Boundary Defense Against Cyber Threat for Power System State Estimation

Ming Jin, Javad Lavaei, Somayeh Sojoudi, and Ross Baldick

***Abstract*—The operation of power grids is becoming increasingly data-centric. While the abundance of data could improve system efficiency, it poses major reliability challenges. In particular, state estimation aims to find the operating state of a network from the telemetered data, but an undetected attack on the data could lead to making wrong operational decisions for the system and trigger a large-scale blackout. Nevertheless, understanding the vulnerability of state estimation with regards to cyberattacks, which is a special instance of graph-structured quadratic sensing problem, has been hindered by the lack of tools for studying the topological and data-analytic aspects of networks. Algorithmic robustness is critical in extracting reliable information from abundant but untrusted grid data. For a large-scale power grid, we quantify, analyze, and visualize the regions of the network that are not robust to cyberattacks in the sense that there exists a data manipulation strategy for each of those local regions that misleads the operator at the global scale and yields a wrong estimation of the state of the network at almost all buses. We also propose an optimization-based graphical boundary defense mechanism to identify the border of the geographical area in which data have been manipulated. The proposed method does not allow a local attack to have a global effect on the data analysis of the entire network, which enhances the situational awareness of the grid, especially in the face of adversity. The developed mathematical framework reveals key geometric and algebraic factors that can affect algorithmic robustness and is used to study the vulnerability of the U.S. power grid in this paper.**

***Index Terms*—Power system state estimation, CPS security, robust algorithm, smart grid**

I. Introduction

While real-world data abound for many complex systems, they are often noisy and corrupted. Acquiring reliable information from abundant but untrusted data is key to enhancing cybersecurity for mission-critical systems such as the power grid [1]. Since many of these systems are inherently network structured, data analytics cannot be satisfactorily understood without incorporating their underlying graph topologies.

For instance, consider the power system state estimation (SE) problem, which constantly monitors the operating status of the grid by filtering and fusing a large volume of data every few minutes [2]. The significance of a functioning SE could be inferred from the 2003 large-scale blackout, in which the failure of SE contributed to the inability of the operator to provide real-time diagnostic support [3]. Despite substantial advances in algorithm design [2], [4]–[17], a major obstacle still remains: the lack of a framework for the design of a robust and scalable algorithm together with a realistic evaluation of its vulnerability. Developing such a framework is challenging for three reasons: (a) the model of a power system is highly nonlinear and nonconvex due to physical laws, (b) computational resources required by existing algorithms grow rapidly with the size of the system, and (c) the number of scenarios involving adversarial conditions is too large for an individual assessment of each scenario to be possible (it is higher than the number of atoms in the observable universe for systems with as low as 500 possible attack points). These challenges have limited the scope of previous studies to simple approximate models or conservative methods that ignore the topology-dependent characterization of vulnerabilities [2], [4]–[21]. Fundamentally, there is a lack of tools to deal with untrusted data associated with nonlinear and structured (rather than random) graphical models.

A. Graph-structured quadratic sensing

The graph-structured quadratic sensing problem includes SE as a special instance and is stated as follows. Let $\boldsymbol{v} \in \mathbb{C}^{n_b}$ be an unknown n_b-dimensional complex-valued state vector. The goal is to find $\boldsymbol{v}$ from a set of noisy quadratic measurements

$$y_i = \boldsymbol{v}^* \boldsymbol{M}_i \boldsymbol{v} + \omega_i + b_i, \qquad \forall i \in [n_m], \tag{1}$$

where $\boldsymbol{v}^*$ indicates the complex conjugate, $\boldsymbol{M}_i$ is a known $n_b \times n_b$ dimensional Hermitian matrix, ω_i denotes a zero-mean Gaussian random noise with standard deviation σ, and b_i denotes bad data that can take arbitrary values. Here, we use the shorthand notation $[n] = \{1, ..., n\}$.

Based on the set of measurement matrices $\{\boldsymbol{M}_i\}_{i\in[n_m]}$, we construct an undirected graph $\mathcal{G} \coloneqq \{\mathcal{N}, \mathcal{L}\}$, where $\mathcal{N} \coloneqq [n_b]$ and $\mathcal{L} \coloneqq [n_l]$ represent the sets of nodes and edges, respectively. The graph is constructed such that there is an edge $\ell \coloneqq \{f, t\}$ that connects nodes f and t if there exists a sensing matrix $\boldsymbol{M}_i$ whose $(f,t)^{\text{th}}$ entry is nonzero, i.e., $[\boldsymbol{M}_i]_{f,t} \neq 0$. We are interested in the case where the measurement matrices are sparse and produce a sparse computational graph $\mathcal{G}$.

For SE, $\boldsymbol{v}$ consists of the voltages at all nodes of the network and the measurements, such as voltage magnitudes and real and reactive power flows over edges, are quadratic [17], [22]. The bad data b_i is either zero corresponding to a correct measurement or nonzero corresponding to cyberattack, communication failure, sensor fault, or deployment of a model (i.e., measurement matrices $\boldsymbol{M}_i$) that does not match the reality (e.g., a disconnected line is wrongly assumed to be in service by the operator). It is not known *a priori* which b_i's are nonzero. The graph $\mathcal{G}$ of SE coincides with or is a subset of the physical topology of the gird, and therefore it is sparse.

†This work was supported by the ONR grant N00014-17-1-2933 and NSF Award 1807260, ARO grant W911NF-17-1-0555, and AFOSR grant FA9550-17-1-0163. M. Jin, J. Lavaei, and S. Sojoudi are with the University of California, Berkeley, CA 94710; R. Baldick is with the University of Texas at Austin, TX 78712, USA.

B. Related work

Different types of quadratic sensing problems have been studied in the literature, which can be cateogrized based on the assumptions made on the measurement matrices $\boldsymbol{M}_i$'s: (i) matrix completion [23], [24] and robust principal component analysis [25], [26] assume that each matrix $\boldsymbol{M}_i$ has a single nonzero element at a random location; and (ii) phase retrieval [27] assumes that each matrix $\boldsymbol{M}_i$ is rank-1. Existing approaches to solve these problems include convex relaxation [23], [25], [26] and iterative algorithms [24]. To obtain guarantees of performance, a common theoretical condition is called restricted isometry property [24]; however, this condition only applies to dense and/or random matrices $\boldsymbol{M}_i$'s, while the measurement matrices in our study are deterministic and structured (due to the existence of an inherent graph structure).

Due to the prominence of SE, extensive works have been conducted in the power system community. These methods include: (i) linearization (a.k.a., DC approximation) [5], [28]; (ii) iterative algorithms such as Newton's method [2], [29], feasible point pursuit [30], and iterative convex program [8]; and (iii) global optimization techniques such as particle swarm optimization [31], and semidefinite relaxation [11], [13], [14]. However, for methods in (i), the approximation error could be arbitrarily large when the unknown voltage vector deviates from the nominal state around which the linearization is performed. For methods in (ii), due to the nonconvexity of solving quadratic measurement equations, the algorithms can become trapped at meaningless local minima or saddle points, which do not provide a useful estimate of the state. For methods in (iii), the primary disadvantage is their heavy computational requirement or lack of theoretical guarantees on their ability to reject bad data. Existing literature on cyberattack and defense, such as the false data injection attack, has also been limited to DC approximation models [4], [6], [12] , with the exception of a few works on the nonlinear AC model [10], [32], [33]. However, it has been found that the mismatch caused by the DC approximation of the AC grid renders either the defense or the attack efforts futile [7], [9], [10]. Recently, a two-stage linear/quadratic programming approach has been proposed in [17], which advances the state of the art by providing a computationally efficient algorithm with theoretical guarantees of recovering the true state. However, the proposed condition is hard to be satisfied and its verification requires the knowledge of the support of the bad data, which is not known *a priori*.

C. Gap in the literature and our contributions

One common drawback of all the existing methods is that the theoretical certificates used to reject bad data are provided on a scenario-by-scenario basis, where each scenario corresponds to one specific set of measurements that are corrupted by bad data. Since there are an exponential number of ways to attack the grid data (namely, 2^m ways to decide on the zero-nonzero pattern of b_i's in the case with m measurements), it is impossible to make a meaningful general assessment of the vulnerability of a grid based on a single scenario.

Another important missing factor is that the prior literature aims to find the state of the system correctly under attacks, while this is theoretically impossible when the data for a sub-network of the system is strategically manipulated. In this case, the state for that region becomes unobservable (not recoverable) from the clean data for the rest of the system. To elaborate, let $\boldsymbol{v}$ and $\hat{\boldsymbol{v}}$ be the true and the estimated states, respectively. Let R denote the subnetwork under a cyberattack, and $\boldsymbol{v}_R$ and $\boldsymbol{v}_{\setminus R}$ be the voltages for the attacked region and the remainder of the system, respectively. The existing works aim to find $\hat{\boldsymbol{v}}$ such that the global metric $\|\boldsymbol{v}-\hat{\boldsymbol{v}}\|$ is minimized, i.e., global recovery, which is not possible since $\|\boldsymbol{v}_R - \hat{\boldsymbol{v}}_R\|$ can be arbitrarily large; therefore, it is more realistic to focus on $\|\boldsymbol{v}_{\setminus R} - \hat{\boldsymbol{v}}_{\setminus R}\|$, i.e., local recovery.

