Identification of Outages in Power Systems With Uncertain States and Optimal Sensor Locations

Yue Zhao, *Member, IEEE*, Jianshu Chen, *Student Member, IEEE*, Andrea Goldsmith, *Fellow, IEEE*, and H. Vincent Poor, *Fellow, IEEE*

***Abstract*—Joint outage identification and state estimation in power systems is studied. A Bayesian framework is employed, and a Gaussian prior distribution of the states is assumed. The joint posterior of the outage hypotheses and the network states is developed in closed form, which can be applied to obtain the optimal joint detector and estimator under any given performance criterion. Employing the minimum probability of error as the performance criterion in identifying outages with uncertain states, the optimal detector is obtained. Efficiently computable performance metrics that capture the probability of error of the optimal detector are developed. Under simplified model assumptions, closed-form expressions for these metrics are derived, and these lead to a mixed integer convex programming problem for optimizing sensor locations. Using convex relaxations, a branch and bound algorithm that finds the globally optimal sensor locations is developed. Significant performance gains from using the optimal detector with the optimal sensor locations are observed from simulations. Furthermore, performance with greedily selected sensor locations is shown to be very close to that with globally optimal sensor locations.**

***Index Terms*—Joint detection and estimation, smart grid, outage identification, state estimation, PMU, sensor placement, Chernoff bound, mixed integer convex programming, branch and bound.**

I. Introduction

THE lack of situational awareness has been identified as a major cause of power system blackouts [3]. In particular, outages of power system components (e.g., transmission lines and transformers), if not timely identified, can escalate to cascading outages that quickly lead to large-scale islanding and loss of loads. Thus, as key functions of wide-area monitoring systems (WAMS), *outage detection and state estimation* are crucial for the grid to monitor its health and quickly react to power failures. For reliable and robust performance, it is clear that WAMS face an increasing need to process a diverse set of measurements arriving at disparate time scales. On the one hand, legacy WAMS have primarily relied on supervisory control and data acquisition (SCADA) systems, with conventional sensors often providing infrequent (ranging from seconds to minutes) and asynchronous measurements. On the other hand, phasor measurement units (PMUs) have also been increasingly deployed in wide area transmission networks. They are able to accurately measure voltage and current phasors at a high frequency (e.g., 30 measurements per second) with synchronized and accurate time stamps [4]. However, the high cost of PMUs has limited the scale of their deployment, and this raises the important question of *where to place* the PMUs to best improve the grid's situational awareness. Clearly, outage detection, state estimation and optimization of sensor locations are strongly coupled problems that jointly influence the performance of WAMS.

Outage detection in transmission networks has received much attention recently. A decision-tree based approach has been developed in [5], where outage detection is based on simulated "$n-1$" (and a few "$n-k$") contingencies and real-time measurements from PMUs. Also using PMU measurements, detection and identification of transmission line outages have been studied in [6], [7] and [1] as a hypothesis testing problem for a power system in a quasi-steady state (i.e., after fast system dynamics converge). In [8], this problem has been formulated as a sparse recovery problem for which efficient algorithms are developed. In all these works, a critical assumption is that the system states and/or power injections at all the buses are accurately and instantaneously known. However, this knowledge may not be accurately available in practical applications.

Indeed, state estimation has been a key function in WAMS for power transmission networks [9], [10]. Basic state estimation assumes prior knowledge of the network topology and parameters, and uses measurements with enough redundancy to infer the complete set of static or quasi-steady states of the network. There is also a bad data detection function that filters out erroneous measurements [9], [11]. However, once an outage happens (e.g., a line trips), the prior information will not reflect the actual network topology anymore, and the resulting state estimates can diverge significantly from the actual states because of model mismatch. Clearly, the coupled tasks of outage detection and state estimation are both inference tasks using *the same set of sensor measurements available*, together with prior information. An interesting issue is thus how these two tasks can

Manuscript received October 02, 2013; revised March 10, 2014; accepted July 01, 2014. Date of publication July 23, 2014; date of current version November 18, 2014. This work was supported in part by the Air Force Office of Scientific Researchunder MURI Grant FA9550-09-1-0643, in part by the Office of Naval Researchunder Grant N00014-12-1-0767, and in part by the DTRA under Grant HDTRA1-08-1-0010. Some preliminary results of this paper were presented in part at the IEEE Power and Energy Society General Meeting, San Diego, CA, 2012 [1]. Some results related to but different from this paper were presented at the IEEE Conference on Decision and Control, Florence, Italy, 2013 [2]. The guest editor coordinating the review of this manuscript and approving it for publication was Dr. Danilo Mandic.

Y. Zhao is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA, and also with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: yuez@stanford.edu).

J. Chen is with the Department of Electrical Engineering, University of California, Los Angeles, CA 90095 USA (e-mail: cjs09@ucla.edu).

A. Goldsmith is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: andrea@stanford.edu).

H. V. Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu).

Digital Object Identifier 10.1109/JSTSP.2014.2342191

be jointly and optimally performed. The problem of identifying outages jointly with state estimation first appeared in the seminal paper [12]. A similar problem has appeared in advanced state estimation with *topological error processing* [13], [14]: there, the state estimator takes into account a set of switches in the network whose status (open/close) is uncertain, and estimates the states together with the status of the switches. Another approach for topological error processing that exploits prior knowledge on sparsity of the errors is developed in [15]. In all these works, heuristics are provided, and optimality is not assured.

For optimizing sensor (and in particular PMU) locations, many studies have addressed this problem for different performance objectives, including network observability [16], outage detection [5]–[8], state estimation [17]–[19], and data attack detection [20], [21]. For each of these objectives, because of the combinatorial nature of the problem, finding the globally optimal sensor locations is in general NP hard, and consequently heuristics have been developed. In [18], it is shown that a good approximation ratio can be achieved by a greedy PMU placement algorithm for appropriate objectives.

In this paper, we model and solve joint outage identification and state estimation for a power system in a quasi-steady state. We focus on using linear approximations of the nonlinear power flow and measurement models. We employ a Bayesian framework and introduce prior distributions on the outage events and the network states. Assuming Gaussian prior conditional distribution of states and Gaussian noises, we develop in closed form the joint posterior distribution of the outage hypotheses and the network states. The joint posterior can then be applied to obtain an optimal joint outage detector and state estimator under any given criterion.

We further consider the performance criterion of *minimum probability of error* in identifying outages in the presence of uncertain states. To capture the probability of error in this multiple hypothesis testing problem, we first develop a Chernoff bound on the pairwise error probability. The developed bound is shown to be easily computable by solving a scalar convex optimization problem via bi-section. Under simplified model assumptions, the bound can be expressed in closed form. Based on these pairwise error bounds, three heuristics are developed for computing appropriate metrics that capture the actual probability of error. Based on the developed metrics, a greedy algorithm is first developed for optimizing sensor locations. Next, with the closed-form metrics derived under simplified model assumptions, the problem of sensor location optimization is formulated as a mixed integer-convex programming problem. Based on a convex relaxation of this problem, a branch and bound algorithm that finds the *globally optimal* sensor locations is then developed. From extensive simulation results, we see that the optimal detector that takes into account the uncertainty of states significantly outperforms the simpler one that *assumes* accurate knowledge of states. Furthermore, the performance gain from optimizing sensor locations is shown to be significant as well. The performance with greedily selected sensor locations is shown to be very close to that with globally optimal locations.

The remainder of the paper is organized as follows. In Section II, the system model is established. We formulate the problem of jointly performing outage identification and state estimation, and the problem of sensor location optimization. In Section III, we derive the joint posterior distribution of the outage hypotheses and the network states. In Section IV, we develop a Chernoff bound on the pairwise error probability of outage identification, based on which metrics that capture the actual probability of error are proposed. In Section V, algorithms for optimizing sensor locations are developed. Extensive simulation results are provided in Section VI. Conclusions are drawn in Section VII.

II. Problem Formulation

In a power transmission network, we consider outage identification and state estimation jointly performed based on a set of sensor measurements. We denote the scenario of a normal condition with no outage by $\mathcal{H}_0$. When outages occur (e.g., tripping of lines or transformers), the relation between sensor measurements and network states changes due to the changes of the network topology and parameters. We denote a set of outage events of interest by $\{\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_K\}$, also referred to as *outage hypotheses*, in the remainder of the paper. The following measurement model captures the coupled relations between measurements, outages and states:

$$\mathcal{H}_k : z = h_k(x) + v, \ k = 0, 1, \ldots, K, \tag{1}$$

where $z \in \mathbb{R}^M$ is the measurement vector, $x \in \mathbb{R}^N$ is the network state vector, v is the measurement noise vector, and h_k is the observation function that relates measurements to states under the particular outage hypothesis $\mathcal{H}_k$. In power systems, the states of the network usually refer to the voltage magnitudes and phase angles at all the buses. The measurements can include power, voltage magnitude and current magnitude measurements, as well as synchronized voltage phasor measurements [9]. The observation function h_k under hypothesis $\mathcal{H}_k$ depends on the topology and parameters of the power grid, as well as the types and locations of the sensors, which are typically known to the system operator. For typical forms of h_k, we refer the readers to [9] (cf. Chapter 2 therein).

If we know exactly the present outage event $\mathcal{H}_k$, estimating the states x from the measurements z is a standard state estimation problem [9]. However, if we mistakenly assume an incorrect outage hypothesis and hence a wrong observation model, the state estimates can arbitrarily differ from the actual states. Thus, in the presence of outages, identifying the true outage event is not only itself of paramount importance, but also a primary task on which state estimation critically depends. Ideally, if we know exactly all the network states x, identifying the true outage event from observing z reduces to a classic multiple hypothesis testing problem [22]. In practice, however, the states are only to be estimated from observing the same set of measurements z, and are not known accurately. Therefore, we study the problem of *identifying outages with uncertain states*, which is the key in joint inference of outage events and network states.

A. Linear Model

The observation functions $\{h_k\}$ are in general nonlinear in power systems, which further complicates outage identification with uncertain states. In this paper, we consider a *linear approximation* of $h_k, k = 0, \dots, K$:

$$\mathcal{H}_k : z = H_k x + v, \ k = 0, 1, \dots, K, \tag{2}$$

where $H_k \in \mathbb{R}^{M \times N}$ is an observation matrix under hypothesis $\mathcal{H}_k$. Such an approximation can be made by either applying a DC power flow model [23], or using the Jacobian of $h_k(x)$ around a crude estimate of the states x [9]. We note that this approximation can also be used as an intermediate step for solving the original nonlinear problem (1) in the following two ways:

- After obtaining an outage detection decision with a linear model (2), finer state estimation can then be performed with the nonlinear model (1) using a standard Gauss-Newton method [9].
- The latter approximation with a Jacobian can be applied iteratively within a standard Gauss-Newton procedure.

B. Power Injections as States in DC Power Flow Model

When applying the DC power flow model, the states are conventionally defined to be $x = \theta \in \mathbb{R}^N$, where θ contains the voltage phase angles at all the buses. Typically, the voltage phase angle at one of the buses (the reference bus) is set to zero, and we effectively have $N - 1$ state variables. From the DC model, we also have $P = B\theta$, where $P \in \mathbb{R}^N$ is the power injection vector at all the buses, and B is a weighted graph Laplacian of the power network based on its topology and line reactances [23]. We note that the power injections fully determine the network states via $\theta = B^\dagger P$, where $B^\dagger$ is the pseudoinverse of B. As a result, *we can also consider P as the network states without any ambiguity*. In this case, the linear model continues to hold with a new observation matrix $H_k B^\dagger$ under outage $\mathcal{H}_k$. Because the power injections must always be balanced (from the lossless assumption in DC power flow), i.e., $\sum_{n=1}^N P_n = 0$, we again have effectively $N - 1$ state variables in the system.

