rate, the execution time of the proposed method is also critical for timely detection. Table IV presents the execution time of Algorithm 1 on various grid systems with different sampling rates.

TABLE IV
EXECUTING TIME (SECOND) OF ALGORITHM 1.

| System (resolution) | $f$-known | PGD | Accelerated PGD |
|---|---|---|---|
| 8-bus (1 hour) | 0.138 | 1.029 | 0.468 |
| 123-bus (1 hour) | 0.453 | 2.019 | 1.176 |
| 8-bus (15 minute) | 0.234 | 1.263 | 0.582 |
| 123-bus (15 minute) | 0.579 | 2.671 | 1.236 |

From the records, less than 3 seconds per sample is needed to obtain the outage detection result when we receive a new sample, even for grid systems with more than 100 buses. This execution time can be negligible compared to the normal smart meter sampling interval, which ranges from 1 minute to 1 hour. Moreover, since the most time-consuming part of our algorithm is the matrix exponential and matrix logarithm operation, we can accelerate the algorithm by approximating these operations based on their Taylor series expansion, as discussed in Section V-B.

To maintain the detection performance, we select an appropriate level of approximation with near-zero errors incurred. In Fig. 8, we choose $K_{\mathrm{exp}} = 12$ because at this approximation level, the executing time is reduced by more than $75\%$ with almost zero errors incurred. Similarly, we choose $K_{\mathrm{log}} = 16$ which is slightly larger than $K_{\mathrm{exp}}$ since the term $\frac{1}{k}$ in (19) decreases slower than term $\frac{1}{k!}$ in (18) as $k$ becomes large. As a result, the accelerated PGD only shows a slight performance degradation, as shown in Fig. 6 and 7. More importantly, in Table IV, the acceleration technique reduces the execution time by more than half, thus achieving better timely outage detection. Evidently, the acceleration technique becomes more valuable in distribution systems with smart meters of a lower sampling rate, which is the trend of the future.
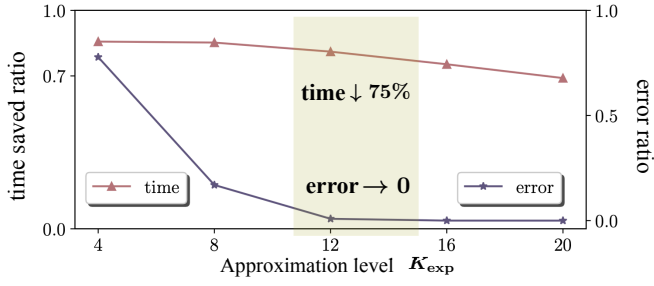
Fig. 8. The ratio of saved execution time versus the ratio of error incurred by the operation $\widehat{\exp}$ against the level of approximation $K_{\exp}$.

Table IV exhibits another phenomenon: as the sampling rate increases, the processing time for accumulated data $\Delta\mathbf{v}^{1:N}$ also increases. Consequently, conducting Algorithm 1 becomes challenging when $N$ grows very large. To address this issue, we discovered that a small window of historical samples can adequately differentiate between the pre- and post-outage distributions. Specifically, instead of using all $N$ samples when $N$ is very large, we can employ the latest $N_0$ samples to represent the entire data stream since they contain nearly identical distribution information in the temporal dimension. By doing so, the time complexity of the algorithm is restricted to a constant number, $N_0$. Through experiments, we determined that $N_0 = 100$ samples are sufficient to maintain the algorithm's effectiveness and accuracy.

### D. Outage Branch Localization with Accuracy

After detecting an outage occurrence, we further compute the conditional correlation between buses to localize the out-of-service branch, following Theorem 3. Here, Fig. 9 demonstrates the absolute conditional correlation of every pair of buses in the loopy 8-bus system before and after a line outage at branch 4-7. Since the value in the red box changes from a non-zero value before the outage ($\rho_{47}^- > \delta_{\max}$) to near zero after the outage ($\rho_{47}^+ < \delta_{\min}$), we localize the out-of-service branch at 4-7, which matches the ground truth. Fig. 9(d) indicates that the localization method using the learned covariance matrix through PGD is as effective as the optimal scenario, and is more effective than using the learned covariance matrix through MLE.

Table V demonstrates the accuracy rate of localization in 1,000 experiments. As shown, our proposed method can accurately localize over 90% of the outage branches, even without the post-outage distribution parameters.