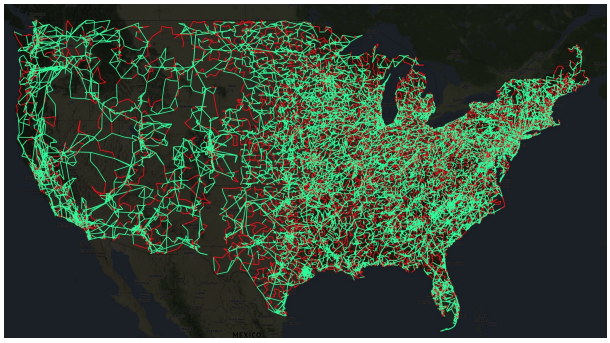

Fig. 1: Vulnerability map of the modified U.S. power grid, which consists of three AC interconnections, West, East, and ERCOT, and shows the lines that satisfy the graphical mutual incoherence condition proposed in the study (green lines) those that do not (red lines).

This paper is the first work to develop a mathematical framework for local recovery. On the application impact, our method provides *the first vulnerability map* for the entire U.S. grid, as shown in Fig. 1. Based on the graphical mutual incoherence condition to be discussed next, we can categorize each edge as either robust or vulnerable. On this map, if the connections between the region R and the rest of the grid are all robust edges, then no matter how the measurements inside the region are modified, the estimation error is only limited to this region and cannot propagate out of the boundary formed by the robust lines to affect the rest of the grid in terms of $\|\boldsymbol{v}_{\setminus R}-\hat{\boldsymbol{v}}_{\setminus R}\|$. If even one edge in the surrounding subnetwork is vulnerable, then it is possible for the estimation error to propagate to the rest of the grid. Importantly, this vulnerability map is obtained without knowing the attack locations, and therefore it provides a universal measure that applies to an exponential number of possible attack scenarios.

The rest of the paper is organized as follows. A two-stage algorithm for quadratic sensing is introduced in Sec. II. Sec. III discusses the graphical mutual incoherence conditions. The boundary defense mechanism is introduced in Sec. IV. Sec. V develops an important application in SE to map the geographic vulnerabilities and how the network and optimization properties can influence vulnerability. Concluding remarks are given in Sec. VI. All the proofs are provided in the Appendix.

Notations: We use $\mathbb{R}$ and $\mathbb{C}$ as the sets of real and complex numbers. The cardinality $|\mathcal{J}|$ of a set $\mathcal{J}$ is the number of elements in a set. The support $\mathrm{supp}(\boldsymbol{x})$ of a vector $\boldsymbol{x}$ is the

set of indices of the nonzero entries of $\boldsymbol{x}$. For a set $\mathcal{J} \subset [m]$, we use $\mathcal{J}^c = [m] \setminus \mathcal{J}$ to denote its complement. The symbols $(\cdot)^\top$ and $(\cdot)^*$ represent the transpose and conjugate transpose operators. We use $\Re(\cdot)$, $\Im(\cdot)$ and $\mathrm{Tr}\,(\cdot)$ to denote the real part, imaginary part and trace of a scalar/matrix. The imaginary unit is denoted as $\sqrt{-1}$. The notations $\angle x$ and $|x|$ indicate the angle and magnitude of a complex scalar. We use $\lambda_{\min}(\boldsymbol{A})$ to denote the smallest eigenvalue of $\boldsymbol{A}$, and $\boldsymbol{A} \succeq 0$ to indicate that $\boldsymbol{A}$ is a positive semidefinite matrix. We denote $\boldsymbol{x}_{\mathcal{R}}$ as the subvector with entries of $\boldsymbol{x}$ indexed by $\mathcal{R}$, $\boldsymbol{A}_{\mathcal{R}}$ as the submatrix with $\mathcal{R}$ rows of $\boldsymbol{A}$, and $\boldsymbol{A}_{\mathcal{R},\mathcal{J}}$ as the submatrix with $\mathcal{R}$ rows and $\mathcal{J}$ columns of $\boldsymbol{A}$. We use $\|\cdot\|_\infty$ to denote the matrix infinity norm and $\|\cdot\|_F$ to denote the Frobenius norm.

II. Two-step pipeline of robust quadratic sensing

This section describes a two-stage robust quadratic sensing algorithm, where the first stage involves a conic optimization and the second stage can be computed with a closed-form equation. We will analyze this algorithm in Sec. III, which will be shown to be more robust to bad data than the algorithm in [17]. Since $\boldsymbol{v}$, $\boldsymbol{b}$ and $\boldsymbol{w}$ are unknown, henceforth we use these notations to show the corresponding variables and use the notations $\boldsymbol{v}_\natural$, $\boldsymbol{b}_\natural$ and $\boldsymbol{w}_\natural$ to denote their true values.

A. Stage 1: Estimation in the lifted space

In the first stage, we estimate a set of variables in a lifted space that are linked through the underlying state $\boldsymbol{v}$. Specifically, for a given computational graph $\mathcal{G} \coloneqq \{\mathcal{N}, \mathcal{L}\}$ based on the measurements $\{\boldsymbol{M}_i\}_{i\in[n_m]}$, we introduce two groups of variables: (i) nodal variables, $x_k^{\mathrm{mg}} \coloneqq |v_k|^2$, for each node $k \in \mathcal{N}$, and (ii) edge variables, denoted as $x_\ell^{\mathrm{re}} \coloneqq \Re(v_f v_t^*)$ and $x_\ell^{\mathrm{im}} \coloneqq \Im(v_f v_t^*)$ for each edge $\ell \in \mathcal{L}$ with the endpoints f and t (note that in this case we do not create separate edge variables corresponding to the (t,f)-th entry). Let $\boldsymbol{x}(\boldsymbol{v}) \in \mathcal{X} \subseteq \mathbb{R}^{n_x}$ be the collection of the lifted variables in a vector form (we often omit the dependence on $\boldsymbol{v}$ and just write $\boldsymbol{x}$), where $\mathcal{X}$ is the corresponding lifted space. By the construction of the lifted variables, the measurement model (1) can be written as:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}(\boldsymbol{v}_\natural) + \mathbf{w}_\natural + \boldsymbol{b}_\natural, \tag{2}$$

where $\boldsymbol{A} \in \mathbb{R}^{n_m \times n_x}$ is the sensing matrix, $\boldsymbol{v}_\natural \in \mathbb{C}^{n_v}$ and $\boldsymbol{x}_\natural \coloneqq \boldsymbol{x}(\boldsymbol{v}_\natural) \in \mathcal{X}$ are the true voltage state and the corresponding state in the lifted space, $\boldsymbol{y} \in \mathbb{R}^{n_m}$ is the set of measurements, $\mathbf{w}_\natural \in \mathbb{R}^{n_m}$ denotes random noise, and $\mathbf{b}_\natural \in \mathbb{R}^m$ is the bad data. Note that the i^{th} row of $\boldsymbol{A}$, denoted by $\boldsymbol{a}_i$, is constructed by taking the entries in $\boldsymbol{M}_i$ corresponding to the lifted variables such that $\boldsymbol{a}_i^*\boldsymbol{x}(\boldsymbol{v}) = \boldsymbol{v}^*\boldsymbol{M}_i\boldsymbol{v}$ for all $\boldsymbol{v}$.

First, we solve a convex optimization to minimize the Huber loss subject to second-order cone constraints (SOCs):

$$\min_{\boldsymbol{x}\in\mathcal{K}} \sum_{i=1}^{n_m} f_{\mathrm{Huber}}(y_i - [\boldsymbol{A}\boldsymbol{x}]_i; \lambda), \tag{3}$$

where $f_{\mathrm{Huber}}(r;\lambda) = \begin{cases} \frac{1}{2}r^2 & |r| \le \lambda \\ \lambda(|r| - \frac{1}{2}\lambda) & |r| > \lambda \end{cases}$ is the standard Huber loss parametrized by λ, and the feasible set $\mathcal{K}$ is

$$\left\{ \boldsymbol{x} \;\middle|\; H_\ell(\boldsymbol{x}) \succeq 0, \quad \forall \ell \coloneqq \{i,j\} \in \mathcal{L} \right\}, \tag{4}$$

where

$$H_\ell(\boldsymbol{x}) = \begin{bmatrix} x_i^{\mathrm{mg}} & x_\ell^{\mathrm{re}} + \sqrt{-1}x_\ell^{\mathrm{im}} \\ x_\ell^{\mathrm{re}} - \sqrt{-1}x_\ell^{\mathrm{im}} & x_j^{\mathrm{mg}} \end{bmatrix} \tag{5}$$

is a 2×2 matrix constructed for line $\ell \coloneqq \{i,j\}$. By standard techniques in convex analysis, the SOC can be equivalently written as:

$$H_\ell(\boldsymbol{x}) \succeq 0 \quad \Longleftrightarrow \quad \boldsymbol{c}_\ell^\top \boldsymbol{x} \ge \|\boldsymbol{D}_\ell \boldsymbol{x}\|_2, \tag{6}$$

where $\boldsymbol{c}_\ell \in \mathbb{R}^{n_x}$ is defined such that $\boldsymbol{c}_\ell^\top \boldsymbol{x} = \frac{1}{\sqrt{2}}(x_i^{\mathrm{mg}} + x_j^{\mathrm{mg}})$ for all $\boldsymbol{x} \in \mathcal{X}$ (i.e., $\boldsymbol{c}_\ell$ has coefficients $\frac{1}{\sqrt{2}}$ at locations corresponding to x_i^{mg} and x_j^{mg}), and $\boldsymbol{D}_\ell \in \mathbb{R}^{4\times n_x}$ is defined such that $\boldsymbol{D}_\ell \boldsymbol{x} = \begin{bmatrix} \frac{1}{\sqrt{2}}x_i^{\mathrm{mg}} & \frac{1}{\sqrt{2}}x_j^{\mathrm{mg}} & x_\ell^{\mathrm{re}} & x_\ell^{\mathrm{im}} \end{bmatrix}^\top$ for all $\boldsymbol{x} \in \mathcal{X}$.