We will see in later sections that it is sometimes more convenient to consider power injections (i.e., generation and loads) as network states. This is because prior information on generation and loads are sometimes easier to obtain than that on voltage phase angles, and do not instantly change as outages occur.

C. A Bayesian Framework

We employ a Bayesian framework for joint outage identification and state estimation. Without loss of generality (WLOG), a joint prior distribution of the outage hypothesis and the network states can be written as

$$p(x, \mathcal{H}_k) = p(\mathcal{H}_k) p(x|\mathcal{H}_k), k = 0, \dots, K, \tag{3}$$

where $p(\mathcal{H}_k)$ denotes the prior probability mass function (PMF) of $\mathcal{H}_k$ and $p(x|\mathcal{H}_k)$ denotes the prior conditional probability density function (PDF) of x given $\mathcal{H}_k$. Similarly, given measurements z, the joint posterior distribution can be written as

$$p(x, \mathcal{H}_k|z) = p(\mathcal{H}_k|z) p(x|\mathcal{H}_k, z), k = 0, \dots, K. \tag{4}$$

The joint posterior is a sufficient statistic for joint detection and estimation. Thus, instead of being functions of z, the optimal decision rule and the optimal estimator need only be functionals of $p(x, \mathcal{H}_k|z)$ (see e.g., (12) below as a functional of $p(\mathcal{H}_k|z)$). After the joint posterior is computed, it can then be used to obtain optimal detection and estimation decisions under any given performance criterion. We will compute closed form expressions for the joint posterior (cf. (4)) in Section III.

Next, we focus on the performance criterion of *minimum probability of detection error* in identifying outages with uncertain states. This is because the outage detection decision to a large extent dictates the state estimation performance. Specifically, we find the optimal detection decision rule, denoted by $\delta^*(p(x, \mathcal{H}_k|z))$, that solves the following problem:

$$\min_{\delta} \mathrm{P_e}(\delta), \tag{5}$$

where $\mathrm{P_e}(\delta) \triangleq \sum_{k=0}^K p(\mathcal{H}_k) p(\delta(p(x, \mathcal{H}_k|z)) \neq k|\mathcal{H}_k)$. After finding δ^*, we characterize its performance by providing an efficiently computable performance metric that captures $\mathrm{P_e}(\delta^*)$. The developed performance metric not only reveals clear intuition, but also provides rigorous and effective guidance for finding the optimal locations to collect sensor measurements.

D. Optimal Sensor Locations

A different set of sensor locations will lead to a different set of observation matrices $\{H_k\}$. Sensor locations thus have a significant impact on the performance of outage identification and state estimation. With a constrained amount of sensing resources, selecting an optimal set of sensor locations becomes a crucial task. In practice, depending on the application scenarios, the time scale of selecting sensor locations can vary greatly. For example, for sensor installations that consider long term performance, sensor locations need to be optimized based on data of network outages and states collected over a long period. In contrast, given a set of sensors already installed, sensor polling for real time measurements operates on a much faster time scale. Both applications of sensor selection have strong economic incentives, as we would like to reduce the cost of sensor installation, as well as that of low latency communications for real time measurements.

To formulate the problem, first consider a *complete set of sensor candidates* with corresponding measurement vector $\bar{z}$ and observation matrices $\{\bar{H}_k\}$. This complete set embodies all possible sensor types and locations that we consider in the network. Specifically, each row of $\bar{z}$ and $\{\bar{H}_k\}$ corresponds to a possible measurement from a potential sensor of a certain type measuring at a particular location. Then, selecting M sensor locations becomes selecting M rows of $\bar{z}$, or equivalently, (the same set of) M rows of $\bar{H}_k$ for each $k = 0, \dots, K$. Because of the combinatorial nature of this problem, finding globally optimal sensor locations is in general NP hard.

Based on the developed performance metric, a greedy algorithm of sensor location selection is first introduced. Next, with simplifications of the metric, we formulate a mixed integer convex programming problem, and develop convex relaxations that can be efficiently solved. We finally exploit these relaxations to develop branch and bound algorithms that can effectively find the globally optimal sensor locations under simplified metrics.

III. Joint Posterior and Optimal Detector

We assume a Gaussian prior conditional PDF of the network states: *under outage hypothesis* $\mathcal{H}_k$,

$$p(x|\mathcal{H}_k) : x \sim \mathcal{N}(x_{k,0}, C_{k,0}), \tag{6}$$

where the conditional prior mean $x_{k,0}$ and covariance matrix $C_{k,0}$ can vary among different hypotheses $\mathcal{H}_k$s. In power systems, such priors can come from many different information sources, such as previous state estimation results, and knowledge of real time generation and loads obtained from sensing and/or forecast (particularly when we view power injections as network states in the DC power flow model). Moreover, as will be shown below, the joint prior distribution (3) and (6) is a *conjugate prior*, as it produces a posterior distribution of the same family. Thus, the prior information here can also be viewed as the posterior distribution obtained in our Bayesian framework from previous joint detection and estimation using past measurements. We assume Gaussian measurement noises $v \sim \mathcal{N}(0, R_v)$. Here, R_v does not depend on the hypothesis $\mathcal{H}_k$. This is because, in practice, R_v is primarily determined by sensor characteristics that are independent of outage events[1].

To compute the joint posterior (4), we first note that estimating x *given a hypothesis* $\mathcal{H}_k$ is a classic linear estimation problem. The posterior conditional PDF of x is given by (see e.g., [24])

$$p(x|\mathcal{H}_k, z) : x \sim N(\hat{x}_{k,\mathrm{MMSE}}, C_{k,\mathrm{MMSE}}), \tag{7}$$

where x's minimum mean square error (MMSE) estimate and its covariance matrix are

$$\begin{aligned} \hat{x}_{k,\mathrm{MMSE}} &\triangleq \left(C_{k,0}^{-1} + I(\hat{x}_{k,\mathrm{ML}})\right)^{-1} \\ &\quad \times \left(C_{k,0}^{-1} x_{k,0} + I(\hat{x}_{k,\mathrm{ML}})\hat{x}_{k,\mathrm{ML}}\right), \\ \text{and } C_{k,\mathrm{MMSE}} &\triangleq \left(C_{k,0}^{-1} + I(\hat{x}_{k,\mathrm{ML}})\right)^{-1}, \end{aligned} \tag{8}$$

with the maximum likelihood (ML) estimate and Fisher information matrix as

$$\begin{aligned} \hat{x}_{k,\mathrm{ML}} &= (H_k^T R_v^{-1} H_k)^{-1} H_k^T R_v^{-1} z, \\ \text{and } I(\hat{x}_{k,\mathrm{ML}}) &= H_k^T R_v^{-1} H_k. \end{aligned}$$

Note that the ML estimate reduces to a weighted least squares solution because of the linear Gaussian model. The derivation of (8) which combines the prior and the ML estimate to form the MMSE estimate can be found in [24] (cf. Chapter 3, Lemma 3.4.1 therein).

Next, we compute the posterior PMF of $\mathcal{H}_k$. Since $p(\mathcal{H}_k|z) = \frac{p(\mathcal{H}_k)p(z|\mathcal{H}_k)}{p(z)}$, it follows that we need to compute $p(z|\mathcal{H}_k)$. As we assumed a Gaussian conditional prior on x and a Gaussian noise of v, it follows that z also has a Gaussian distribution conditioned on each hypothesis $\mathcal{H}_k$. It is thus sufficient to compute the mean and covariance matrix of z given each $\mathcal{H}_k$. With some simple algebra, we have, $\forall k = 0, \ldots, K$,

$$\zeta_k \triangleq \mathbb{E}(z|\mathcal{H}_k) = H_k x_{k,0}, \tag{9}$$

$$\Sigma_k \triangleq \mathrm{Cov}(z|\mathcal{H}_k) = H_k C_{k,0} H_k^T + R_v. \tag{10}$$

[1]For the general case in which R_v depends on $\mathcal{H}_k$, the developed methods in this paper apply as well.

Accordingly, we have the following lemma:

Lemma 1: The posterior PMF of $\mathcal{H}_k$ can be computed by

$$p(\mathcal{H}_k|z) = \frac{p(\mathcal{H}_k)}{f(z)\det(\Sigma_k)^{1/2}} \exp\left(-\frac{1}{2}\|z - \zeta_k\|^2_{\Sigma_k^{-1}}\right), \tag{11}$$

where $f(z)$ is a normalization factor that keeps $\sum_{k=0}^{K} p(\mathcal{H}_k|z) = 1$, ζ_k and Σ_k are given by (9) and (10), and the notation $\|x\|^2_\Sigma$ denotes $x^T \Sigma x$ for any positive definite matrix Σ.

We note that, alternatively, (11) can be derived by computing $p(z|\mathcal{H}_k) = \int p(x|\mathcal{H}_k)p(z|x, \mathcal{H}_k)dx$. For details of this integration as well as the case of uninformative prior (i.e., $C_{k,0} \to \infty$), we refer the reader to [2]. In summary, Lemma 1 and (7) together give the complete expression of the joint posterior distribution of the outage events and the network states (4).

From now on, we employ the performance criterion of *minimum probability of error* for outage detection with uncertain states (5). The optimal detection is achieved by applying the maximum a-posteriori probability (MAP) rule [22]. From computing the posterior PMF of the outage hypotheses by Lemma 1, the optimal detection rule is thus

$$\delta^* \left(p(\mathcal{H}_k|z)\right) = \operatorname*{argmax}_k p(\mathcal{H}_k|z). \tag{12}$$

IV. Performance Metrics

Given the observation matrices $\{H_k\}$, we have developed the optimal outage detector based on the joint posterior of the outage hypotheses and the network states. The next design goal for a monitoring system is to *optimize the observation matrices* $\{H_k\}$ via, e.g., appropriately selecting sensor locations, in order to further improve the outage identification performance. For this, we first need to understand how the observation matrices $\{H_k\}$ affect such performance. This motivates the derivations of performance metrics in this section.

A. Bound on Pairwise Error Probability

The optimal detection rule (12) can have complicated decision regions that render computing the probability of error difficult. To characterize the probability of error, we start with bounding the *pairwise error probability* between any two of the $K + 1$ hypotheses: Suppose $\mathcal{H}_i$ and $\mathcal{H}_j$ are the only two hypotheses considered, and define

$$\mathrm{P}(\mathcal{H}_i \to \mathcal{H}_j) \triangleq \mathrm{P}(\delta^*(p(\mathcal{H}_k|z)) = j|\mathcal{H}_i), j \neq i, \tag{13}$$

i.e., the probability of claiming $\mathcal{H}_j$ when $\mathcal{H}_i$ is the ground truth. From the optimal detection rule (12), when $\mathcal{H}_i$ is true, $\mathcal{H}_j$ is declared if and only if $p(\mathcal{H}_j|z)/p(\mathcal{H}_i|z) > 1$,[2] and vice versa. Substituting (11), we arrive at the following lemma, whose proof is relegated to Appendix A.