TABLE V
LOCALIZATION ACCURACY (%) IN ALGORITHM 1.

| System | $f$-known | MLE | PGD |
|---|---|---|---|
| 8-bus | 98.3 | 92.5 | 93.6 |
| 123-bus | 94.7 | 88.2 | 91.8 |
| LV suburban | 95.6 | 90.9 | 92.3 |
| MV urban | 95.1 | 92.3 | 94.6 |

### E. Sensitivity Analysis to Data Noise and Data Coverage

Smart-meter data can be noisy and corrupted. Besides, smart-meter data may not be accessible in every household of the distribution grid. Thus, an analysis of our proposed method under different levels of data noise and data coverage is critical to gain a better understanding of its effectiveness in real-world outage scenarios.
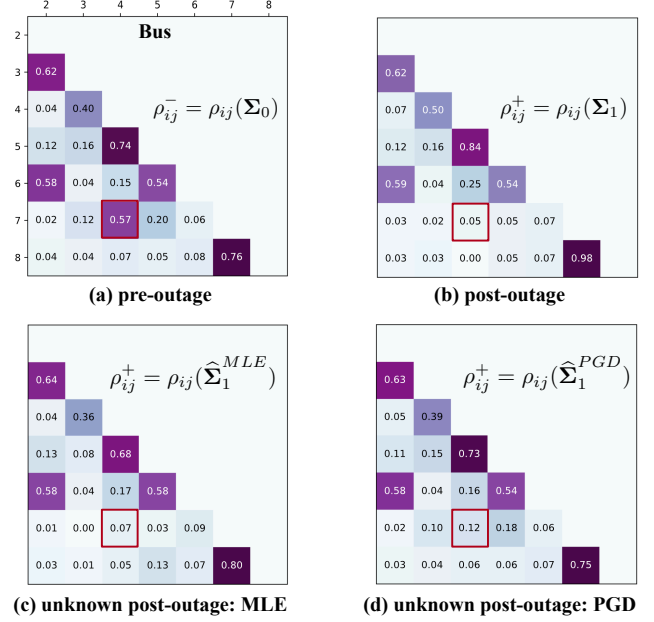


Fig. 9. Absolute conditional correlation of the loopy 8-bus system before and after an outage in branches 4-7. We choose $\delta_{\max} = 0.5$ and $\delta_{\min} = 0.1$.

In the U.S., ANSI C12.20 standard permits the utility smart meters to have an error within $\pm 0.5\%$ [40]. Hence, we simulate such noise in our smart-meter voltage measurements and then evaluate the corresponding detection performance. Table VI shows both average detection delay and false alarm rate under our method with different noise levels. As we see, when the noise level is $0.5\%$, one more sample (compared to noiseless case) is needed for the detection, while the false alarm rate is also slightly increased. In fact, we are able to quantify the increase in detection delay by analyzing the change of KL divergence between the pre- and post-outage distributions affected by noisy data. In doing so, we are able to better understand and control real-world line outage detection.

TABLE VI
OUTAGE DETECTION PERFORMANCE WITH NOISY DATA, $\alpha = 1\%$.

| Noise level | Average Detection Delay | False Alarm Rate |
|---|---|---|
| noiseless | 2.51 | 1.26% |
| 0.1% | 2.76 | 1.36% |
| 0.5% | 3.64 | 2.88% |

Another concern regarding the smart meter data is that it may not be accessible for every household in the distribution grid, particularly in certain situations. For instance, (1) in rural areas, some households may not have installed smart meters, (2) the voltage data for certain households may be lost due to technical issues, and (3) some households may refuse to provide their voltage data due to privacy concerns. Although the new generation of smart meters is developing very fast, an analysis of incomplete coverage of smart meters data is needed to evaluate our algorithm. We first emphasize that our proposed method does not rely on the assumption of 100% coverage of smart meters data in the grid. In fact, a power line outage will influence almost all buses in the system, while the degree of influence depends on the distance between a bus and the source of the outage. Hence, we can reveal the outage by detecting the distribution change of some (not necessarily all) voltage data collected nearby the outage source.