Let the solution to (3) be denoted by $\hat{\boldsymbol{x}}$. We can estimate the bad data vector by

$$\hat{b}_i = \mathrm{sign}(y_i - \boldsymbol{a}_i^*\hat{\boldsymbol{x}}) \max\left(0, |y_i - \boldsymbol{a}_i^*\hat{\boldsymbol{x}}| - \lambda\right),$$

which turns out to be optimal for a mixed ℓ_1, ℓ_2 optimization that is equivalent to (3) (see Lemma 4 in the appendix).

Remark: Despite the wide usage of Huber loss in the literature, existing studies are limited to unconstrained cases and are ignorant of the computational graph [2], [16], [34]. We will analyze the SOC constrained case and study its robustness on a graph.

B. Stage 2: Projection to the lower-dimensional space

Based on the solution in Stage 1, the next stage reconstructs the state by projecting a solution $\hat{\boldsymbol{x}}$ of (3) back to the original lower-dimensional space. To do so, we construct a vector $\hat{\boldsymbol{v}}$ such that: (i) the magnitude at each node $k \in \mathcal{N}$ can be obtained by $|\hat{v}_k| = \sqrt{\hat{x}_k^{\mathrm{mg}}}$; and (ii) select an acyclic subgraph of $\mathcal{G}$ with the maximum number of edges and define the phase difference along each edge $\ell \coloneqq (i,j)$ of this subgraph as $\hat{\theta}_{ij} = \arctan \hat{x}_\ell^{\mathrm{im}} / \hat{x}_\ell^{\mathrm{re}}$. To estimate the phases at all nodes, we compute the following least-square solution:

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{L}^\top \boldsymbol{L})^{-1} \boldsymbol{L}^\top \boldsymbol{\theta}_\Delta, \tag{7}$$

where $\boldsymbol{\theta}_\Delta$ is the collection of $\hat{\theta}_{ij}$ and $\boldsymbol{L} \in \mathbb{R}^{\tilde{n}_l \times n_b}$ is a sparse matrix with $L(\ell, i) \coloneqq 1$ and $L(\ell, j) \coloneqq -1$ for each edge $\ell \coloneqq \{i,j\}$ of the acyclic subgraph and zero elsewhere ($\tilde{n}_l$ denotes the number of edges of the subgraph). Finally, we can reconstruct an estimate of the true state $\boldsymbol{v}_\natural$, denoted as $\hat{\boldsymbol{v}}$, via the formula:

$$\hat{v}_k = |\hat{v}_k| e^{\sqrt{-1}\hat{\theta}_k}, \qquad k \in \mathcal{N}. \tag{8}$$

If the regression vector from Step 1 is exact, i.e., $\hat{\boldsymbol{x}} = \boldsymbol{x}_\natural$, then (8) accurately recovers the system state, i.e., $\hat{\boldsymbol{v}} = \boldsymbol{v}_\natural$. If the $\hat{\boldsymbol{x}}$ is not exact, as long as the effect of bad data is significantly controlled, (7) has favorable properties and allows controlling the estimation error.

III. Graphical mutual incoherence

In this section, we discuss the proposed graphical mutual incoherence condition. A node k is said to be under attack and is denoted as $k \in \mathcal{N}_{\mathrm{at}}$ if any measurement that depends on

the nodal variable $x_k^{\rm mg}$ or edge variables $x_\ell^{\rm re}$ and $x_\ell^{\rm im}$ (with ℓ incident with k) is corrupted by bad data. For a given attack scenario, we define a partition of the network below.

Definition 1. *Given a measurement graph $\mathcal{G} := \{\mathcal{N}, \mathcal{L}\}$, we partition the graph as follows:*

- *Attacked region $\mathcal{B}_{\rm at} := \{\mathcal{N}_{\rm at}, \mathcal{L}_{\rm at}\}$ is the subgraph induced by attacked nodes $\mathcal{N}_{\rm at}$*
- *Inner boundary $\mathcal{B}_{\rm bi}$ is the subgraph induced by the nodes adjacent to the attacked nodes, defined as $\mathcal{N}_{\rm bi} := \{i \in \mathcal{N} \setminus \mathcal{N}_{\rm at} \mid \exists j \in \mathcal{N}_{\rm at}, \text{ s.t. } \{i,j\} \in \mathcal{L}\}$*
- *Outer boundary $\mathcal{B}_{\rm bo}$ is the subgraph induced by the set of nodes adjacent to the inner boundary nodes but not including the attacked nodes, defined as $\mathcal{N}_{\rm bo} := \{i \in \mathcal{N} \setminus (\mathcal{N}_{\rm at} \cup \mathcal{N}_{\rm bi}) \mid \exists j \in \mathcal{N}_{\rm bi}, \text{ s.t. } \{i,j\} \in \mathcal{L}\}$*
- *Boundary region $\mathcal{B}_{\rm bd} := \{\mathcal{N}_{\rm bd}, \mathcal{L}_{\rm bd}\}$ is the subgraph induced by the nodes in $\mathcal{N}_{\rm bd} := \mathcal{N}_{\rm bi} \cup \mathcal{N}_{\rm bo}$*
- *Safe region $\mathcal{B}_{\rm sf} := \{\mathcal{N}_{\rm sf}, \mathcal{L}_{\rm sf}\}$ is the subgraph induced by the remaining nodes, i.e., $\mathcal{N}_{\rm sf} := \mathcal{N} \setminus (\mathcal{N}_{\rm at} \cup \mathcal{N}_{\rm bd})$*

Moreover, we define $\mathcal{L}_{\rm at\cap bi}$ as the set of edges that connect nodes in $\mathcal{N}_{\rm at}$ to nodes in $\mathcal{N}_{\rm bi}$, and $\mathcal{L}_{\rm bi\cap bo}$ as the set of edges that connect nodes in $\mathcal{N}_{\rm bi}$ to nodes in $\mathcal{N}_{\rm bo}$.

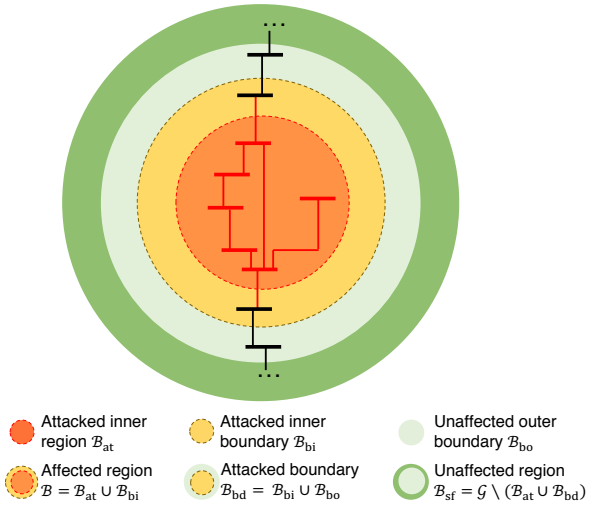

Fig. 2: Illustration of the partition introduced in Def. 1. Lines or buses whose measurements are under attack are marked red.

The partition set notations are illustrated in Fig. 2. Now, we introduce a partition of the measurements and variables.

Definition 2. *Given a partition of the graph according to Def. 1, we partition the variables in $\boldsymbol{x}$ as follows:*

- *Attacked variables $\boldsymbol{x}_{\rm at}$, consisting of nodal variables for $\mathcal{N}_{\rm at}$ (i.e., $x_k^{\rm mg}$ for $k \in \mathcal{N}_{\rm at}$) and edge variables for $\mathcal{L}_{\rm at} \cup \mathcal{L}_{\rm at\cap bi}$ (i.e., $x_\ell^{\rm re}$ and $x_\ell^{\rm im}$ for $\ell \in \mathcal{L}_{\rm at} \cup \mathcal{L}_{\rm at\cap bi}$)*
- *Boundary variables $\boldsymbol{x}_{\rm bd}$, consisting of nodal variables for $\mathcal{N}_{\rm bd}$ and edge variables for $\mathcal{L}_{\rm bd}$*
- *Safe variables $\boldsymbol{x}_{\rm sf}$, consisting of all other variables*

Accordingly, we denote $\boldsymbol{x}_{\natural {\rm at}}$, $\boldsymbol{x}_{\natural {\rm bd}}$ and $\boldsymbol{x}_{\natural {\rm sf}}$ as the partition of the true state vector $\boldsymbol{x}_\natural$; and $\hat{\boldsymbol{x}}_{\rm at}$, $\hat{\boldsymbol{x}}_{\rm bd}$ and $\hat{\boldsymbol{x}}_{\rm sf}$ as the partition of the estimated state vector $\hat{\boldsymbol{x}}$. We also denote $n_{\rm at}$, $n_{\rm bd}$ and $n_{\rm sf}$ as the number of variables in $\boldsymbol{x}_{\rm at}$, $\boldsymbol{x}_{\rm bd}$ and $\boldsymbol{x}_{\rm sf}$, respectively. The measurements are partitioned as follows:

- *Attacked measurements $\mathcal{M}_{\rm at}$, consisting of those that depend on $x_k^{\rm mg}$ for some $k \in \mathcal{N}_{\rm at}$, and/or $x_\ell^{\rm re}$ and $x_\ell^{\rm im}$ for some $\ell \in \mathcal{L}_{\rm at}$*
- *Inner boundary measurements $\mathcal{M}_{\rm bi}$, consisting of those that depend on $x_k^{\rm mg}$ for some $k \in \mathcal{N}_{\rm bi}$ and $x_\ell^{\rm re}$ and $x_\ell^{\rm im}$ for some $\ell \in \mathcal{L}_{\rm at\cap bi}$*
- *Outer boundary measurements $\mathcal{M}_{\rm bo}$, consisting of those that depend on $x_k^{\rm mg}$ for some $k \in \mathcal{N}_{\rm bo}$ and $x_\ell^{\rm re}$ and $x_\ell^{\rm im}$ for some $\ell \in \mathcal{L}_{\rm bd}$*
- *Boundary measurements $\mathcal{M}_{\rm bd} := \mathcal{M}_{\rm bi} \cup \mathcal{M}_{\rm bo}$, including both the inner and outer boundary measurements*
- *Safe measurements $\mathcal{M}_{\rm sf}$, consisting of the remaining measurements*

The above definition allows one to "rearrange" the matrix $\boldsymbol{A}$ in the following form such that the attacked and safe regions become "weakly coupled" through the boundary region:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{\mathcal{M}_{\rm sf},\mathcal{X}_{\rm sf}} & \boldsymbol{A}_{\mathcal{M}_{\rm sf},\mathcal{X}_{\rm bd}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\rm bo},\mathcal{X}_{\rm bd}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\rm bi},\mathcal{X}_{\rm bd}} & \boldsymbol{A}_{\mathcal{M}_{\rm bi},\mathcal{X}_{\rm at}} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\rm at},\mathcal{X}_{\rm at}} \end{bmatrix}. \tag{9}$$

Remark: The above partition exploits the sparsity of the computational graph to weakly isolate the attacked region and the safe region. We will illustrate the direct benefit of such partitioning below. However, the challenge is that the partition is not known *a priori*, which motivates the graphical mutual incoherence condition in Sec. III-B.

A. Preliminary results

We first introduce some regularity conditions.

Condition 1 (Measurement normalization)**.** *Let $\boldsymbol{a}_i$ be the i^{th} row of $\boldsymbol{A}$. Assume that all rows are normalized, i.e., $\|\boldsymbol{a}_i\|_2 = 1$ for all $i \in [n_m]$.*

This condition is straightforward to implement in practice, since the sensing matrix $\boldsymbol{A}$ is fixed for a given set of measurements. This is also known as preconditioning, which assists with the statistical performance of regression.

Condition 2 (Lower eigenvalue)**.** *Let $\boldsymbol{Q}_{\mathcal{M}_{\rm bd},\mathcal{X}_{\rm bd}} := \begin{bmatrix} \boldsymbol{A}_{\mathcal{M}_{\rm bd},\mathcal{X}_{\rm bd}} & \boldsymbol{I}_{\mathcal{M}_{\rm bi}}^\top \end{bmatrix}$. Then, the lower eigenvalue bou d $C_{\rm min}$ is defined as*

$$\min \left\{ \begin{aligned} &\lambda_{\min}\left(\boldsymbol{Q}_{\mathcal{M}_{\rm bd},\mathcal{X}_{\rm bd}}^\top \boldsymbol{Q}_{\mathcal{M}_{\rm bd},\mathcal{X}_{\rm bd}}\right), \\ &\lambda_{\min}\left(\boldsymbol{A}_{\mathcal{M}_{\rm bo},\mathcal{X}_{\rm bd}}^\top \boldsymbol{A}_{\mathcal{M}_{\rm bo},\mathcal{X}_{\rm bd}}\right), \\ &\lambda_{\min}\left(\boldsymbol{A}_{\mathcal{M}_{\rm sf},\mathcal{X}_{\rm sf}}^\top \boldsymbol{A}_{\mathcal{M}_{\rm sf},\mathcal{X}_{\rm sf}}\right) \end{aligned} \right\} \geq C_{\rm min}. \tag{10}$$

The value $C_{\rm min}$ characterizes the identifiability of $\boldsymbol{x}_\natural$ *outside* the attacked region for a given set of measurements. If $C_{\rm min}$ is strictly positive and one can accurately detect the support of bad data, then this condition ensures the accurate estimation of $\boldsymbol{x}_\natural$ outside the attacked region. To analyze the algorithm that incorporates SOCs, we also need to introduce the following condition.

Condition 3 (Non-binding SOCs at boundary)**.** *Define the sets*

$$\mathcal{K}_{\rm bd} \coloneqq \left\{ \boldsymbol{x}_{\rm bd} \;\middle|\; H_\ell(\boldsymbol{x}) \succeq 0, \quad \forall \ell \in \mathcal{L}_{\rm bd} \right\}, \tag{11}$$

$$\mathcal{K}_{\rm at} \coloneqq \left\{ \boldsymbol{x}_{\rm at} \;\middle|\; H_\ell(\boldsymbol{x}) \succeq 0, \quad \forall \ell \in \mathcal{L}_{\rm at} \cup \mathcal{L}_{\rm at\cap bi} \right\}, \tag{12}$$

for boundary and attacked variables, respectively, and let

$$\tilde{\mathcal{K}}_{\rm at}(\hat{\boldsymbol{x}}_{\rm bd}) = \Bigg\{ \boldsymbol{x}_{\rm at} \;\Big|\; H_\ell(\boldsymbol{x}) \succeq 0, \forall \in \mathcal{L}_{\rm at} \cup \mathcal{L}_{\rm at\cap bi}, \text{ and } \boldsymbol{x} \text{ such that } \boldsymbol{x}_{\rm bd} = \hat{\boldsymbol{x}}_{\rm bd} \Bigg\},$$

be the confined feasible set for $\boldsymbol{x}_{\rm at}$ with the boundary variables fixed at $\hat{\boldsymbol{x}}_{\rm bd}$ in the SOC constraints. We say that the solution $\hat{\boldsymbol{x}}_{\rm at}$ satisfies the non-binding condition if $\hat{\boldsymbol{x}}_{\rm at} \in \tilde{\mathcal{K}}_{\rm at}(\boldsymbol{x}_{\natural \rm bd})$.

This condition simply requires that the values of the estimated attack variable and the true boundary variable lie within the set defined by SOCs. When there is an attack on a local region, a subset of the local measurements are compromised. Our goal is to recover the states outside the attacked region, namely *local recovery*, rather than for the entire network, namely *global recovery*. The following lemma provides a preliminary result for solving the estimation problem in the absence of dense noise.

Lemma 1 (SOCP)**.** *Suppose that there is no dense noise (i.e., $\boldsymbol{w} = \boldsymbol{0}$ in (2)), and that the lower eigenvalue condition is satisfied. Assume that for an arbitrary $\boldsymbol{b}_{\mathcal{M}_{\rm bd}}$ with its support limited to the inner boundary, the solution $\hat{\boldsymbol{x}}_{\rm bd} \in \mathcal{X}_{\rm bd}$ to the program*

$$\min_{\boldsymbol{x}_{\rm bd} \in \mathcal{K}_{\rm bd}} \|\boldsymbol{z}_{\mathcal{M}_{\rm bd}} - \boldsymbol{A}_{\mathcal{M}_{\rm bd}, \mathcal{X}_{\rm bd}} \boldsymbol{x}_{\rm bd}\|_1, \tag{13}$$

is unique and satisfies $\hat{\boldsymbol{x}}_{\rm bd} = \boldsymbol{x}_{\natural \rm bd}$, where $\boldsymbol{z}_{\mathcal{M}_{\rm bd}} = \boldsymbol{A}_{\mathcal{M}_{\rm bd}, \mathcal{X}_{\rm bd}} \boldsymbol{x}_{\natural \rm bd} + \boldsymbol{b}_{\mathcal{M}_{\rm bd}}$, and that the solution $\hat{\boldsymbol{x}}_{\rm at}$ to

$$\min_{\boldsymbol{x}_{\rm at} \in \mathcal{K}_{\rm at}} \|\boldsymbol{y}_{\mathcal{M}_{\rm at}} - \boldsymbol{A}_{\mathcal{M}_{\rm at}, \mathcal{X}_{\rm at}} \boldsymbol{x}_{\rm at}\|_1, \tag{14}$$

also satisfies the non-binding condition. Then, the solution $\hat{\boldsymbol{x}}$ to the conic program:

$$\min_{\boldsymbol{x} \in \mathcal{K}} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_1 \tag{15}$$

satisfies $\hat{\boldsymbol{x}}_{\rm bd} = \boldsymbol{x}_{\natural \rm bd}$ and $\hat{\boldsymbol{x}}_{\rm sf} = \boldsymbol{x}_{\natural \rm sf}$.

Intuitively, conditioning on the boundary variables $\boldsymbol{x}_{\rm bd}$, the attacked variables $\boldsymbol{x}_{\rm at}$ and safe variables $\boldsymbol{x}_{\rm sf}$ are independent, which can be regarded as decoupling the "weakly coupled" system. Therefore, if the boundary variables are correct (by assumption), then the safe variables can be recovered. The proof of this lemma is based on carefully analyzing the Karush-Kuhn-Tucker (KKT) conditions. Since the proof can be derived based on the proof of Theorem 1 in Sec. IV, which is more general, we omitted the details.