Lemma 2 (Linear Quadratic Detector): The optimal detector is a linear quadratic detector:

$$\delta^*(z) = \begin{cases} \mathcal{H}_j, & \text{if } T(z) > \tau, \\ \mathcal{H}_i, & \text{otherwise}, \end{cases} \tag{14}$$

[2]We neglect the case $p(\mathcal{H}_j|z) = p(\mathcal{H}_i|z)$, since it happens with probability zero in our context.

where

$$T(z) \triangleq z^T \left(\Sigma_i^{-1} - \Sigma_j^{-1}\right) z + 2 \left(\zeta_j^T \Sigma_j^{-1} - \zeta_i^T \Sigma_i^{-1}\right) z, \tag{15}$$

$$\tau = -2 \ln \frac{p(\mathcal{H}_j)}{p(\mathcal{H}_i)} - \ln \frac{\det(\Sigma_i)}{\det(\Sigma_j)} + \|\zeta_j\|^2_{\Sigma_j^{-1}} - \|\zeta_i\|^2_{\Sigma_i^{-1}}, \tag{16}$$

with $\zeta_i, \zeta_j, \Sigma_i, \Sigma_j$ computed by (9) and (10).

We note that the above binary hypothesis testing problem falls into the category of the *general Gaussian problem* as introduced in [25], and linear quadratic detectors have appeared in many other applications as well. From Lemma 2,

$$\mathrm{P}(\mathcal{H}_i \to \mathcal{H}_j) = \mathrm{P}(T(z) > \tau | \mathcal{H}_i). \tag{17}$$

Now, we provide an upper bound on $\mathrm{P}(\mathcal{H}_i \to \mathcal{H}_j)$ (17) by developing a Chernoff bound: $\forall s > 0$,

$$\begin{aligned} \mathrm{P}(T(z) > \tau | \mathcal{H}_i) &= \mathrm{P}(e^{sT(z)} > e^{s\tau} | \mathcal{H}_i) \\ &\leq e^{-s\tau} \mathbb{E}\left(e^{sT(z)} | \mathcal{H}_i\right) = \exp(-s\tau + \mu_{T,i}(s)), \end{aligned} \tag{18}$$

where $\mu_{T,i}(s) \triangleq \ln\left(\mathbb{E}\left(e^{sT(z)} | \mathcal{H}_i\right)\right)$, and (18) is from Markov's Inequality. Note that, when $s = 0$, we have that $\mu_{T,i}(s) = 0$, and $\exp\left(-s\tau + \mu_{T,i}(s)\right) = 1$ always upper bounds $\mathrm{P}(T(z) > \tau | \mathcal{H}_i)$. It is immediate to prove that $\mu_{T,i}(s)$ is a *convex* function of $s (\geq 0)$ in the support of s such that $\mu_{T,i}(s) < \infty$. After substituting (15) for $T(z)$, the following theorem provides a closed form expression of $\mu_{T,i}(s)$, whose proof is relegated to Appendix B.

Theorem 1:

$$\mu_{T,i}(s) = \begin{cases} \frac{1}{2} b^T(s) A(s) b(s) + \frac{1}{2} \ln \det(A(s)) + c, \\ \qquad\qquad\qquad\qquad \text{if } 0 \leq s < \bar{s}, \\ \infty, s \geq \bar{s}, \end{cases} \tag{19}$$

where

$$\bar{s} = \begin{cases} \lambda_0^{-1}, \text{if } \lambda_0 \triangleq \lambda_{\max}\left(2(\Sigma_i^{-1} - \Sigma_j^{-1}), \Sigma_i^{-1}\right) > 0, \\ \infty, \text{if } \lambda_0 \leq 0, \end{cases} \tag{20}$$

$$A(s) = \left(\Sigma_i^{-1} - 2s\left(\Sigma_i^{-1} - \Sigma_j^{-1}\right)\right)^{-1}, \tag{21}$$

$$b(s) = 2s\left(\Sigma_j^{-1}\zeta_j - \Sigma_i^{-1}\zeta_i\right) + \Sigma_i^{-1}\zeta_i, \tag{22}$$

$$c = -\frac{1}{2}\left(\zeta_i^T \Sigma_i^{-1} \zeta_i + \ln \det(\Sigma_i)\right), \tag{23}$$

with the notation $\lambda_{\max}(X, Y)$ in (20) denoting the largest generalized eigenvalue of the matrix pair (X, Y).

Because $\mu_{T,i}(s)$ is a convex function, the optimal s that minimizes the error bound (18),

$$s_{ij}^* \triangleq \operatorname*{argmin}_{s \geq 0} \{-s\tau + \mu_{T,i}(s)\}, \tag{24}$$

can be found efficiently with a bi-section algorithm [26]. Accordingly, we have the following Chernoff bound on the pairwise error probability:

$$\mathrm{P}(\mathcal{H}_i \to \mathcal{H}_j) \leq \exp\left(-s_{ij}^* \tau + \mu_{T,i}(s_{ij}^*)\right) \triangleq P_{ij}. \tag{25}$$

Here, s_{ij}^* does not yet have an explicit form. To analytically understand how the observation matrices $\{H_k\}$ affect the detection performance, we seek approximate but explicit form of the pairwise error bound (25) in the next subsection.

B. Performance Metrics With Simplified Model Assumptions

We now consider a simpler case in which we assume a *perfect* prior on network states x as follows:

$$C_{k,0} = \mathbf{0}, \forall k = 0, 1, \ldots, K, \tag{26}$$

where $\mathbf{0}$ is the all-zero matrix. Assuming accurate knowledge in state estimates, we particularly focus on the problem of outage detection. In the DC power flow model, if we view power injections as states (cf. Section II-B), (26) can also mean that we assume accurate knowledge of the generation and loads in the network. We note that this assumption is an approximation, and is made for a power system in a quasi-steady state [6], [10]. This is a good approximation when we have real time sensor measurements that can closely track the network states prior to an outage, and then the main issue becomes the potential network topological change due to outages.

Equation (26) leads to a simpler optimal detector as follows. From (10), $\Sigma_k = R_v, \forall k = 0, \ldots, K$, and the linear quadratic detector (14) (cf. Lemma 2) reduces to a linear detector with

$$\begin{aligned} T(z) &= 2(\zeta_j - \zeta_i)^T R_v^{-1} z, \\ \tau &= -2 \ln \frac{p(\mathcal{H}_j)}{p(\mathcal{H}_i)} + \|\zeta_j\|^2_{R_v^{-1}} - \|\zeta_i\|^2_{R_v^{-1}}. \end{aligned}$$

The computation of error bounds also greatly simplifies to an explicit form (as opposed to being algorithmically solved by a convex optimization), as shown in the following corollary whose proof is relegated to Appendix C:

Corollary 1: With a perfect prior on network states (26),

$$\mathrm{P}(\mathcal{H}_i \to \mathcal{H}_j) \leq \exp\left(-\frac{\left(\|\zeta_j - \zeta_i\|^2_{R_v^{-1}} - 2\ln\frac{p(\mathcal{H}_j)}{p(\mathcal{H}_i)}\right)_+^2}{8\|\zeta_j - \zeta_i\|^2_{R_v^{-1}}}\right), \tag{27}$$

where $(x)_+ \triangleq \max\{0, x\}$.

Furthermore, if we have no knowledge about the outage hypotheses, it is typical to assume a *uniform* prior on the outage hypotheses. Now that $p(\mathcal{H}_j) = p(\mathcal{H}_i)$, we have

$$\mathrm{P}(\mathcal{H}_i \to \mathcal{H}_j) \leq \exp\left(-\frac{1}{8}\|\zeta_j - \zeta_i\|^2_{R_v^{-1}}\right). \tag{28}$$

The intuition behind (28) is clear: as a *weighted distance* between the two measurement mean vectors under the two hypotheses increases, the error probability decreases monotonically. The lower the measurement noise is, the larger the weighted distance is.

Substituting (9) into (28), the relation between the observation matrices $\{H_k\}$ and the pairwise error bound is then crystalized. As will be shown later, these simplifications on the performance metric provide valuable insight and great convenience in the optimization of sensor locations.

C. From Pairwise Error Probability to Probability of Error

We have discussed bounds on pairwise error probabilities. From these, we essentially obtain a matrix of pairwise error bounds, $[\mathrm{P}_{ij}] \in \mathbb{R}^{(K+1)\times(K+1)}$, where P_{ij} is given by the right hand side of (25), which upper bounds $\mathrm{P}(\mathcal{H}_i \to \mathcal{H}_j), \forall i, j$. Based on this pairwise error bound matrix, the next step is to approximate the probability of error $\mathrm{P_e}(\delta^*)$ (5) in the original problem. A series of heuristics for computing performance metrics on $\mathrm{P_e}(\delta^*)$ are as follows:

- *"Sum-Sum" metric (union bound)*: A standard technique is to apply a union bound on all the error events $\mathcal{H}_i \to \mathcal{H}_j, j \neq i$ for each particular $\mathcal{H}_i$, and we have

$$p\left(\delta^*(z) \neq i | \mathcal{H}_i\right) \leq \sum_{j\neq i} \mathrm{P}(\mathcal{H}_i \to \mathcal{H}_j) \leq \sum_{j\neq i} \mathrm{P}_{ij},$$
$$\Rightarrow \mathrm{P_e}\left(\delta^*(z)\right) \leq \sum_{i=0}^{K} p(\mathcal{H}_i) \sum_{j\neq i} \mathrm{P}_{ij}. \tag{29}$$

  Since this bound is the weighted sum of the sums within each row of $[\mathrm{P}_{ij}]$, we term it the "Sum-Sum" metric, denoted by $\mathrm{P}_{\mathrm{e}_{\mathrm{sum}}}^{\mathrm{sum}}$.
- *"Sum-Max" metric*: As opposed to summing over all error events, we can instead use the following approximation:

$$p\left(\delta^*(z) \neq i | \mathcal{H}_i\right) \approx \max_{j\neq i} \mathrm{P}(\mathcal{H}_i \to \mathcal{H}_j) \leq \max_{j\neq i} \mathrm{P}_{ij},$$
$$\Rightarrow \mathrm{P_e}\left(\delta^*(z)\right) \approx \sum_{i=0}^{K} p(\mathcal{H}_i) \max_{j\neq i} \mathrm{P}_{ij}. \tag{30}$$

  This metric chooses a single dominant error event under each hypothesis, and then computes across all hypotheses a weighted sum of $K+1$ dominant pairwise error bounds. Accordingly, we term it the "Sum-Max" metric, denoted by $\mathrm{P}_{\mathrm{e}_{\max}}^{\mathrm{sum}}$.
- *"Max-Max" metric*: Finally, a cruder approximation of $\mathrm{P_e}(\delta^*(z))$ is as follows:

$$\begin{aligned}\mathrm{P_e}\left(\delta^*(z)\right) &\approx \max_{i=0,\ldots,K} \left(p(\mathcal{H}_i) \max_{j\neq i} \mathrm{P}(\mathcal{H}_i \to \mathcal{H}_j)\right) \\ &\leq \max_{i=0,\ldots,K} \left(p(\mathcal{H}_i) \max_{j\neq i} \mathrm{P}_{ij}\right).\end{aligned} \tag{31}$$

  This metric chooses one single dominant error event (weighted by prior probabilities of the hypotheses) among all the $(K+1)K$ error events, and uses its error bound to compute a proxy of the probability of error. Accordingly, we term it the "Max-Max" metric, denoted by $\mathrm{P}_{\mathrm{e}_{\max}}^{\max}$.