According to [41], over 107 million smart meters were deployed by 2021, covering 75% of U.S. households. Hence, we simulate this scenario where only a ratio of buses is randomly selected to provide its voltage measurements in the grid system to detect the outage. Fig. 10 demonstrates both the average detection delay and the false alarm rate of our method under different levels of coverage ratio. In comparison to the scenario where voltage data is available for all buses, the detection delay increases by 1.2 units of time step. This means that an extra 1.2 samples of data are needed to detect the outage in the 75% data coverage scenario. Similarly, when the data coverage ratio drops to 50%, an additional 6.9 samples are required for detection. Furthermore, as the data coverage ratio decreases to only 50%, the false alarm rate increases from 0.7% to 21.9%.
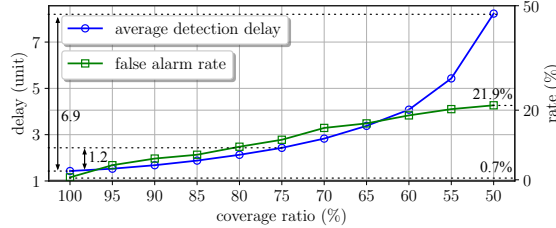


Fig. 10. Average detection delay (unit) and false alarm rate (%) under different levels of data coverage in loopy 123-bus system, $\alpha = 1\%$.

### F. Sensitivity Analysis to Hyper-parameters

Our detection procedure involves certain hyper-parameters that have the potential to influence the detection performance, such as the geometric distribution parameter $\rho$. Therefore, conducting a sensitivity analysis pertaining to these hyper-parameters is crucial to assess the robustness of our proposed method.

During our experiments, we randomly simulated the outage time $\lambda$ using a geometric prior distribution denoted as $\text{Geo}(\rho)$. This distribution aligns with our assumption that outages can take place in any time step with an equal probability of $\rho$. Fig. 11 illustrates the effect of the parameter $\rho$ on the performance of our detection method. It can be observed that choosing different values of $\rho$ within the range of 0.004 to 0.05 has a negligible impact on both the false alarm rate (approximately 1.65%) and the localization accuracy (approximately 92.8%). Additionally, decreasing the value of $\rho$ leads to a slight increase in the average detection delay.
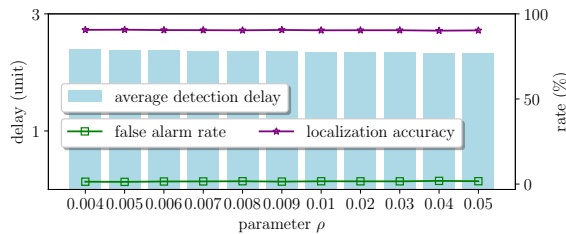


Fig. 11. Average detection delay (unit), false alarm rate (%) and localization accuracy (%) under different levels of $\rho$ in loopy 123-bus system, $\alpha = 1\%$.

### VII. LIMITATIONS

While this paper has some performance guarantee, we also encounter some of the limitations that we look forward to address in the future. For instance, while the proposed approach requires only voltage magnitude data, it may be limited by the quality and availability of this data. As shown in Section VI-E, noise or incomplete data will lead to additional detection delay. Future research could investigate how to leverage additional types of data to improve outage detection and localization. Another aspect worth investigating is the ability to withstand diverse outage scenarios. For instance, if an outage occurs in an insignificant branch of the grid, resulting in minimal fluctuations in voltage data, detecting such subtle outages remains a challenge. Hence, further research is necessary to improve the detection performance in such cases. Lastly, although sensor readings facilitate line outage detection, they pose privacy concerns since they can disclose sensitive information like household occupancy and economic status to potential adversaries. An open problem is how to identify outages accurately without compromising the customer's data.

### VIII. CONCLUSION

This paper resolves three challenges in the line outage identification problem: data availability, unknown outage pattern, and timely operation. Our approach for detecting and localizing line outages only utilizes voltage magnitude. To handle unknown outage patterns, we propose a Projected Gradient Descent framework that can learn the unknown post-outage distribution parameters with a feasibility guarantee. We demonstrate the convergence guarantee of our method and further accelerate the proposed algorithm for timely operation, resulting in a reduction of more than 75% of execution time with minimal errors. Empirical results on representative grid systems confirm that our proposed method is suitable for timely outage detection and localization, even in the absence of prior knowledge about outage patterns.