The key assumption in the previous results is the uniqueness and correctness of solving (13). However, verifying this assumption requires enumerating over the support of bad data, which can have an exponential number of possibilities. This motivates the development of a new condition below.

B. Graphical mutual incoherence

We propose the notion of graphical mutual incoherence (gMI) as a sufficient condition to certify the recovery of the boundary variables in the presence of arbitrary bad data. To this end, we introduce the following concepts.

Definition 3. *Given an edge $\ell \coloneqq \{i, j\}$ with node i in the attacked region and node j in the inner boundary, define local partitions as follows:*

- *Local attack region $\mathcal{B}_{\rm at}^{i\to j} \coloneqq \{\{i\}, \emptyset\}$ has only one node*
- *Local inner boundary $\mathcal{B}_{\rm bi}^{i\to j} \coloneqq \{\{j\}, \emptyset\}$ has only one node*
- *Local outer boundary $\mathcal{B}_{\rm bo}^{i\to j} \coloneqq \{\mathcal{N}_{\rm bo}^{i\to j}, \mathcal{L}_{\rm bo}^{i\to j}\}$ is the subgraph induced by nodes other than i that are connected to j, i.e., $\mathcal{N}_{\rm bo}^{i\to j} \coloneqq \{k \in \mathcal{N} \setminus \{i\} \mid \{j, k\} \in \mathcal{L}\}$*
- *We also use $\mathcal{L}_{\rm bd}^{i\to j}$ to represent the union of edges that connect nodes in $\mathcal{B}_{\rm bi}^{i\to j}$ and those in $\mathcal{B}_{\rm bo}^{i\to j}$*

Similarly, we introduce the local versions of the partitions of variables and measurements:

- *Local boundary variables $\mathcal{X}_{\rm bd}^{i\to j}$ include $\{x_k^{\rm mg}\}_{k \in \mathcal{B}_{\rm bi}^{i\to j} \cup \mathcal{B}_{\rm bo}^{i\to j}}$ and $\{x_\eta^{\rm re}, x_\eta^{\rm im}\}_{\eta \in \mathcal{L}_{\rm bd}^{i\to j}}$*
- *Local boundary measurements $\mathcal{M}_{\rm bd}^{i\to j} \coloneqq \mathcal{M}_{{\rm bd}\checkmark}^{i\to j} \cup \mathcal{M}_{{\rm bd}\times}^{i\to j}$ include those that depend only on the boundary variables $\mathcal{X}_{\rm bd}^{i\to j}$, denoted by $\mathcal{M}_{{\rm bd}\checkmark}^{i\to j}$, and those that depend on both $\mathcal{X}_{\rm bd}^{i\to j}$ and variables $\{x_\ell^{\rm re}, x_\ell^{\rm im}\}$ for $\ell \coloneqq \{i, j\}$, denoted by $\mathcal{M}_{{\rm bd}\times}^{i\to j}$*

We also let $n_\checkmark^{i\to j} = |\mathcal{M}_{{\rm bd}\checkmark}^{i\to j}|$, $n_\times^{i\to j} = |\mathcal{M}_{{\rm bd}\times}^{i\to j}|$, and $n_{\mathcal{L}}^{i\to j} = |\mathcal{L}_{\rm bd}^{i\to j}|$ be the number of correct measurements, the number of wrong measurements, and the number of boundary lines, respectively. The above terms can be similarly defined for the direction $j \to i$. Thus, for each line, we have two sets of boundary variables and measurements.

With the above notations, we can now define the graphical mutual incoherence (gMI). To begin with, we introduce the gMI for the estimation problem (3) without the SOCs, which coincides with the algorithm in [17].

Definition 4 (gMI for estimation without SOCs)**.** *For each line $\ell = \{i, j\} \in \mathcal{L}$, define the graphical mutual incoherence $\alpha_{i\to j}$ along the direction $i \to j$ as the globally optimal objective value of the following optimization problem:*

$$\max_{\boldsymbol{\xi} \in \{-1,+1\}^{n_\times^{i\to j}}} \quad \min_{\alpha \in \mathbb{R}, \boldsymbol{h} \in \mathbb{R}^{n_\checkmark^{i\to j}}} \quad \alpha \tag{16a}$$

$$\text{s.t.} \quad \boldsymbol{A}_{\mathcal{M}_{{\rm bd}\checkmark}^{i\to j}, \mathcal{X}_{\rm bd}^{i\to j}}^\top \boldsymbol{h} + \boldsymbol{A}_{\mathcal{M}_{{\rm bd}\times}^{i\to j}, \mathcal{X}_{\rm bd}^{i\to j}}^\top \boldsymbol{\xi} = \boldsymbol{0} \tag{16b}$$

$$\|\boldsymbol{h}\|_\infty \le \alpha \tag{16c}$$

Similarly, we can define gMI $\alpha_{j\to i}$ for the direction $j \to i$.

Intuitively, gMI measures the correlation between the correct data and the corrupted data. The name "mutual incoherence" originates from the compressed sensing literature [35], [36]. However, the gMI proposed in this study is different. First, gMI is defined on a single line, and we build a theoretical certificate from bottom up by leveraging the graph topology. This alleviates the dependence on each instance of the bad data support. Second, as we will introduce in Sec. IV, gMI can

be applied to local recovery, while existing conditions in the literature are all designed for global recovery. Also, gMI can be solved efficiently (see Sec. III-C), while other conditions cannot be easily verified for large-scale systems. Moreover, we show that gMI is much less conservative than the existing conditions. Next, we extend the definition to the estimation problem (3) proposed in this work.

Definition 5 (gMI for estimation with SOCs)**.** *For each edge $\ell = \{i, j\} \in \mathcal{L}$ and a given $\boldsymbol{x} \in \mathcal{K}$, define the gMI $\alpha_{i\to j}^{\mathrm{SOCP}}(\boldsymbol{x})$ along the direction $i \to j$ as the globally optimal value of the following optimization problem:*

$$\underset{\boldsymbol{\xi}\in\{-1,+1\}^{n_{\times}^{i\to j}}}{max} \quad \underset{\alpha\in\mathbb{R},\boldsymbol{\omega}\in\mathbb{R}^{n_{\mathcal{L}}^{i\to j}},\boldsymbol{h}\in\mathbb{R}^{n_{\checkmark}^{i\to j}}}{min} \alpha \tag{17a}$$

$$s.t. \qquad \|\boldsymbol{h}\|_\infty \le \alpha \tag{17b}$$

$$\omega_\ell \ge 0, \qquad \forall \ell \in \mathcal{L}_{\mathrm{bd}}^{i\to j} \tag{17c}$$

$$\boldsymbol{A}_{\mathcal{M}_{\mathrm{bd}\checkmark}^{i\to j},\mathcal{X}_{\mathrm{bd}}^{i\to j}}^{\top}\boldsymbol{h} + \boldsymbol{A}_{\mathcal{M}_{\mathrm{bd}\times}^{i\to j},\mathcal{X}_{\mathrm{bd}}^{i\to j}}^{\top}\boldsymbol{\xi} + \sum_{\ell\in\mathcal{L}_{\mathrm{bd}}^{i\to j}} \omega_\ell \boldsymbol{T}_\ell \boldsymbol{x} = \boldsymbol{0} \tag{17d}$$

where $\boldsymbol{T}_\ell = \boldsymbol{c}_\ell\boldsymbol{c}_\ell^\top - \boldsymbol{D}_\ell^\top\boldsymbol{D}_\ell$. The gMI $\alpha_{j\to i}^{\mathrm{SOCP}}(\boldsymbol{x})$ for direction $j \to i$ can be defined similarly.

The closest condition that measures the alignment of the sensing directions of the corrupted measurements (i.e., $\boldsymbol{A}_{\mathcal{J}}$, where $\mathcal{J}$ is the support of the bad data) with those of the clean data (i.e., $\boldsymbol{A}_{\mathcal{J}^c}$) has been proposed in [17]. For each edge $\ell \in \mathcal{L}$, the mutual incoherence metric defined in [17] is given by:

$$\rho(\mathcal{M}_{\mathrm{bd}\times}^{i\to j}) = \|\boldsymbol{A}_{\mathcal{M}_{\mathrm{bd}\checkmark}^{i\to j},\mathcal{X}_{\mathrm{bd}}^{i\to j}}^{\top+}\boldsymbol{A}_{\mathcal{M}_{\mathrm{bd}\times}^{i\to j},\mathcal{X}_{\mathrm{bd}}^{i\to j}}^{\top}\|_\infty,$$

where $\boldsymbol{A}_{\mathcal{J}}^{+} = (\boldsymbol{A}_{\mathcal{J}}^\top\boldsymbol{A}_{\mathcal{J}})^{-1}\boldsymbol{A}_{\mathcal{J}}^\top$ denotes the pseudo-inverse. We next show the relationship among these measures.

Proposition 1. *For each edge $\ell \in \mathcal{L}$, it holds that*

$$\rho(\mathcal{M}_{\mathrm{bd}\times}^{i\to j}) \ge \alpha_{i\to j} \ge \alpha_{i\to j}^{\mathrm{SOCP}}(\boldsymbol{x})$$

for any $\boldsymbol{x} \in \mathcal{K}$.