When evaluating the above performance metrics, it is important to note that the behavior of the measurement means $\{\zeta_k\}$ and covariances $\{\Sigma_k\}$ from all the outage hypotheses heavily depends on the physical network topology and parameters, the set of outage events, the prior on the states, and the sensor locations. This behavior can be quite arbitrary with a large number of outage hypotheses: For example, the constellations of the hypothesis means $\{\zeta_k\}$ can have irregular shapes, (which are quite unlike, e.g., the well designed regular signal constellations in communication systems). Furthermore, as we identify outages with uncertain states and have limits on sensor accuracies, $\{\Sigma_k\}$ can be relatively sizable in practice. As a result, we must sometimes work in the non-asymptotic regime of the detection problem in which the probability of error is not very close to zero (see further details in Section VI below).

Because of the above practical observations, evaluating how well the approximate metrics work analytically is difficult. Instead, as shown later in Section VI, we use extensive simulations to evaluate how different metrics capture the actual probabilities of error. The ultimate purpose of deriving the metrics in this section is to provide a guidance with which sensor locations can be optimized.

V. Finding the Optimal Sensor Locations

Any potential sensor at any location, if selected, can contribute a measurement that is a linear function of the network states (plus noise) given each outage hypothesis (each outage leads to a different linear observation function due to the specific topological change that it causes). We denote the set of all potential locations to collect sensor measurements by $\bar{\mathcal{M}}$ with $|\bar{\mathcal{M}}| = \bar{M}$. We define $\bar{z} \in \mathbb{R}^{\bar{M}}, \bar{H}_k \in \mathbb{R}^{\bar{M}\times N}$, and $\bar{v} \in \mathbb{R}^{\bar{M}}$ with

$$\mathcal{H}_k : \bar{z} = \bar{H}_k x + \bar{v}, k = 0, 1, \ldots, K, \tag{32}$$

to capture all the $\bar{M}$ potential sets of measurement equations (each sensor gives one set of $K+1$ equations for the $K+1$ hypotheses). In other words, a sensor at a particular location provides *one row in the observation matrix* $\bar{H}_k$ under hypothesis $\mathcal{H}_k, k = 0, \ldots, K$. If we are limited to $M \leq \bar{M}$ sensors, choosing a set of locations is equivalent to choosing M among all the $\bar{M}$ sets of measurement equations. Given a set of sensor locations, a set of observation matrices $\{H_k\}$ is determined, and the detection performance with this set of sensors can be evaluated using the metrics discussed in the last section. Clearly, given M, finding the set of M locations with the best detection performance is a combinatorial optimization that has a worst case of $\binom{\bar{M}}{M}$ complexity.

A. A Greedy Algorithm With General Metrics

With the three metrics developed in the last section, we first have a greedy algorithm as in Algorithm 1 that generates a series of sensor location sets $\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_{\bar{M}}$ for the number of sensors $M = 1, 2, \ldots, \bar{M}$ respectively, that satisfy the following consistency property:

$$\mathcal{M}_1 \subset \mathcal{M}_2 \subset \ldots \subset \mathcal{M}_{\bar{M}} = \bar{\mathcal{M}}. \tag{33}$$

We illustrate the algorithm with the Sum-Max metric as a function of the sensor locations $\mathcal{M}$: $\mathrm{P}_{\mathrm{e}_{\max}}^{\mathrm{sum}}(\mathcal{M})$.

Algorithm 1: Greedy Sensor Location Selection

Initialize the set of sensors $\mathcal{M}_0 = \emptyset$,
and the number of sensors $M = 0$.
Repeat
$M \leftarrow M + 1$,

$$\mathcal{M}_M = \mathcal{M}_{M-1} \cup \left\{ \underset{m\in\bar{\mathcal{M}}\setminus\mathcal{M}_{M-1}}{\operatorname{argmin}} \ \mathrm{P}_{\mathrm{e}_{\max}}^{\mathrm{sum}}\left(\mathcal{M}_{M-1} \cup \{m\}\right) \right\} \tag{34}$$

Until $M = \bar{M}$.

In other words, we choose up to $\bar{M}$ sensor locations one by one: At each step, we keep the already chosen locations; from the remaining locations, we choose the one that *minimizes the current step's Sum-Max metric on probability of error*, and include it in the set of the chosen locations.

Now, toward finding *globally* optimal sensor locations, the general pairwise error bounds (25) as building blocks of the performance metrics, although efficient to compute, do not explicitly exhibit how the sensor locations affect the performance, and hence do not render a tractable combinatorial optimization. However, with the simplified metrics developed in Section IV-B, we show next that the problem can be formulated as a mixed integer convex programming (MICP) problem, for which convex relaxation and branch and bound algorithms are developed.

B. Mixed Integer Convex Programming

Consider the case in which we assume a perfect prior on network states (26) and a uniform prior on the outage hypotheses, the pairwise error bound is given by (28). We further assume that the measurement noises v are independently (but not necessarily identically) distributed across different sensors, i.e., R_v is a diagonal covariance matrix.

Now, given the set of all potential sensor locations $\bar{\mathcal{M}}$, we define $\forall k, \bar{\zeta}_k = \mathbb{E}(\bar{z}|\mathcal{H}_k) = \bar{H}_k x_{k,0} \in \mathbb{R}^{\bar{M}}$ as in (9), and $\bar{r}_v = [r_1^2, r_2^2, \ldots, r_{\bar{M}}^2]^T$ as the vector of measurement noise variances from all the potential sensors. Given a set of sensor locations $\mathcal{M}$, we define a diagonal *location indicator matrix*:

$$W = \operatorname{diag}(w_1, \ldots, w_{\bar{M}}), \text{with } w_n = \begin{cases} 1, & \text{if } n \in \mathcal{M} \\ 0, & \text{otherwise} \end{cases},$$

and term $\{w_n\}$ the *location indicator variables*. Further define $\zeta_k^{\mathcal{M}} \in \mathbb{R}^M$ and $r_v^{\mathcal{M}} \in \mathbb{R}^M$ as the sub-vectors of $\bar{\zeta}_k$ and $\bar{r}_v$ by extracting the M entries whose indices are in $\mathcal{M}$, respectively. It is then straightforward to show that, $\forall 0 \leq i < j \leq K$,

$$\begin{aligned} &\|\zeta_j^{\mathcal{M}} - \zeta_i^{\mathcal{M}}\|^2_{\operatorname{diag}^{-1}(r_v^{\mathcal{M}})} \\ =&(\bar{\zeta}_j - \bar{\zeta}_j)^T \operatorname{diag}^{-1/2}(\bar{r}_v)\, W \operatorname{diag}^{-1/2}(\bar{r}_v)(\bar{\zeta}_j - \bar{\zeta}_j). \end{aligned} \tag{35}$$

Note that (35) is a *linear* function of the location indicator variables $\{w_n\}$ in the diagonal of W. Thus, the pairwise error bound (28) $\mathrm{P}_{ij} = \exp\left(-\frac{1}{8}\|\zeta_j^{\mathcal{M}} - \zeta_i^{\mathcal{M}}\|^2_{\operatorname{diag}^{-1}(r_v^{\mathcal{M}})}\right)$ is a *convex* function of $\{w_n\}$.

Because summation and pointwise maximum both preserve convexity [26], the Sum-Sum, Sum-Max and Max-Max metrics (cf. (29), (30) and (31)) are all convex functions of $\{w_n\}$. Consequently, for finding the optimal set of M sensor locations, we arrive at the following mixed integer convex programming problem (illustrated with the Sum-Max metric):

$$\min_{w_n, n=1,\ldots,\bar{M}} \sum_{i=0}^{K} p(\mathcal{H}_i) \max_{j \neq i} \mathrm{P}_{ij} \tag{36a}$$

$$s.t.\ w_n \in \{0,1\}, n = 1, \ldots, \bar{M}, \tag{36b}$$

$$\sum_{n=1}^{\bar{M}} w_n = M. \tag{36c}$$

C. Convex Relaxation and a Branch and Bound Algorithm

Noting that the only nonconvexity in the above problem (36a) lies in the integer constraints on $\{w_n\}$ (36b), we have the following relaxation of it as a convex optimization that can be solved efficiently:

$$\min_{w_n, n=1,\ldots,\bar{M}} \sum_{i=0}^{K} p(\mathcal{H}_i) \max_{j \neq i} \mathrm{P}_{ij} \tag{37a}$$

$$s.t.\ 0 \leq w_n \leq 1, n = 1, \ldots, \bar{M},$$

$$\sum_{n=1}^{\bar{M}} w_n = M. \tag{37b}$$

Accordingly, the optimal value of (37a) serves as a *lower bound*, denoted by L_1, on the global optimum of (36a). Meanwhile, Algorithm 1 provides an *upper bound*, denoted by U_1.

Remark 1 (A Note on the Rounding Heuristic): Another upper bounding heuristic is to find an integral solution by *rounding* the fractional solution obtained from the relaxed problem (37a). In particular, we consider the heuristic by rounding *the M largest* fractional entries to 1, and the others to 0. In our simulations, we found that this rounding heuristic is outperformed by the greedy algorithm. The reason is that, in the relaxed fractional solution $\{w_n\}$, it is often some *very small* non-zero w_n that is *critical* in the following sense: losing it will drastically reduce the weighted mean distance (cf. (28)) in some critical pairwise error event, and hence leads to significant increase in the error probability between this critical hypothesis pair. Consequently, for upper bounding the global optimum, we propose the greedy algorithm instead of the rounding heuristic, as the former is both much cheaper computationally and much better in performance. Thus, we use the relaxation technique *not* to provide a fractional solution to round, but to *lower bound* the global optimum and to develop a branch and bound algorithm in the following that can significantly improve the greedy solutions in a few iterations.

For any sensor location m, (36a) can be split into two sub-problems by fixing w_m to be either 0 or 1:

$$\min_{w_n, n=1,\ldots,\bar{M}} \sum_{i=0}^{K} p(\mathcal{H}_i) \max_{j \neq i} \mathrm{P}_{ij} \tag{38a}$$

$$s.t.\ w_n \in \{0,1\}, n = 1, \ldots, \bar{M},$$

$$\sum_{n=1}^{\bar{M}} w_n = M, w_m = 0. \tag{38b}$$

and

$$\min_{w_n, n=1,\ldots,\bar{M}} \sum_{i=0}^{K} p(\mathcal{H}_i) \max_{j \neq i} \mathrm{P}_{ij} \tag{39a}$$

$$s.t.\ w_n \in \{0,1\}, n = 1, \ldots, \bar{M},$$

$$\sum_{n=1}^{\bar{M}} w_n = M, w_m = 1. \tag{39b}$$

Similarly to (36a), relaxations of these two sub-problems can be formed by replacing (38b) and (39b) with (37b), and they provide lower bounds, denoted by $l_2^{(0)}$ and $l_2^{(1)}$, on the global optimum of (38a) and (39a) respectively. Meanwhile, applying the greedy heuristic under the constraint $w_m = 0$ or $w_m =$

1 provides upper bounds, denoted by $u_2^{(0)}$ and $u_2^{(1)}$, on these sub-problems' global optima. Define

$$L_2 \triangleq \min\{l_2^{(0)}, l_2^{(1)}\}, \text{and } U_2 \triangleq \min\{u_2^{(0)}, u_2^{(1)}\}. \tag{40}$$

Then, L_2 and U_2 are new lower and upper bounds on the original global optimum (36a) [27].

More generally, the above splitting procedure with relaxations and greedy heuristics can be applied on the sub-problems themselves to form more children sub-problems with lower and upper bounds. For example, for any location $s, (s \neq m,)$ (36a) can be further split into two sub-problems by adding yet another constraint $w_s = 0$ or $w_s = 1$ respectively.