### APPENDIX A
### PROOF OF LEMMA 1

*Proof.* In (1), the incidence matrix $\boldsymbol{A}_{\mathcal{E},\mathcal{G}}$ has rank $M - 1$ in the connected grid [42]. The diagonal and full-rank complex matrix $\boldsymbol{Y}_{\mathcal{E}}$ can be decomposed as $\boldsymbol{Y}_{\mathcal{E}} = \boldsymbol{B}^{\top}\boldsymbol{B}$ where $\boldsymbol{B}$ is also full-rank. Then, $\text{rank}(\boldsymbol{A}_{\mathcal{E},\mathcal{G}}^{\top}\boldsymbol{Y}_{\mathcal{E}}\boldsymbol{A}_{\mathcal{E},\mathcal{G}}) = M - 1$ since

$$\boldsymbol{A}_{\mathcal{E},\mathcal{G}}^{\top}\boldsymbol{Y}_{\mathcal{E}}\boldsymbol{A}_{\mathcal{E},\mathcal{G}} = \boldsymbol{A}_{\mathcal{E},\mathcal{G}}^{\top}\boldsymbol{B}^{\top}\boldsymbol{B}\boldsymbol{A}_{\mathcal{E},\mathcal{G}} = (\boldsymbol{B}\boldsymbol{A}_{\mathcal{E},\mathcal{G}})^{\top}(\boldsymbol{B}\boldsymbol{A}_{\mathcal{E},\mathcal{G}}).$$

Hence, we have $\text{rank}(\boldsymbol{Y}_{\mathcal{G}}) = M - 1$ considering zero shunt admittance and $\text{rank}(\boldsymbol{Y}_{\mathcal{G}}) = M$ otherwise. In both cases, $\boldsymbol{Y}_{\mathcal{G}}$ has at least one non-singular $(M - 1) \times (M - 1)$ sub-matrix, which can be viewed as the remaining matrix after eliminating one column and one row (of a slack-bus) in $\boldsymbol{Y}_{\mathcal{G}}$. $\square$

### APPENDIX B
### PROOF OF THEOREM 5

*Proof.* Since $L(\boldsymbol{\mu}_1)$ is restricted convex, we apply this convexity property in constraint set $\mathcal{U}$ to give an upper bound to the level of sub-optimality of the $e^{th}$ iterate as

$$L_e = L(\boldsymbol{\mu}_1^{(e)}) - L(\boldsymbol{\mu}_1^{*}) \leq \langle \nabla L(\boldsymbol{\mu}_1^{(e)}), \boldsymbol{\mu}_1^{(e)} - \boldsymbol{\mu}_1^{*} \rangle. \quad (20)$$

We can obtain an upper bound of $\langle \nabla L(\boldsymbol{\mu}_1^{(e)}), \boldsymbol{\mu}_1^{(e)} - \boldsymbol{\mu}_1^{*} \rangle$ as

$$\frac{1}{2\eta}\left(\|\boldsymbol{\mu}_1^{(e)} - \boldsymbol{\mu}_1^{*}\|_2^2 + \eta^2 U^2 - \|\boldsymbol{\mu}_1^{e+1} - \boldsymbol{\mu}_1^{*}\|_2^2\right), \quad (21)$$

where the inequality holds since the gradient $\nabla L(\boldsymbol{\mu}_1)$ can be bounded on constraint set $\mathcal{U}$, i.e., $\|\nabla L(\boldsymbol{\mu}_1)\|_2 \leq U$ for all $\boldsymbol{\mu}_1 \in \mathcal{U}$. Combining equations (20) and (21), we arrive at

$$L_e \leq \frac{1}{2\eta}\left(\|\boldsymbol{\mu}_1^{(e)} - \boldsymbol{\mu}_1^*\|_2^2 - \|\boldsymbol{\mu}_1^{e+1} - \boldsymbol{\mu}_1^*\|_2^2\right) + \frac{\eta U^2}{2},$$

which upper bounds the sub-optimality at every $e^{th}$ iterate. Suppose we initialize the mean vector as $\boldsymbol{\mu}_1^1$, we can sum the sub-optimality across iterates and average it by dividing the total iterate number $E$ as $\frac{1}{E}\sum_{e=1}^{E} L_e \leq \frac{1}{2\sqrt{E}}\left(\|\boldsymbol{\mu}_1^* - \boldsymbol{\mu}_1^1\|_2^2 + U^2\right)$, where step size $\eta = \frac{1}{\sqrt{E}}$. Therefore, for any $\varepsilon > 0$, we can always use at most $E = \mathcal{O}(\frac{1}{\varepsilon^2})$ total iterates to make sure $\frac{1}{E}\sum_{e=1}^{E} L_e \leq \varepsilon$ since $\|\boldsymbol{\mu}_1^* - \boldsymbol{\mu}_1^1\|_2^2 + U^2$ is a constant number.