As we will see in Sec. IV, boundary defense requires a low value for gMI or mutual incoherence. The above result implies that gMI is always less conservative than the mutual incoherence proposed in [17], and the incorporation of SOCs can certifiably improve robustness.

C. Computational aspect

The minimax programs (16) and (17) used to define gMIs consist of a convex optimization in the inner minimization and a discrete optimization in the outer maximization. For problems where the number of measurements in $\mathcal{M}_{\mathrm{bd}\times}^{i\to j}$ is not large, it is computationally tractable to enumerate all the feasible points in the outer maximization. This is the case in our experiments for the power system state estimation. Based on standard convex analysis, we have developed a more scalable method for the case with a large $n_{\times}^{i\to j}$ by reformulating the problem as a linear complimentarity problem (LCP) [37], which can be solved readily using off-the-shelf solvers such as PATH Solver [38] or YALMIP. Alternatively, we can reformulate the complimentarity slackness conditions as a mixed-integer program, and solve the problem using standard packages such as Gurobi.

IV. Boundary defense mechanism

In this section, we first establish that global defense is certified if the gMI conditions are satisfied for all the lines on the boundary. Then, we derive theoretical guarantees for bad data detection and estimation error. Although we focus on the estimation problem (3), the technical proof can be easily adapted to the case without the SOCs, which corresponds to the unconstrained program of (3).

Define

$$\boldsymbol{A}^\circ = \begin{bmatrix} \boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}},\mathcal{X}_{\mathrm{sf}}} & \boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}},\mathcal{X}_{\mathrm{bd}}} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\mathrm{bo}},\mathcal{X}_{\mathrm{bd}}} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\mathrm{bi}},\mathcal{X}_{\mathrm{bd}}} \end{bmatrix}$$

as a subset of $\boldsymbol{A}$ that removes the rows and columns corresponding to $\mathcal{M}_{\mathrm{at}}$ and $\boldsymbol{x}_{\mathrm{at}}$, respectively. Similarly, define $\boldsymbol{c}_\ell^\circ$ and $\boldsymbol{D}_\ell^\circ$ as the subvector and submatrix of $\boldsymbol{c}_\ell$ and $\boldsymbol{D}_\ell$ that remove the entries or rows corresponding to $\boldsymbol{x}_{\mathrm{at}}$, respectively. We also define $\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^\circ = \begin{bmatrix}\boldsymbol{A}^\circ & \boldsymbol{I}_{\mathcal{M}_{\mathrm{bi}}}^\top\end{bmatrix}$ and $\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ+} = (\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^\circ)^{-1}\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}$ as its pseudo-inverse, and let $\boldsymbol{I}_x$ and $\boldsymbol{I}_b$ be the matrices that consist of the first $n_{\mathrm{bd}}+n_{\mathrm{sf}}$ rows and the last $|\mathcal{M}_{\mathrm{bi}}|$ rows of the identity matrix of size $n_{\mathrm{bd}}+n_{\mathrm{sf}}+|\mathcal{M}_{\mathrm{bi}}|$, respectively.

Condition 4 (gMI condition)**.** *The gMI condition is satisfied if $\alpha_{i\to j}^{\mathrm{SOCP}} \le 1-\gamma$ for all $\{i, j\} \in \mathcal{L}_{\mathrm{at}\cap\mathrm{bi}}$ with $i \in \mathcal{N}_{\mathrm{at}}$ and $j \in \mathcal{N}_{\mathrm{bi}}$, for some positive constant γ.*

Given an attack scenario $\mathcal{N}_{\mathrm{at}}$, the gMI condition can be verified by checking every line in the boundary. Since the gMI considers the worst-case guarantee and is independent of the attack scenario, a single map of gMI can be used to verify an exponential number of attack scenarios. If the condition is not satisfied, one can artificially increase the set $\mathcal{N}_{\mathrm{at}}$ by adding the nodes of violated gMI into the attack set until the condition is met. We will provide a vulnerability map of the U.S. grid in Sec. V based on this concept. The main result of this paper is provided below.

Theorem 1. *Assume that the gMI condition is satisfied with a constant $\gamma > 0$, and that the lower eigenvalue condition and the non-binding SOCs condition are satisfied. Suppose that the hyperparameter λ in Huber loss is chosen such that*

$$\lambda > \frac{2}{n_m\gamma}\sqrt{2\sigma^2\log n_m}. \tag{18}$$

Then, the following properties hold for the solution to (3)*, denoted as $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{b}})$:*

1. *The solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{b}})$ has no false bad data inclusion (i.e., $\mathrm{supp}(\hat{\boldsymbol{b}}) \subseteq \mathrm{supp}(\boldsymbol{b}_\natural)$) with probability greater than $1 - \frac{c_0}{n_m}$, for some constant $c_0 > 0$.*
2. *Define $g(\lambda)$ as*

$$n_m\lambda\left(\frac{1}{2\sqrt{C_{min}}} + \|\boldsymbol{I}_b\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ+}\|_\infty\right),$$

and let $\tilde{\boldsymbol{b}}_{\mathcal{M}_{\mathrm{bi}}} = \boldsymbol{A}_{\mathcal{M}_{\mathrm{bi}},\mathcal{X}_{\mathrm{at}}}(\boldsymbol{x}_{\natural\mathrm{at}} - \hat{\boldsymbol{x}}_{\mathrm{at}})$ be the mismatch at the boundary caused by a potentially false estimate

of $\hat{\boldsymbol{x}}_{\rm at}$*. Then, all bad data with magnitude greater than* $g(\lambda)$ *will be detected (i.e., if* $|\tilde{b}_i| > g(\lambda)$*, then* $|\hat{b}_i| > 0$*) with probability greater than* $1 - \frac{c_2}{m}$*.*

(3) *(Bounded error) The estimator error is bounded by*

$$\|\boldsymbol{x}_{\natural\mathcal{X}_{\rm sf}\cup\mathcal{X}_{\rm bd}} - \hat{\boldsymbol{x}}_{\mathcal{X}_{\rm sf}\cup\mathcal{X}_{\rm bd}}\|_2 \leq t\frac{\sqrt{|\mathcal{X}_{\rm sf}| + |\mathcal{X}_{\rm bd}| + |\mathcal{M}_{\rm bi}|}}{C_{min}} + n_m\lambda\|\boldsymbol{I}_x\boldsymbol{Q}^{\circ+}_{\mathcal{M}_{\rm bi}}\|_{\infty,2}$$

with probability greater than $1 - \exp\left(-\frac{c_1 t^2}{\sigma^4}\right)$.

The proof of the theorem is shown in Appendix C. Inspired by Sec. III-A, a key step in establishing the result is to ensure that local defense is sufficient to guard against attacks when solving the problem globally. In appendix B, we prove that as long as the gMI condition is met, we have a desirable property in terms of defending against bad data on the boundary. While the theorem only focuses on (3), the result for the unconstrained optimization can be derived similarly. The main advantage of (3) over the case without SOCs is that the gMI condition is more likely to be satisfied due to Prop. 1.

Theorem 1 provides formal guarantees of bad data detection. From the measurements, we first estimate the variables in the lifted space using first-stage algorithm (3). Then, we threshold the estimated bad data vector to determine its support. By result (1) above, for large n_m, with high probability the support will be confined within the attacked region; by (2), corrupted measurements with large enough magnitudes will be detected. Thus, our approach can be used to detect the attacked region and guarantee the lifted variables in the boundary and safe region can be recovered accurately by result (3). Finally, the recovered lifted variables are fed into the second-stage algorithm (7) and (8) to produce a state estimation.

V. Experiments

A. Power system state estimation

Power system state estimation is an important instance of graph-structured quadratic sensing. The electric grid is modeled as a graph $\mathcal{G} := \{\mathcal{N}, \mathcal{L}\}$, where $\mathcal{N} := [n_b]$ and $\mathcal{L} := [n_l]$ represent its sets of buses (i.e., nodes) and branches (i.e., edges). The power system state is described by the complex voltage $\boldsymbol{v} = \left[v_1, ..., v_{n_b}\right]^\top \in \mathbb{C}^{n_b}$, where $v_k \in \mathbb{C}$ is the complex voltage at bus $k \in \mathcal{N}$ with magnitude $|v_k|$ and phase $\theta_k := \angle v_k$. By Ohm's law, the measurements obtained by the supervisory control and data acquisition (SCADA) system, including voltage magnitude squares, real and reactive power injection at each bus, and real and reactive power flow along each branch can be represented in the form of quadratic measurements (1). The Hermitian matrix $\boldsymbol{M}_i$ follows the graph-induced sparsity pattern; therefore, the physical graph coincides with the computational graph. The goal of SE is to reliably infer about the underlying state $\boldsymbol{v}$ given noisy and corrupted measurements y_i.

Here, we focus on the U.S. grid, which is the largest machine on earth with more than 200,000 miles of transmission lines. Due to confidentiality requirements, we report our findings on modified grids provided in [39], which match the size, complexity, and characteristics of actual grids.

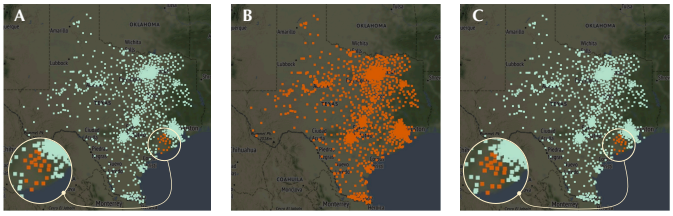

Fig. 3: Evaluation of the boundary defense mechanism. **(A)** The grid is under "zonal attack," where the measurements within a zone are corrupted (shown in red). SE based on **(B)** Newton's method for nonlinear least squares, and **(C)** the proposed method with SOCs, where in both cases, buses with an estimation error greater than 0.002 are marked in red.