We define the following lower and upper bounding oracles, as well as an oracle that returns the next location to split:

Definition 1: Oracle $LB(\mathcal{C})$ takes a *constraint set* $\mathcal{C}$ as input, where $\mathcal{C}$ specifies a set of locations whose indicator variables are pre-determined to be either 0 or 1. An MICP under the constraints $\mathcal{C}$ is formed, a relaxation is solved, and the optimum of this relaxation is output by $LB(\mathcal{C})$ as a lower bound on the optimum of the constrained MICP.

For example, in (38a) and (39a), the constraint sets are $\mathcal{C}^{(0)} = \{w_m = 0\}$ and $\mathcal{C}^{(1)} = \{w_m = 1\}$, respectively.

Definition 2: Oracle $UB(\mathcal{C})$ takes a constraint set $\mathcal{C}$ as input. An MICP under the constraints $\mathcal{C}$ is formed, a greedy solution is found by Algorithm 1, and the achieved objective value is output by $UB(\mathcal{C})$ as an upper bound on the optimum of the constrained MICP.

Definition 3: Based on the *order* of the locations chosen by Algorithm 1, Oracle $\mathrm{next}(\mathcal{C})$ outputs *the first* location that is chosen by this heuristic.

When a sub-problem with constraints $\mathcal{C}$ needs to be split further, $\mathrm{next}(\mathcal{C})$ is the location we choose to perform the splitting by fixing $w_{\mathrm{next}(\mathcal{C})}$ to be either 0 or 1.

We now provide a branch and bound algorithm as in Algorithm 2 where $i_{\max}$ is the maximum number of iterations allowed. As the algorithm progresses, *a binary tree is developed where each node represents a constraint set*. The *leaf nodes* are kept in $\mathcal{S}$. The tree starts with a single node with an empty constraint set. When a sub-problem corresponding to a leaf node $\mathcal{C}^*$ is split into two new sub-problems, the two new constraint sets $\mathcal{C}^{(0)}$ and $\mathcal{C}^{(1)}$ become the children of the parent constraint set $\mathcal{C}^*$.

In Algorithm 2, (42) is a generalization of (40). This means that the current global lower bound equals the *lowest* lower bound among all the *leaf node* constraint sets. This is true because all the leaf nodes $\mathcal{S}$ represent a *complete partition* of the original parameter space [27]. At the beginning of every iteration, in choosing which leaf node to split (41), we select the one that gives the *lowest* lower bound (i.e., the current *global* lower bound). It is a heuristic based on the reasoning that, by further splitting this critical leaf node, a higher global lower bound may be obtained (whereas splitting any other node will leave the global lower bound unchanged). At iteration i, the current lower and upper bounds on the global optimum are available as L_i and U_i. When these two bounds meet, i.e., $U_i - L_i < \epsilon$, the solution that achieves the current upper bound is guaranteed to be globally optimal.

Algorithm 2: Sensor Location Selection using Branch and Bound

Initial step: $i = 1$,
 the initial constraint set: $\mathcal{C}_1 = \emptyset$,
 the initial set of leaves of the tree of constraint sets
 (initially a single node): $\mathcal{S} = \{\mathcal{C}_1\}$.
 Compute $L_1 = LB(\mathcal{C}_1), U_1 = UB(\mathcal{C}_1)$.
While $U_i - L_i > \epsilon$ or $i < i_{\max}$, repeat
 Choose which leaf node constraint set to split:

$$\mathcal{C}^* = \operatorname*{argmin}_{\mathcal{C}\in\mathcal{S}}\{LB(\mathcal{C})\}. \tag{41}$$

 Choose the next location to split, $m = \mathrm{next}(\mathcal{C}^*)$,
 Form two new constraint sets,

$$\mathcal{C}_{i+1}^{(0)} = \mathcal{C}^* \cup \{w_m = 0\}, \mathcal{C}_{i+1}^{(1)} = \mathcal{C}^* \cup \{w_m = 1\}.$$

 In the set of leaves $\mathcal{S}$, replace the parent constraint set $\mathcal{C}^*$ with the two children $\mathcal{C}_{i+1}^{(0)}$ and $\mathcal{C}_{i+1}^{(1)}$:

$$\mathcal{S} \leftarrow (\mathcal{S}\backslash\{\mathcal{C}^*\}) \cup \{\mathcal{C}_{i+1}^{(0)}\} \cup \{\mathcal{C}_{i+1}^{(1)}\}.$$

 Compute new lower and upper bounds for the two new constrained MICP:

$$LB(\mathcal{C}_{i+1}^{(0)}), LB(\mathcal{C}_{i+1}^{(1)}), UB(\mathcal{C}_{i+1}^{(0)}), UB(\mathcal{C}_{i+1}^{(1)}),$$

 Update the global lower and upper bounds,

$$L_{i+1} = \min_{\mathcal{C}\in\mathcal{S}}\{LB(\mathcal{C})\}, U_{i+1} = \min_{\mathcal{C}\in\mathcal{S}}\{UB(\mathcal{C})\}. \tag{42}$$

$i \leftarrow i + 1$.
Choose the best achieved solution so far:

$$\hat{\mathcal{C}} = \operatorname*{argmin}_{\mathcal{C}\in\mathrm{S}} UB(\mathcal{C}).$$

 Return the greedy solution under the constraint set $\hat{\mathcal{C}}$.

We note that Algorithm 1 is a degraded version of Algorithm 2 with just one iteration. As the total number of possible constraint sets is $2^{\bar{M}}$ (corresponding to the $2^{\bar{M}}$ location indicator matrices W), Algorithm 2 is guaranteed to converge in $2^{\bar{M}}$ iterations (and in practice many fewer as will be shown in the next section). To limit the algorithm's run time, a maximum number of iterations $i_{\max}$ can be enforced as in Algorithm 2.

VI. Simulation

In this section, we evaluate the developed optimal detection rule (cf. (12)), approximate performance metrics (cf. (29), (30) and (31)), and algorithms for optimizing sensor locations (cf. Algorithm 1 and 2), via simulations in the IEEE 14 bus system (cf. Fig. 1 [28]) using the software toolbox MATPOWER [29].

A. Set-Up and System Parameters

We employ a DC power flow model in the simulations. As explained in Section II-B, here we view power injections as states of the network. Note that power injections do not automatically change when the network topologies change as a result of outages, as long as the power balance is still satisfied, and before generation control or load shedding is performed. In contrast, voltage phase angles (i.e., the conventional "states") do instantly

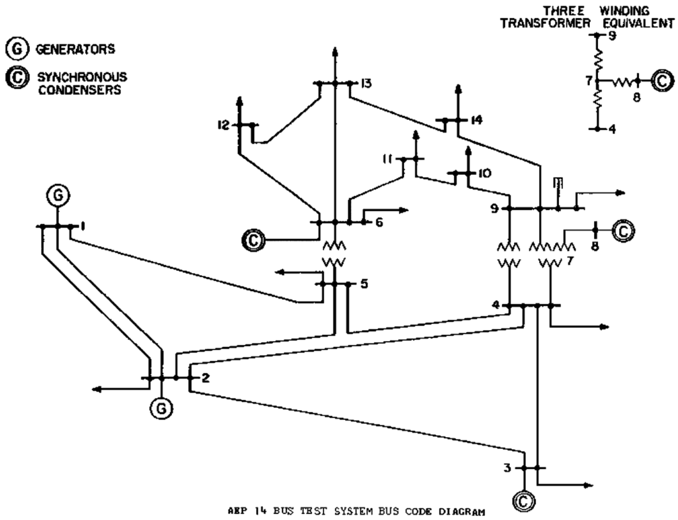


Fig. 1. IEEE 14 bus system.

change once there is a topological change. Thus, viewing power injections as states provides us the convenience that the prior distributions of power injections do *not* need to depend on the outage hypothesis, in contrast to the general case of (6). For the prior distributions, we employ the typical real power injections from the IEEE 14 bus system data as the prior means, and we employ a diagonal prior covariance matrix, implying independently (but not identically) distributed power injections. We denote by κ *the ratio between the standard deviation and the mean of a power injection*. The lower κ is, the more information we know about the states (i.e., the power injections). We assume all the power injections have the same κ. We will let $\kappa = 0.1$ for the most part in our simulations, and finally vary κ to evaluate its effect on optimal detection performance. We observe from the test data that buses 7 and 8 do not have power injections. This leads to the fact that bus 8 can be merged into bus 7, since bus 7 is the only bus that bus 8 connects to (cf. Fig. 1). Indeed, as there is no power flow on the line that connects buses 7 and 8, an outage on this line is intrinsically undetectable. Equivalently, we work with a network having 13 buses[3].

We simulate with an outage event set that includes all the 19 single line outages in the network, and apply a uniform probability distribution on these 19 outage events. We consider using only PMUs that measure voltage phase angles at the buses in real time. We assume PMU measurement noise to be independent and identically distributed (IID) Gaussian noise, with zero mean and standard deviation of 0.005 rad. This degree of accuracy conforms to the IEEE standard for PMUs [30]. We note that all our developed methodologies apply to any general set of outage events and any type of sensors as well.

B. Three Metrics for Optimizing Sensor Locations

We have proposed three performance metrics (cf. Section IV-C) that approximate the probability of error as guidance for optimizing sensor locations, To compare the three metrics, we use exhaustive search to find the *globally optimal* sensor locations that minimize each of the three metrics, with the number of sensors M sweeping from the minimum 2 (we consider that there is always a sensor at the reference bus 1) to the maximum 13 (buses 7 and 8 are merged). For each set of sensor locations, we evaluate the actual probability of error using the Monte Carlo method, and run 10^5 sets of realizations of the random outage events, power injections and noises. For each realization, the optimal outage detector with uncertain states (12) is applied.

Remark 2: The above simulation setting evaluates the case in which the number of PMUs is no greater than the number of network states. A natural question arises as to whether "network observability" [9] can be achieved based only on these sensor measurements. When there is no prior information at all about the network states, with the number of sensors less than the number of states to be inferred, network observability indeed cannot be achieved. In the Bayesian framework considered in this paper, however, prior information in principle eliminates the concern of network observability. As indicated at the beginning of Section III, such prior information can come from many different sources, including previously computed state estimates. For example, conventional SCADA can provide state estimates in relatively slower time scales that set the priors on the states, while PMUs can provide measurements in *much faster* time scales, enabling *real time identification of outages*. In these simulations, we are interested in the performance of real time outage identification based on *a few PMU measurements* as well as priors on the network states obtained via all other information sources in slower time scales. We note that the developed methodologies apply to general simulation settings (including the case of using redundant measurements as opposed to just a few PMUs).

In Fig. 2, the three metrics achieved (each with a corresponding optimal set of sensor locations) as a function of the number of sensors M are plotted as dash-dotted curves. The simulated probabilities of errors with the optimal sets of sensor locations under each metric are plotted in solid lines. We observe that, in terms of the actual probability of error, the optimal sensor locations obtained with the Sum-Sum and Sum-Max metric perform almost the same. The Max-Max metric, on the other hand, leads to sensor locations that are less optimal as the actual $\mathrm{P_e}$s are higher. Thus, the Sum-Sum and Sum-Max metric provide better guidance for optimizing sensor locations. From the curves, $\mathrm{P_e} \geq 2\%$ in all the simulated cases. We see that a moderate probability of error is likely to be a typical regime in which the detector operates.