Then, we prove that the averaged and best update both converge to $\boldsymbol{\mu}_1^*$ after $E$ updates. Applying Jensen's inequality, we derive $L(\boldsymbol{\mu}_1^{\text{avg}}) = L(\frac{1}{E}\sum_{e=1}^{E}\boldsymbol{\mu}_1^{(e)}) \leq \frac{1}{E}\sum_{e=1}^{E} L(\boldsymbol{\mu}_1^{(e)}) \leq L(\boldsymbol{\mu}_1^*) + \varepsilon$. Since $L(\boldsymbol{\mu}_1^{\text{best}}) \leq L(\boldsymbol{\mu}_1^{(e)})$ for every $e^{th}$ iterate, we have $L(\boldsymbol{\mu}_1^{\text{best}}) \leq \frac{1}{E}\sum_{e=1}^{E} L(\boldsymbol{\mu}_1^{(e)}) \leq L(\boldsymbol{\mu}_1^*) + \varepsilon$. $\qquad\square$

## References

[1] A. N. Samudrala, M. H. Amini, S. Kar, and R. S. Blum, "Distributed outage detection in power distribution networks," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5124–5137, 2020.

[2] Energy Information Administration, "Average frequency and duration of electric distribution outages vary by states," https://www.eia.gov/todayinenergy/detail.php?id=35652, Accessed Sep. 2022.

[3] Y. Liao, Y. Weng, C.-W. Tan, and R. Rajagopal, "Quick line outage identification in urban distribution grids via smart meters," *CSEE Journal of Power and Energy Systems*, 2021.

[4] D. Bakken, *Smart grids: clouds, communications, open source, and automation*. CRC Press, 2014.

[5] A. Meier, T. Ueno, and M. Pritoni, "Using data from connected thermostats to track large power outages in the united states," *Applied Energy*, vol. 256, p. 113940, 2019.

[6] M. He and J. Zhang, "Fault detection and localization in smart grid: A probabilistic dependence graph approach," in *IEEE International Conference on Smart Grid Communications*, 2010, pp. 43–48.

[7] M. Babakmehr, F. Harirchi, A. Al-Durra, S. Muyeen, and M. G. Simões, "Compressive system identification for multiple line outage detection in smart grids," *IEEE Transactions on Industry Applications*, 2019.

[8] R. A. Sevlian, Y. Zhao, R. Rajagopal, A. Goldsmith, and H. V. Poor, "Outage detection using load and line flow measurements in power distribution systems," *IEEE Transactions on Power Systems*, 2017.

[9] H. Sun, Z. Wang, J. Wang, Z. Huang, N. Carrington, and J. Liao, "Data-driven power outage detection by social sensors," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2516–2524, 2016.

[10] P. Kankanala, S. Das, and A. Pahwa, "Adaboost: An ensemble learning approach for estimating weather-related outages in distribution systems," *IEEE Transactions on Power Systems*, vol. 29, no. 1, pp. 359–367, 2013.

[11] A. Dwivedi and A. Tajer, "Scalable quickest line outage detection and localization via graph spectral analysis," *IEEE Transactions on Power Systems*, vol. 37, no. 1, pp. 590–602, 2021.

[12] A. N. Shiryaev, "On optimum methods in quickest detection problems," *Theory of Probability & Its Applications*, vol. 8, no. 1, pp. 22–46, 1963.

[13] C. Wei, A. Wiesel, and R. S. Blum, "Change detection in smart grids using errors in variables models," in *IEEE Sensor Array and Multichannel Signal Processing Workshop*, 2012, pp. 17–20.

[14] Y. C. Chen, T. Banerjee, A. D. Domínguez-García, and V. V. Veeravalli, "Quickest line outage detection and identification," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 749–758, 2016.

[15] K. Gajula, V. Le, X. Yao, S. Zou, and L. Herrera, "Quickest detection of series arc faults on dc microgrids," in *IEEE Energy Conversion Congress and Exposition*, 2021, pp. 796–801.

[16] A. G. Tartakovsky and V. V. Veeravalli, "General asymptotic bayesian theory of quickest change detection," *Theory of Probability &amp; Its Applications*, vol. 49, no. 3, pp. 458–497, 2005.

[17] Y. Cao, L. Xie, Y. Xie, and H. Xu, "Sequential change-point detection via online convex optimization," *Entropy*, vol. 20, no. 2, p. 108, 2018.

[18] D. Siegmund and E. Venkatraman, "Using the generalized likelihood ratio statistic for sequential detection of a change-point," *The Annals of Statistics*, pp. 255–271, 1995.