B. Adversary model and zonal attack

We are concerned with the scenario where the data for an entire subregion are compromised. We assume that the attacker has access to the model and can manipulate every measurement within the region under attack in an arbitrary way. Specifically, we consider the "zonal attack" (Fig. 3), where all measurements within a zone—usually governed by a single utility—are corrupted. In this example, we consider the ERCOT network with 2,000 nodes, where a subgraph around Houston with about 19 nodes are under attack. In this case, Newton's method is seriously affected by the bad data, whereas our proposed method can recover the state outside the attacked zone correctly. In the case of a stealth attack, there is a problem of symmetry, namely, without additional information, it is impossible to decide which zone is under attack since the only inconsistency is observed at the boundary. To avoid this case, we arbitrarily break the symmetry by introducing some sensors within the attacked zone that are more secure than others in such a way that their values cannot be modified.

C. Geographic mapping of vulnerabilities

Based on the mathematical tools developed in the study, we assess the robustness of the synthetic U.S. grid.

Definition 6. *A line* $\{i, j\} \in \mathcal{L}$ *is said to be a robust line if* $\alpha_{i\to j} < 1$ *and* $\alpha_{j\to i} < 1$*; otherwise, it is said to be a vulnerable line (V-line).*

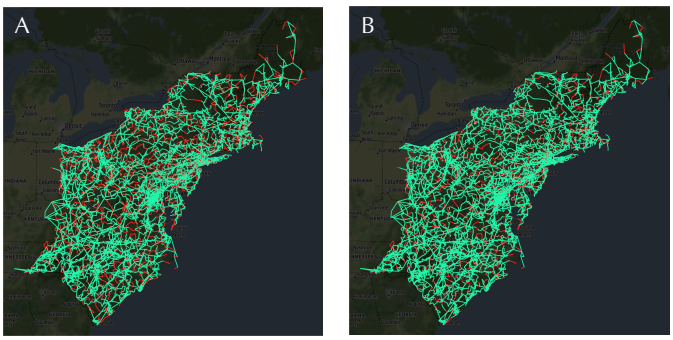

Fig. 4: Vulnerability maps based on the proposed gMI **(A)** without SOCs (16) and **(B)** with SOCs (17), which marked robust lines (in green) and vulnerable lines (in red).

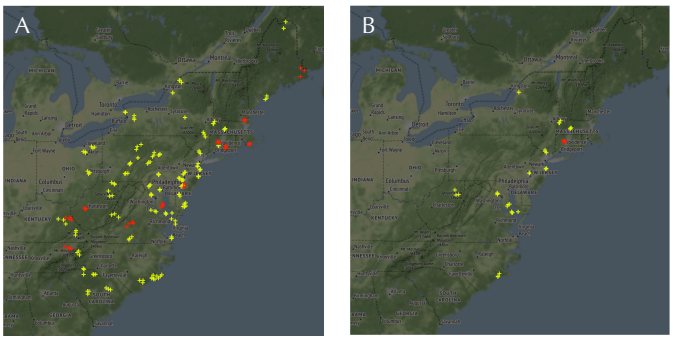

Fig. 5: Comparison of bus critical index with and without SOCs. Since the bus critical indices are no larger than 3, we only show the locations with values 2 (yellow) and 3 (red) for gMIs **(A)** without SOCs and **(B)** with SOCs.

First, we visualize the geographic distribution of robust and vulnerable lines for the Eastern U.S. grid in Fig. 4. It can be seen that the density of vulnerable lines is relatively high for populated areas, such as Boston and New York, where we also observe a high density of robust lines. On average, 59% lines are robust across the states, which are then split further into independent synchronous regions, as shown in Table I. In addition, the map validates Proposition 1 that the incorporation of SOCs can help rectify SE and better detect bad data.

The vulnerability map can be used in various ways. For instance, it can be used to investigate whether topological errors for a line or a substation can be contained locally, in which case the associated measurements are largely biased.

Definition 7 (Critical bus and critical line)**.** *For a node $i \in \mathcal{N}$, if there exists a neighboring node j such that $\alpha_{i\to j} \geq 1$, then the node i is called a critical bus (C-bus). A branch $\{i,j\} \in \mathcal{L}$ is a critical branch (C-line) if there exists a node k adjacent to either i or j such that $\alpha_{i\to k} \geq 1$ or $\alpha_{j\to k} \geq 1$ (or both conditions are satisfied).*

Specifically, if the erroneous line/substation is surrounded by robust lines, then it is guaranteed by Theorem 1 that the error will be contained locally via the boundary defense mechanism. Otherwise, there is a possibility that the error will "escape" outside the boundary to affect the outside region. Summary statistics are shown in Table I.

D. Criticality index for substations under cyberattack

Furthermore, we can extend the case study by defining a criticality index (CI) for each substation.

Definition 8 (Criticality index)**.** *Given a node i_1, an arbitrary node i_n is path-connected to i_1 if there is a path $i_1, i_2, ..., i_n$ such that $\alpha_{i_k\to i_{k+1}} \geq 1$ for $k = 1, ..., n-1$. The criticality index at node i is defined as the number of nodes that are path-connected to i.*

The CI gauges how many nodes near a substation will be affected if the substation is under attack. The higher the value is, the more crucial the situation is when the substation is compromised. This situation is analogous to the cascading failures of generators, but the difference is clear—our focus is on the algorithmic robustness rather than the physical dynamics. We visualize the distribution of the CIs on the map shown in Fig. 5. It can be observed that (i) the highest number is 3; and (ii) the incorporation of SOCs improve CIs.

E. Network and optimization properties

So far, our study has been conducted with respect to a specific measurement profile. Important questions are: How do the number and locations of measurement sensors affect line vulnerability? In particular, does decreasing the number of sensors make the network significantly more vulnerable? What type of sensor measurements can bolster boundary defenses?

For this purpose, we examine three methods used for "measurement selection." The first method (Method 1) starts from a spanning tree of the network and adds a set of lines to the tree incrementally to obtain a subgraph that will be used for taking measurements. In this method, each bus is equipped with only voltage magnitude measurements and each line has three out of four branch flow measurements. The second method (Method 2) starts with the full network, where each node has voltage magnitude measurements, and each line has one real and one reactive power measurement, and it grows the set of sensors by randomly adding branch measurements. The third method (Method 3) differs from Method 2 only in that it grows the set of sensors by randomly adding branch measurements as well as nodal power injections.

To evaluate these three methods, we devise a "scattered attack" strategy, where we randomly select 25 lines from the 2000-bus Texas interconnection and corrupt all of its branch measurements, which amounts to roughly 100 bad pieces of data. We then employ our proposed SE method. The observation is that, in general, both the root mean squared error (RMSE) and the F1 score for bad data detection are enhanced as more sensors are added to the network, as shown in Fig. 6. The F1 score is given by $\frac{2\text{precision}\times\text{recall}}{\text{precision}+\text{recall}}$, where precision is the rate of true positives (i.e., correctly identified bad data) among all data that are claimed to be bad, whereas recall is given by the percentage of true positives identified as bad data among all ground truth bad data. Specifically, an F1 score close to 1 indicates that the algorithm detects all corrupted data (high recall rate) and does not falsely blame the correct data (high precision rate).

There is also a major discrepancy among the above methods for the same level of measurement redundancy. For instance, Method 1 significantly outperforms the other two methods at a low redundancy rate, whereas Method 2 steadily outmatches Method 3 with more sensors. To explain this phenomenon, we need to examine the types of available measurements. Thus, we select five typical measurement profiles as snapshots of Fig. 6 and calculate the percentage of V-lines and C-lines, and the average CI in each case (Fig. 7). It turns out that the inclusion of voltage magnitude or branch flow measurements can enhance the robustness, whereas the addition of nodal power injections is a major factor in weakening the defense. For example, with only voltage magnitude and branch flow measurements, the network is almost "everywhere defendable." On the contrary, with the inclusion of nodal injections, even

TABLE I: **Summary statistics of network properties and vulnerability characteristics.** We show the percentage of V-lines and C-lines among all network lines, and the percentage of C-buses among all network buses for QP and SOCP. We also show the average bus critical index, which measures the influence of a single-bus attack on the rest of the network.

	Basic properties		Properties of gMI (16)				Properties of gMI (17)			
	Buses	Lines	V-lines	C-lines	C-bus	Bus CI	V-lines	C-lines	C-bus	Bus CI
Texas	2,000	3,206	.3762	.4251	.4775	.20	.2979	.3674	.4225	.06
Western	10,000	12,706	.4715	.5231	.5313	.15	.3979	.4636	.4860	.06
Eastern	70,000	88,207	.4932	.5415	.5327	.14	.4104	.4780	.4810	.05

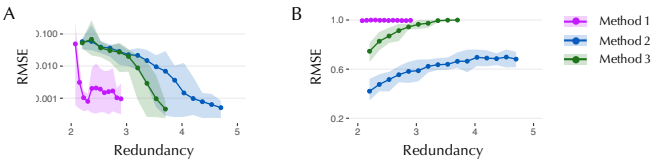

Fig. 6: Comparison of different measurement profiles and redundancy. The redundancy value is calculated as the number of sensors divided by $2 \times n_b$(number of buses) $- 1$, which represents the degrees of freedom in the traditional power flow problem. Each point for the **(A)** RMSE and **(B)** F1 score is obtained by averaging over 100 independent simulations. The average value is shown by the solid line, and the 5% and 95% quantiles are shown by the shaded region.