For the metrics themselves, the Sum-Sum metric stays much above the actual probabilities of error, meaning that the union bound used in this metric is quite loose. The Sum-Max metric stays closer to the actual $\mathrm{P_e}$. The Max-Max metric is much below the actual $\mathrm{P_e}$, meaning that the error events other than the most dominant one are not really negligible. We observe that the three metrics themselves are not particularly accurate in terms of approximating the actual probability of error (e.g., the Sum-Sum metric is off by approximately a constant factor for all the simulated cases). Nonetheless, the primary purpose of deriving the metrics is to provide analytical bases for optimally selecting sensor locations. It will be shown in the next

[3]We note that this merge of buses 7 and 8 is specific for the test data that we used, and is not necessary for other test cases.

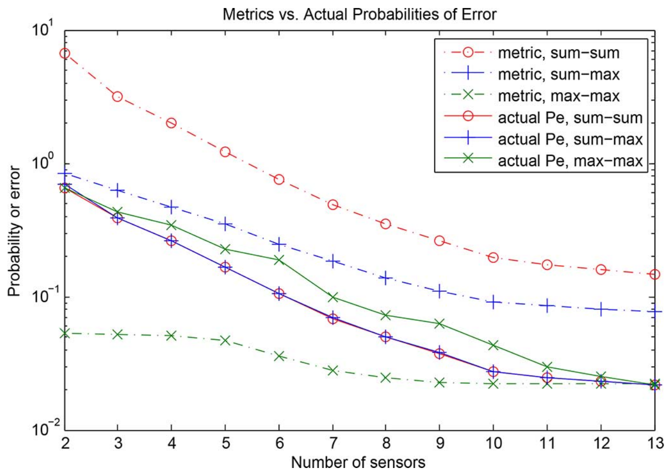


Fig. 2. Comparison of the proposed metrics, with each of which the sensor locations are optimized and the actual probabilities of error are evaluated via Monte Carlo simulations.

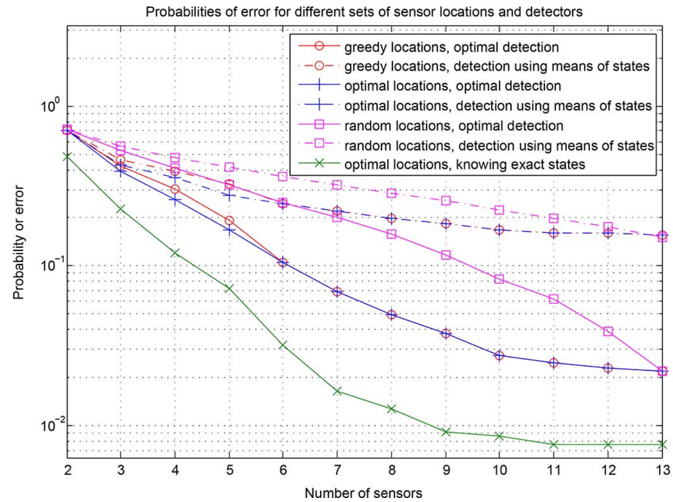


Fig. 3. Performance comparison between the optimal detector and simple detector, and between the optimal sensor locations, greedy locations and random locations.

subsection that the sensor locations optimized accordingly indeed provide great performance gain in outage identification. For the rest of this section, we will demonstrate simulations using the *Sum-Max* metric.

C. Impact of Detectors and Sensor Locations

The optimal detector (cf. (12) and (11)) takes full account of the uncertainty of states. In comparison, a simpler detector assumes that the knowledge of states is accurate, and takes prior conditional means as the actual states by substituting (26) into (10), (11) and (12). For sensor location optimization, we compare three different strategies: uniformly random locations, greedily selected locations (cf. Algorithm 1), and globally optimal sensor locations via exhaustive search. With each strategy, we sweep the number of sensors M from 2 to 13 and obtain sensor locations. For each greedily or optimally selected set of sensor locations, we evaluate the actual probability of error using the Monte Carlo method, and run 10^5 sets of realizations of the random outages, power injections and noises. For the random location strategy, we randomly draw a set of sensor locations *in each Monte Carlo run*. For each realization of all the random quantities, we apply the optimal detector and the above simple detector for comparison.

In Fig. 3, for the three strategies of sensor location selection, the $\mathrm{P_e}$ achieved with the optimal detector are plotted in solid lines, and that with the simple detector in dash-dotted curves. With all three strategies, we observe significant $\mathrm{P_e}$ reduction, exhibited by the optimal detector over the simple detector: The optimal detector can achieve as much as 7 times lower probability of error. These results demonstrate that the uncertainty of states cannot be simply neglected. We also observe that optimizing sensor locations makes a huge difference in terms of performance, as random sensor locations perform much worse than either greedily or optimally selected locations. For example, with 8 sensors, optimally located sensors can achieve as much as 3 times lower probability of error than randomly located ones. Interestingly, we observe that the performance gap between greedy and optimal sensor locations is quite small, and greedy solutions of sensor locations perform optimally for $M \geq 6$.

TABLE I
NUMBER OF ITERATIONS TO REACH THE GLOBAL OPTIMUM, IEEE 14 BUS

M	2	3	4	5	6	7	8	9	10
i_{achieve}	1	1	1	1	1	1	1	1	1
i_{prove}	11	22	23	13	10	8	9	8	10
M	11	12	13						
i_{achieve}	1	1	1						
i_{prove}	1	1	1						

Finally, we evaluate the case when there is indeed no uncertainty about the states, namely, we know the ground truth of the states exactly when doing outage identification. In this case, the simplified performance metric with (28) applies, and the globally optimal sensor locations can now be found via the proposed branch and bound algorithm (Algorithm 2). We again evaluate the probability of error with 10^5 Monte Carlo runs. In this case, the simple detector is indeed optimal, as the states are known exactly. The simulated curve shows that much lower $\mathrm{P_e}$ can be achieved in this case. A more detailed study of the effect of the accuracy of the prior knowledge is described later in Section VI-E.

D. Performance of the Branch and Bound Algorithm

To examine the effectiveness of the proposed branch and bound algorithm, we define i_{achieve} to be the number of iterations used by Algorithm 2 to *achieve* the globally optimal solution, and i_{prove} the number of iterations used to *prove* its global optimality. In other words, it takes i_{achieve} iterations for the *upper bound* to reach the global optimum, while it takes i_{prove} iterations for *both* the upper and lower bounds to reach the global optimum. We summarize in Table I the actual i_{achieve} and i_{prove} used to find the globally optimal sensor locations when we know the states accurately.

Interestingly, since $i_{\text{achieve}} = 1$ for all M, *the greedily selected sensor locations by Algorithm 1 are globally optimal* in this case of exact state knowledge. (We note that, the phenomenon that i_{prove} falls from 10 to 1 when M increases from 10 to 11 is a quantization effect due to a stopping rule of $\frac{U_i-L_i}{|U_i|} <$

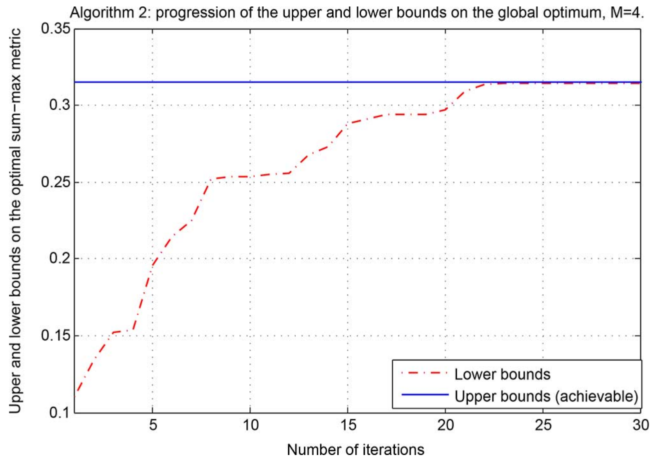


Fig. 4. The instant upper and lower bounds on the optimal achievable Sum-Max metric as Algorithm 2 iterates; $M = 4$.

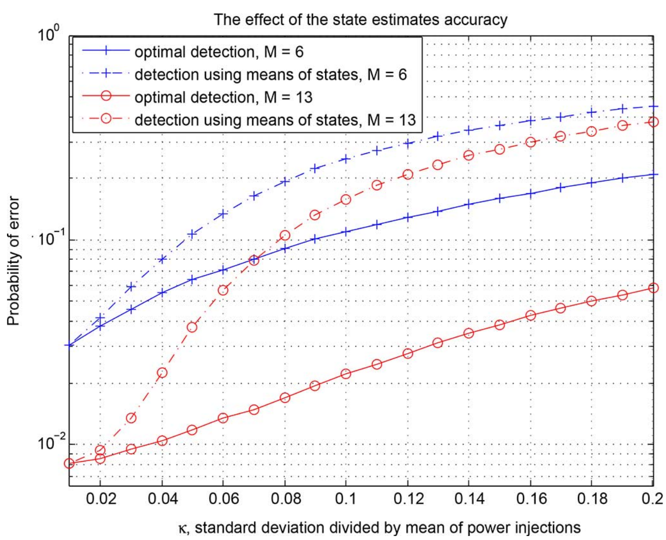


Fig. 5. Detection performance as the accuracy of the prior information on the states varies.

10^{-3}.) An illustration of the convergence of the lower bound is plotted in Fig. 4 for the case of $M = 4$, for which 23 iterations are needed.

E. Effect of the Accuracy of Prior

We now examine the effect of the accuracy of prior characterized by κ, i.e., the ratio between the prior standard deviation and the prior mean of power injections. We vary κ from 1% to 20%, and evaluate the probabilities of error with the globally optimal sensor locations obtained in Section VI-C, for the cases of $M = 6$ and $M = 13$. We evaluate both the optimal detector (12) and the simple detector which assumes perfect priors (26). Again, 10^5 Monte Carlo runs are used. The average performance is computed and plotted in Fig. 5. We see that with κ as low as 1%, the optimal detector and the simple detector perform almost the same. As κ increases, $\mathrm{P_e}$ increases steadily, and the performance gap between the two detectors increases significantly.

VII. Conclusion

We have studied the problem of outage identification in power systems with uncertain states, formulated as a joint detection and estimation problem. The joint posterior of the outage hypotheses and the network states have been derived in closed-form and have been used to develop the minimum probability of error detector. To capture how the observation matrices affect the optimal detection performance, we have developed three performance metrics by first deriving a Chernoff bound on the pairwise error probability, and then applying three heuristics to combine the pairwise error bounds to capture the actual probability of error. Based on the three metrics, a greedy strategy has been developed to optimize the sensor locations. Moreover, assuming perfect prior knowledge on the network states and a uniform prior on the outage events, a branch and bound algorithm based on relaxation of a mixed integer convex programming has been developed to find the globally optimal sensor locations.

The developed optimal detector and sensor location optimization algorithms have been tested via simulation in the IEEE 14 bus system. It has been shown that, under state uncertainty, the optimal detector significantly outperforms the simple detector that does not consider state uncertainty: The optimal detector can achieve as much as 7 times lower probability of error. Furthermore, optimizing the sensor locations has also been shown to be critical for improving the detection performance, as optimally located sensors can achieve as much as 3 times lower probability of error than randomly located ones. Interestingly, the proposed greedy algorithm for sensor location optimization has been shown to have near-optimal performance, suggesting that it is an effective algorithm with low complexity.