[19] S. B. Fotopoulos, V. K. Jandhyala, and E. Khapalova, "Exact asymptotic distribution of change-point mle for change in the mean of gaussian sequences," *The Annals of Applied Statistics*, 2010.

[20] N. Duan, C. Huang, C.-C. Sun, and L. Min, "Smart meters enabling voltage monitoring and control: The last-mile voltage stability issue," *IEEE Transactions on Industrial Informatics*, 2021.

[21] L. M. Bregman, "Finding the common point of convex sets by the method of successive projection," in *Doklady Akademii Nauk*, vol. 162, no. 3. Russian Academy of Sciences, 1965, pp. 487–490.

[22] A. M. Kettner and M. Paolone, "On the properties of the power systems nodal admittance matrix," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 1130–1131, 2017.

[23] D. Deka, S. Backhaus, and M. Chertkov, "Structure learning in power distribution networks," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 3, pp. 1061–1074, 2017.

[24] S. Bolognani, N. Bof, D. Michelotti, R. Muraro, and L. Schenato, "Identification of power distribution network topology via voltage correlation analysis," in *IEEE Conference on Decision and Control*, 2013.

[25] Y. Liao, Y. Weng, G. Liu, and R. Rajagopal, "Urban mv and lv distribution grid topology estimation via group lasso," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 12–27, 2018.

[26] S. Jamali, A. Bahmanyar, and E. Bompard, "Fault location method for distribution networks using smart meters," *Measurement*, vol. 102, pp. 150–157, 2017.

[27] S. S. Gururajapathy, H. Mokhlis, and H. A. Illias, "Fault location and detection techniques in power distribution systems with distributed generation: A review," *Renewable and sustainable energy reviews*, vol. 74, pp. 949–958, 2017.

[28] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[29] K. Tsuda, G. Rätsch, and M. K. Warmuth, "Matrix exponential gradient updates for on-line learning and bregman projection," in *Advances in Neural Information Processing Systems*, 2004, pp. 1425–1432.

[30] P. Jain and P. Kar, "Non-convex optimization for machine learning," *arXiv preprint arXiv:1712.07897*, 2017.

[31] W. H. Kersting, "Radial distribution test feeders," *IEEE Transactions on Power Systems*, vol. 6, no. 3, pp. 975–985, 1991.

[32] C. Mateo, G. Prettico, T. Gómez, R. Cossent, F. Gangale, P. Frías, and G. Fulli, "European representative electricity distribution networks," *International Journal of Electrical Power & Energy Systems*, 2018.

[33] A. P. Dobos, "Pvwatts version 5 manual," National Renewable Energy Lab.(NREL), Golden, CO (United States), Tech. Rep., 2014.

[34] Y. Liao, A. S. Kiremidjian, R. Rajagopal, and C.-H. Loh, "Structural damage detection and localization with unknown postdamage feature distribution using sequential change-point detection method," *Journal of Aerospace Engineering*, vol. 32, no. 2, p. 04018149, 2019.

[35] Y. C. Chen, T. Banerjee, A. D. Dominguez-Garcia, and V. V. Veeravalli, "Quickest line outage detection and identification," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 749–758, 2015.

[36] G. Rovatsos, X. Jiang, A. D. Domínguez-García, and V. V. Veeravalli, "Comparison of statistical algorithms for power system line outage detection," in *2016 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2016, pp. 2946–2950.

[37] T. S. Lau, W. P. Tay, and V. V. Veeravalli, "Quickest change detection with unknown post-change distribution," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 3924–3928.

[38] J. Z. Hare, I. Kaplan, and V. V. Veeravalli, "Toward uncertainty aware quickest change detection," in *2021 IEEE 24th International Conference on Information Fusion (FUSION)*. IEEE, 2021, pp. 1–8.

[39] X. Ma, L. Lai, and S. Cui, "A deep q-network based approach for online bayesian change point detection," in *2021 IEEE 31st International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2021, pp. 1–6.

[40] J. Zheng, D. W. Gao, and L. Lin, "Smart meters in smart grid: An overview," in *IEEE Green Technologies Conference*, 2013, pp. 57–64.

[41] Institution for Electronic Innovation, "Electric company smart meter deployments: Foundation for a smart grid," https://www.edisonfoundation.net/-/media/Files/IEI/publications/IEI_Smart_Meter_Report_April_2021.ashx, Accessed Sep. 2022.

[42] R. B. Bapat, *Graphs and matrices*. Springer, 2010, vol. 27.