Appendix A
Proof of Lemma 2

Noting that the optimal detector has a form of

$$\frac{p(\mathcal{H}_j|\boldsymbol{z})}{p(\mathcal{H}_i|\boldsymbol{z})} \underset{\mathcal{H}_i}{\overset{\mathcal{H}_j}{\gtrless}} 1 \Leftrightarrow \ln\frac{p(\mathcal{H}_j|\boldsymbol{z})}{p(\mathcal{H}_i|\boldsymbol{z})} \underset{\mathcal{H}_i}{\overset{\mathcal{H}_j}{\gtrless}} 0, \tag{43}$$

we now evaluate the log-posterior-ratio on the right hand side. By (11), we have

$$\begin{aligned}
&\ln\frac{p(\mathcal{H}_j|\boldsymbol{z})}{p(\mathcal{H}_i|\boldsymbol{z})}\\
&\quad = \ln\frac{p(\mathcal{H}_j)}{p(\mathcal{H}_i)} + \frac{1}{2}\ln\frac{\det(\Sigma_i)}{\det(\Sigma_j)}\\
&\qquad + \frac{1}{2}\|z-\zeta_i\|^2_{\Sigma_i^{-1}} - \frac{1}{2}\|z-\zeta_j\|^2_{\Sigma_j^{-1}}\\
&\quad = \frac{1}{2}\left\{z^T\left(\Sigma_i^{-1}-\Sigma_j^{-1}\right)z + 2\left(\zeta_j^T\Sigma_j^{-1}-\zeta_i^T\Sigma_i^{-1}\right)z\right\}\\
&\qquad - \frac{1}{2}\left\{-2\ln\frac{p(\mathcal{H}_j)}{p(\mathcal{H}_i)} - \ln\frac{\det(\Sigma_i)}{\det(\Sigma_j)} + \|\zeta_j\|^2_{\Sigma_j^{-1}} - \|\zeta_i\|^2_{\Sigma_i^{-1}}\right\}\\
&\quad = \frac{1}{2}T(\boldsymbol{z}) - \frac{1}{2}\tau
\end{aligned} \tag{44}$$

where $T(\boldsymbol{z})$ and τ are defined in (15) and (16), respectively. Substituting (44) into (43), we complete our proof.

APPENDIX B
PROOF OF THEOREM 1

To evaluate $\mu_{T,i}(s) = \ln(\mathbb{E}(e^{sT(\boldsymbol{z})}|\mathcal{H}_i))$, we first evaluate the exponential moment $\mathbb{E}(e^{sT(\boldsymbol{z})}|\mathcal{H}_i)$:

$$\begin{aligned}
&\mathbb{E}\left(e^{sT(\boldsymbol{z})}|\mathcal{H}_i\right)\\
&= \int_{\mathbb{R}^M} \exp(sT(z)) \cdot \frac{1}{\sqrt{(2\pi)^M \cdot \det(\Sigma_i)}}\\
&\quad \times \exp\left(-\frac{1}{2}(z-\zeta_i)^T\Sigma_i^{-1}(z-\zeta_i)\right) dz\\
&= \int_{\mathbb{R}^M} \frac{1}{\sqrt{(2\pi)^M \cdot \det(\Sigma_i)}}\\
&\quad \times \exp\left(sT(z) - \frac{1}{2}(z-\zeta_i)^T\Sigma_i^{-1}(z-\zeta_i)\right) dz.
\end{aligned} \tag{45}$$

Substituting the expression (15) for $T(\boldsymbol{z})$ into the right-hand side of (45), we can express the exponent in (45) as

$$\begin{aligned}
sT(z) &- \frac{1}{2}(z-\zeta_i)^T\Sigma_i^{-1}(z-\zeta_i)\\
&= s\cdot z^T\left(\Sigma_i^{-1} - \Sigma_j^{-1}\right)z + 2s\cdot\left(\zeta_j^T\Sigma_j^{-1} - \zeta_i^T\Sigma_i^{-1}\right)z\\
&\quad - \frac{1}{2}(z-\zeta_i)^T\Sigma_i^{-1}(z-\zeta_i)\\
&= -\frac{1}{2}z^T\left(\Sigma_i^{-1} - 2s\cdot\left(\Sigma_i^{-1}-\Sigma_j^{-1}\right)\right)z\\
&\quad + \left(2s\cdot\left(\Sigma_j^{-1}\zeta_j - \Sigma_i^{-1}\zeta_i\right) + \Sigma_i^{-1}\zeta_i\right)^T z\\
&\quad - \frac{1}{2}\zeta_i^T\Sigma_i^{-1}\zeta_i.
\end{aligned} \tag{46}$$

Substituting (46) into (45), we observe that, as long as the matrix $\Sigma_i^{-1} - 2s\cdot(\Sigma_i^{-1} - \Sigma_j^{-1})$ is not positive definite, the integral in (45) will diverge to infinity, which means that $\mathbb{E}(e^{sT(\boldsymbol{z})}|\mathcal{H}_i)$ will go to infinity so that the exponential moment $\mu_{T,i}(s) \triangleq \ln\left(\mathbb{E}(e^{sT(\boldsymbol{z})}|\mathcal{H}_i)\right)$ also goes to infinity. Then, over this particular range of s, the Chernoff bound would become infinity and is thus not tight. Next, we focus on the case in which s satisfies that $\Sigma_i^{-1} - 2s\cdot(\Sigma_i^{-1} - \Sigma_j^{-1})$ is positive definite. Later, we will provide a condition on s that guarantees such positive definiteness. We introduce the following quantities

$$A(s) \triangleq \left(\Sigma_i^{-1} - 2s\cdot\left(\Sigma_i^{-1}-\Sigma_j^{-1}\right)\right)^{-1} \tag{47}$$

$$b(s) \triangleq 2s\cdot\left(\Sigma_j^{-1}\zeta_j - \Sigma_i^{-1}\zeta_i\right) + \Sigma_i^{-1}\zeta_i \tag{48}$$

$$\tilde{c} \triangleq -\frac{1}{2}\zeta_i^T\Sigma_i^{-1}\zeta_i. \tag{49}$$

With the above definitions, expression (46) can be written as

$$\begin{aligned}
sT(z) &- \frac{1}{2}(z-\zeta_i)^T\Sigma_i^{-1}(z-\zeta_i)\\
&= -\frac{1}{2}z^TA^{-1}(s)z + b^T(s)z + \tilde{c}\\
&= -\frac{1}{2}(z - A(s)b(s))^TA^{-1}(s)(z - A(s)b(s))\\
&\quad + \frac{1}{2}b^T(s)A(s)b(s) + \tilde{c}.
\end{aligned} \tag{50}$$

where in the last step we completed the square. Substituting (50) into (45), we obtain

$$\begin{aligned}
&\mathbb{E}\left(e^{sT(\boldsymbol{z})}|\mathcal{H}_i\right)\\
&= \frac{\sqrt{\det(A(s))}}{\sqrt{\det(\Sigma_i)}}\cdot\exp\left(\frac{1}{2}b^T(s)A(s)b(s)+\tilde{c}\right)\\
&\quad \cdot\int_{\mathbb{R}^M}\frac{1}{\sqrt{(2\pi)^M\cdot\det(A(s))}}\\
&\quad \times\exp\left(-\frac{1}{2}(z-A(s)b(s))^TA^{-1}(s)(z-A(s)b(s))\right)dz.
\end{aligned} \tag{51}$$

Since $A(s)$ is positive definite, the expression inside the integral of (51) is a Gaussian PDF with mean $A(s)b(s)$ and covariance $A(s)$, which means that the integral of it over the entire domain is one. Therefore, we obtain

$$\mathbb{E}\left(e^{sT(\boldsymbol{z})}|\mathcal{H}_i\right) = \begin{cases}\frac{\sqrt{\det(A(s))}}{\sqrt{\det(\Sigma_i)}}\exp\left(\frac{1}{2}b^T(s)A(s)b(s)+\tilde{c}\right), & A(s)>0\\ \infty, & \text{otherwise.}\end{cases}$$

Finally, we derive a condition on s that guarantees

$$\Sigma_i^{-1} - 2s\cdot\left(\Sigma_i^{-1}-\Sigma_j^{-1}\right) > 0, \tag{52}$$

which is equivalent to requiring that

$$x^T\left[\Sigma_i^{-1} - 2s\cdot\left(\Sigma_i^{-1}-\Sigma_j^{-1}\right)\right]x > 0, \forall x \neq 0. \tag{53}$$

Solving the above inequality for $1/s, s>0$, we obtain

$$\frac{1}{s} > \frac{2x^T\left(\Sigma_i^{-1}-\Sigma_j^{-1}\right)x}{x^T\Sigma_i^{-1}x}, \quad \forall x\neq 0. \tag{54}$$

Since the above inequality (54) is required to be hold for all $x \neq 0$, it is equivalent to requiring

$$\frac{1}{s} > \max_{x\neq 0}\frac{2x^T\left(\Sigma_i^{-1}-\Sigma_j^{-1}\right)x}{x^T\Sigma_i^{-1}x}. \tag{55}$$

Note that the ratio on the right-hand side of (55) is a generalized Rayleigh quotient. The maximum of a generalized Rayleigh quotient y^TAy/y^TBy for positive definite matrix B and symmetric matrix A is equal to the largest generalized eigenvalue of the corresponding matrix pair:

$$\max_{y\neq 0}\frac{y^TAy}{y^TBy} = \lambda_{\max}(A,B) \tag{56}$$

where $\lambda_{\max}(A,B)$ denotes the largest generalized eigenvalue of the matrix pair (A,B), i.e., there exists $y \neq 0$ such that

$$Ay = \lambda_{\max}(A,B)\cdot By. \tag{57}$$

For this reason, we have

$$\max_{x\neq 0}\frac{2x^T\left(\Sigma_i^{-1}-\Sigma_j^{-1}\right)x}{x^T\Sigma_i^{-1}x} = \lambda_{\max}\left(2(\Sigma_i^{-1}-\Sigma_j^{-1}),\Sigma_i^{-1}\right),$$

so that inequality (55) is equivalent to

$$\frac{1}{s} > \lambda_{\max}\left(2\left(\Sigma_i^{-1} - \Sigma_j^{-1}\right), \Sigma_i^{-1}\right) \triangleq \lambda_0. \tag{58}$$

If λ_0 on the right-hand side of (58) is negative or zero, then (58) holds automatically for $s > 0$ so that the condition for $A(s) > 0$ is

$$0 < s < +\infty. \tag{59}$$

Otherwise, if $\lambda_0 > 0$, then the condition for $A(s)$ becomes

$$0 < s < \frac{1}{\lambda_0}. \tag{60}$$

Summarizing all the above cases, we can conclude our proof of Theorem 1.

Appendix C
Proof of Corollary 1

With a perfect prior, i.e., $C_i = C_j = \mathbf{0}$, we have $\Sigma_i = \Sigma_j = R_v$ according to (10), and

$$\tau = -2\ln\frac{p(\mathcal{H}_j)}{p(\mathcal{H}_i)} + \|\zeta_j\|^2_{R_v^{-1}} - \|\zeta_i\|^2_{R_v^{-1}} \tag{61}$$

$$\bar{s} = +\infty \tag{62}$$

$$A(s) = R_v \tag{63}$$

$$b(s) = R_v^{-1}\left(2s\cdot(\zeta_j - \zeta_i) + \zeta_i\right) \tag{64}$$

$$c = -\frac{1}{2}\left(\|\zeta_i\|^2_{R_v^{-1}} + \ln\det(R_v)\right), \tag{65}$$

according to (20)–(23). Substituting the above expressions into (19), we obtain

$$\begin{aligned} -s\tau &+ \mu_{T,i}(s) \\ &= -s\tau + \frac{1}{2}\|2s\cdot(\zeta_j - \zeta_i) + \zeta_i\|^2_{R_v^{-1}} - \frac{1}{2}\|\zeta_i\|^2_{R_v^{-1}} \\ &= 2s^2\cdot\|\zeta_j - \zeta_i\|^2_{R_v^{-1}} - s\cdot\left(\|\zeta_j - \zeta_i\|^2_{R_v^{-1}} - 2\ln\frac{p(\mathcal{H}_j)}{p(\mathcal{H}_i)}\right). \end{aligned} \tag{66}$$

Note that the above expression is a quadratic function of s and is convex. Therefore, optimizing the above expression with respect to s over $[0, +\infty)$, we obtain the optimal s as

$$s^*_{ij} = \left(\frac{\|\zeta_j - \zeta_i\|^2_{R_v^{-1}} - 2\ln\frac{p(\mathcal{H}_j)}{p(\mathcal{H}_i)}}{4\|\zeta_j - \zeta_i\|^2_{R_v^{-1}}}\right)_+.$$

Substituting the above s^*_{ij} into (66), we obtain the minimum value for (66) as

$$-s^*_{ij}\tau + \mu_{T,i}\left(s^*_{ij}\right) = -\frac{\left(\|\zeta_j - \zeta_i\|^2_{R_v^{-1}} - 2\ln\frac{p(\mathcal{H}_j)}{p(\mathcal{H}_i)}\right)_+^2}{8\|\zeta_j - \zeta_i\|^2_{R_v^{-1}}}. \tag{67}$$

Substituting the above expression into the right-hand side of (18), we conclude our proof.

References

[1] Y. Zhao, A. Goldsmith, and H. V. Poor, "On PMU location selection for line outage detection in wide-area transmission networks," in *Proc. IEEE Power Energy Soc. General Meeting*, Jul. 2012, pp. 1–8.

[2] J. Chen, Y. Zhao, A. Goldsmith, and H. V. Poor, "Optimal joint detection and estimation in linear models," in *Proc. IEEE Conf. Decision and Control*, Dec. 2013, pp. 4416–4421.

[3] "Final Report on August 14, 2003 the Blackout in the United States and Canada," US-Canada Power System Outage Task Force, 2004.

[4] A. G. Phadke, "Synchronized phasor measurements in power systems," *IEEE Comput. Appl. Power*, vol. 6, no. 2, pp. 10–15, Apr. 1993.

[5] K. Sun, S. Likhate, V. Vittal, V. S. Kolluri, and S. Mandal, "An online dynamic security assessment scheme using phasor measurements and decision trees," *IEEE Trans. Power Syst.*, vol. 22, no. 4, pp. 1935–1943, Nov. 2007.

[6] J. E. Tate and T. J. Overbye, "Line outage detection using phasor angle measurements," *IEEE Trans. Power Syst.*, vol. 23, no. 4, pp. 1644–1652, Nov. 2008.

[7] J. E. Tate and T. J. Overbye, "Double line outage detection using phasor angle measurements," in *Proc. IEEE PES General Meeting*, Jul. 2009, pp. 1–5.

[8] H. Zhu and G. B. Giannakis, "Sparse overcomplete representations for efficient identification of power line outages," *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 2215–2224, Nov. 2012.

[9] A. Abur and A. Gomez-Exposito, *Power System State Estimation: Theory and Implementation*. New York, NY, USA: Marcel Dekker, 2004.

[10] Y. F. Huang, S. Werner, J. Huang, N. Kashyap, and V. Gupta, "State estimation in electric power grids: Meeting new challenges presented by the requirements of the future grid," *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 33–43, Sep. 2012.

[11] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on smart grid state estimation: Attack strategies and countermeasures," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm.)*, Oct. 2010, pp. 222–225.

[12] F. C. Schweppe and J. Wildes, "Power system static-state estimation, Part I: Exact model," *IEEE Trans. Power Apparatus Syst.*, vol. 89, no. 1, pp. 120–125, Jan. 1970.

[13] F. F. Wu and W. H. E. Liu, "Detection of topology errors by State Estimation," *IEEE Trans. Power Sys.*, vol. 4, pp. 176–183, Aug. 1989.

[14] K. A. Clements and P. W. Davis, "Detection and identification of topology errors in electric power systems," *IEEE Trans. Power Syst.*, vol. 3, no. 4, pp. 1748–1753, Nov. 1988.

[15] V. Kekatos and G. B. Giannakis, "Joint power system state estimation and breaker status identification," in *Proc. North Amer. Power Symp. (NAPS)*, Sep. 2012, pp. 1–6.

[16] T. L. Baldwin, L. Mili, Jr., M. B. Boisen, and R. Adapa, "Power system observability with minimal phasor measurement placement," *IEEE Trans. Power Syst.*, vol. 8, no. 2, pp. 707–715, May 1993.

[17] L. Zhao and A. Abur, "Multi area state estimation using synchronized phasor measurements," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 611–617, May 2005.

[18] Q. Li, R. Negi, and M. D. Ilic, "Phasor measurement units placement for power system state estimation: A greedy approach," in *Proc. IEEE Power Energy Soc. General Meeting*, Jul. 2011, pp. 1–8.

[19] X. Li, A. Scaglione, and T. H. Chang, "A framework for phasor measurement placement in hybrid state estimation via Gauss-Newton," *eprint arXiv:1309.4911*.

[20] A. Giani, E. Bitar, M. Garcia, M. McQueen, P. Khargonekar, and K. Poolla, "Smart grid data integrity attacks: Characterizations and countermeasures," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm.)*, Oct. 2011, pp. 232–237.

[21] T. T. Kim and H. V. Poor, "Strategic protection against data injection attacks on power grids," *IEEE Trans. Smart Grid*, vol. 2, no. 2, pp. 326–333, Jun. 2011.

[22] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York, NY, USA: Springer-Verlag, 1994.

[23] A. J. Wood and B. F. Wollenberg, *Power Generation, Operation, and Control*, 2nd ed. New York, NY, USA: Wiley, 1996.

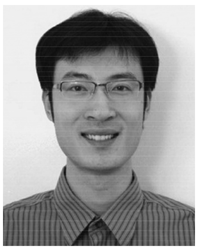
[24] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Upper Saddle River, NJ, USA: Prentice-Hall, 2000.

[25] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*. New York, NY, USA: Wiley, 2004.

[26] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[27] V. Balakrishnan, S. Boyd, and S. Balemi, "Branch and bound algorithm for computing the minimum stability degree of parameter-dependent linear systems," *Int. J. Robust Nonlinear Control*, vol. 1, no. 4, pp. 295–317, Oct.–Dec. 1991.

[28] "Power systems test case archive," Univ. of Washington Elect. Eng..
[29] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, "MATPOWER steady-state operations, planning and analysis tools for power systems research and education," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, Feb. 2011.
[30] *IEEE Standard for Synchrophasors for Power Systems*, IEEE Standard C37.118-2005, 2006.



Yue Zhao (S'06–M'11) received the B.E. degree in electronic engineering from Tsinghua University, Beijing, China, in 2006, and the M.S. and Ph.D. degrees in electrical engineering, both from the University of California, Los Angeles (UCLA), Los Angeles, in 2007 and 2011, respectively.

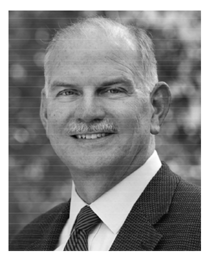
He is currently a postdoctoral scholar with the Department of Electrical Engineering, Stanford University, Stanford, CA, and with the Department of Electrical Engineering, Princeton University, Princeton, NJ. His research interests include statistical signal processing, optimization and game theory, information theory, and smart grid.

Jianshu Chen (S'10) received his B.S. and M.S. degrees in electronic engineering from Harbin Institute of Technology in 2005 and Tsinghua University in 2009, respectively. He received his Ph.D. degree in electrical engineering from University of California, Los Angeles (UCLA) in 2014.

Between 2009 and 2014, he was a member of the Adaptive Systems Laboratory (ASL) at UCLA. His research interests include statistical signal processing, stochastic approximation, parallel and distributed optimization, large-scale machine learning, and smart grid.

Andrea Goldsmith (S'90–M'93–SM'99–F'05) is the Stephen Harris professor in the School of Engineering and a professor of Electrical Engineering at Stanford University. She was previously on the faculty of Electrical Engineering at Caltech. Her research interests are in wireless information and communication theory, smart grid design and optimization, and application of signal processing, control and communications to biology and neuroscience. Dr. Goldsmith has co-founded and served as CTO for two wireless companies: Accelera, Inc. and Quantenna Communications, Inc. and she has also held industry positions at Maxim Technologies, Memorylink Corporation, and AT&T Bell Laboratories. She is a Fellow of the IEEE and of Stanford, and has received several awards for her work, including the IEEE Communications Society and Information Theory Society joint paper award, the IEEE Communications Society Best Tutorial Paper Award, the National Academy of Engineering Gilbreth Lecture Award, the IEEE ComSoc Communications Theory Technical Achievement Award, the IEEE ComSoc Wireless Communications Technical Achievement Award, the Alfred P. Sloan Fellowship, and the Silicon Valley/San Jose Business Journal's Women of Influence Award. She is author of the book "Wireless Communications" and co-author of the books "MIMO Wireless Communications" and "Principles of Cognitive Radio," all published by Cambridge University Press. She received the B.S., M.S. and Ph.D. degrees in electrical engineering from U.C. Berkeley.

Dr. Goldsmith has served as editor for the IEEE TRANSACTIONS ON INFORMATION THEORY, the *Journal on Foundations and Trends in Communications and Information Theory and in Networks*, the IEEE TRANSACTIONS ON COMMUNICATIONS, and the IEEE WIRELESS COMMUNICATIONS MAGAZINE as well as on the Steering Committee for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. She participates actively in committees and conference organization for the IEEE Information Theory and Communications Societies and has served on the Board of Governors for both societies. She has also been a Distinguished Lecturer for both societies, served as President of the IEEE Information Theory Society in 2009, founded and chaired the student committee of the IEEE Information Theory Society, and chaired the Emerging Technology Committee of the IEEE Communications Society. At Stanford she received the inaugural University Postdoc Mentoring Award and has been active in committees to innovate and revise both graduate and undergraduate education university-wide. She served as Chair of Stanford's Faculty Senate in 2009 and currently serves on its Faculty Senate and on its Budget Group.



H. Vincent Poor (S'72–M'77–SM'82–F'87) received the Ph.D. degree in EECS from Princeton University in 1977. From 1977 until 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990 he has been on the faculty at Princeton, where he is the Michael Henry Strater University Professor of Electrical Engineering and Dean of the School of Engineering and Applied Science. Dr. Poor's research interests are in the areas of stochastic analysis, statistical signal processing, and information theory, and their applications in wireless networks and related fields such as social networks and smart grid. Among his publications in these areas are the recent books Principles of Cognitive Radio (Cambridge, 2013) and Mechanisms and Games for Dynamic Spectrum Allocation (Cambridge University Press, 2014).

Dr. Poor is a member of the National Academy of Engineering and the National Academy of Sciences, and a foreign member of Academia Europaea and the Royal Society. He is also a fellow of the American Academy of Arts and Sciences, the Royal Academy of Engineering (U.K), and the Royal Society of Edinburgh. He received the Technical Achievement and Society Awards of the IEEE Signal Processing Society in 2007 and 2011, respectively. Recent recognition of his work includes the 2014 URSI Booker Gold Medal, and honorary doctorates from Aalborg University, the Hong Kong University of Science and Technology, and the University of Edinburgh.