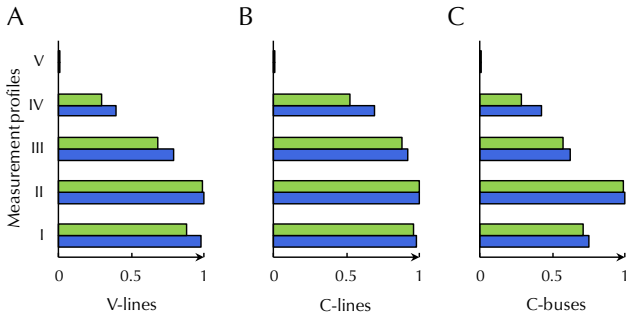

Fig. 7: Characterization of vulnerability based on measurement profiles. The five measurement profiles are: full nodal measurements and two/three/four branch flows per line (I/III/IV); real and reactive power injections per bus and three branch flows per line (II); and voltage magnitude per bus and three branch flows per line (V). For SEs with (green bar) and without (red bar) SOCs, we show the percentage of **(A)** V-lines, **(B)** C-lines, and **(C)** C-buses within the Texas interconnection.

with a high rate of branch flow measurements, the network is still vulnerable. Intuitively, this situation occurs because nodal power injections are highly coupled measurements that depend on state variables for all lines connected to the node. In contrast, voltage magnitudes and branch flows are more localized in nature, and, when corrupted, they have a smaller effect on adjacent buses/lines.

In addition to the measurement set, network vulnerability also depends on topological properties. In particular, our findings show that the connectivity degree for each node is positively correlated with line vulnerability (Fig. 8(A)). A

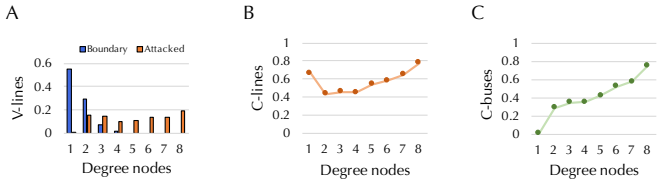

Fig. 8: Characterization of vulnerability through nodal degrees. **(A)** Percentage of V-lines when the nodes are at the boundary or in the attacked region. In this case, we distinguish the two directions of a line. Percentage of **(B)** C-lines and **(C)** C-buses averaged over nodes with the same degree. Since the distribution of nodal degrees is light-tailed, we group nodes of degree eight or higher in the same bin.

boundary defense node is increasingly likely to defend against attacks as the degree increases. However, this trend is less obvious when the node is under attack, since high-degree nodes have more measurements from the region not under attack to leverage in order to rectify the corrupted lines. On the other hand, it is more likely that a line will be critical if it is connected to a high-degree bus, as shown in Fig. 8(B). This criticality can be explained via the definition of a critical line, and as long as at least one of the remaining lines incident to that bus is vulnerable, the error will propagate out through that vulnerable line. Similarly, a high-degree node is more likely to be a critical bus.

As for the optimization property, Proposition 1 indicates that the incorporation of SOCs always improves line robustness, which can be verified visually in Fig. 4 and observed in Fig. 7 for different measurement profiles.

VI. Conclusion

Our vulnerability analysis of graph-structured quadratic sensing is distinguished from previous works by its scalability but also by the strong formal guarantees of a boundary defense against cyberattacks and a localized vulnerability assessment that accounts for network and optimization properties. This study provides a set of notions and tools—the development of graphical mutual incoherence, the boundary defense mechanism, and the analysis of topological and optimization relations to vulnerability—that are applicable to a wide range of graph-structured data. Furthermore, our result offers a scientific foundation for vulnerability-based resource allocation, which, in the case of a power grid, would be based on prioritizing the upgrade of sensing infrastructure for critical locations.

REFERENCES

[1] E. National Academies of Sciences, Medicine *et al.*, *Enhancing the resilience of the Nation's electricity system*. National Academies Press, 2017.

[2] A. Abur and A. G. Exposito, *Power system state estimation: theory and implementation*. CRC press, 2004.

[3] U.S.-Canada Power System Outage Task Force, "Final report on the August 14, 2003 blackout in the United States and Canada: Causes and recommendations," 2004.

[4] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on smart grid state estimation: Attack strategies and countermeasures," in *IEEE International Conference on Smart Grid Communications*, 2010, pp. 220–225.

[5] 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 Processing Magazine*, vol. 29, no. 5, pp. 33–43, 2012.

[6] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security*, vol. 14, no. 1, p. 13, 2011.

[7] W. Wang and Z. Lu, "Cyber security in the smart grid: Survey and challenges," *Computer networks*, vol. 57, no. 5, pp. 1344–1371, 2013.

[8] W. Xu, M. Wang, J.-F. Cai, and A. Tang, "Sparse error correction from nonlinear measurements with applications in bad data detection for power networks," *IEEE Transactions on Signal Processing*, vol. 61, no. 24, pp. 6175–6187, 2013.

[9] M. A. Rahman and H. Mohsenian-Rad, "False data injection attacks against nonlinear state estimation in smart power grids," in *IEEE Power & Energy Society General Meeting*, 2013, pp. 1–5.

[10] J. Liang, L. Sankar, and O. Kosut, "Vulnerability analysis and consequences of false data injection attack on power system state estimation," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3864–3872, 2015.

[11] Y. Weng, M. D. Ilić, Q. Li, and R. Negi, "Convexification of bad data and topology error detection and identification problems in ac electric power systems," *IET Generation, Transmission & Distribution*, vol. 9, no. 16, pp. 2760–2767, 2015.

[12] G. Liang, J. Zhao, F. Luo, S. R. Weller, and Z. Y. Dong, "A review of false data injection attacks against modern power systems," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1630–1638, 2016.

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

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

[15] D. K. Molzahn, I. A. Hiskens *et al.*, "A survey of relaxations and approximations of the power flow equations," *Foundations and Trends® in Electric Energy Systems*, vol. 4, no. 1-2, pp. 1–221, 2019.

[16] G. Wang, G. B. Giannakis, and J. Chen, "Robust and scalable power system state estimation via composite optimization," *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 6137–6147, 2019.

[17] M. Jin, I. Molybog, R. Mohammadi-Ghazi, and J. Lavaei, "Scalable and robust state estimation from abundant but untrusted data," *IEEE Transactions on Smart Grid*, 2019.

[18] G. Dan and H. Sandberg, "Stealth attacks and protection schemes for state estimators in power systems," in *IEEE International Conference on Smart Grid Communications*, 2010, pp. 214–219.

[19] H. Sandberg, A. Teixeira, and K. H. Johansson, "On security indices for state estimators in power networks," in *First Workshop on Secure Control Systems*, 2010.

[20] A. Teixeira, G. Dán, H. Sandberg, and K. H. Johansson, "A cyber security study of a scada energy management system: Stealthy deception attacks on the state estimator," *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 11 271–11 277, 2011.

[21] M. Jin, I. Molybog, R. Mohammadi-Ghazi, and J. Lavaei, "Towards robust and scalable power system state estimation," in *IEEE Conference on Decision and Control*, 2019.

[22] D. Bienstock, *Electrical transmission system cascades and vulnerability: an operations research viewpoint*. SIAM, 2015, vol. 22.

[23] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, p. 717, 2009.

[24] Y. Chi, Y. M. Lu, and Y. Chen, "Nonconvex optimization meets low-rank matrix factorization: An overview," *arXiv preprint arXiv:1809.09573*, 2018.

[25] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Advances in neural information processing systems*, 2009, pp. 2080–2088.

[26] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM (JACM)*, vol. 58, no. 3, pp. 1–37, 2011.

[27] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: a contemporary overview," *IEEE signal processing magazine*, vol. 32, no. 3, pp. 87–109, 2015.

[28] R. Baldick, K. Clements, Z. Pinjo-Dzigal, and P. Davis, "Implementing nonquadratic objective functions for state estimation and bad data rejection," *IEEE Transactions on Power Systems*, vol. 12, no. 1, pp. 376–382, 1997.

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

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

[31] S. Naka, T. Genji, T. Yura, and Y. Fukuyama, "A hybrid particle swarm optimization for distribution state estimation," *IEEE Transactions on Power Systems*, vol. 18, no. 1, pp. 60–68, 2003.

[32] G. Hug and J. A. Giampapa, "Vulnerability assessment of ac state estimation with respect to false data injection cyber-attacks," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1362–1370, 2012.

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

[34] P. J. Huber, *Robust statistics*. John Wiley & Sons, 2004, vol. 523.

[35] J.-J. Fuchs, "Recovery of exact sparse representations in the presence of bounded noise," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3601–3608, 2005.

[36] M. J. Wainwright, "Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso)," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2183–2202, 2009.

[37] R. W. Cottle, J.-S. Pang, and R. E. Stone, *The linear complementarity problem*. SIAM, 2009.

[38] M. C. Ferris and T. S. Munson, "Complementarity problems in GAMS and the PATH solver," *Journal of Economic Dynamics and Control*, vol. 24, no. 2, pp. 165–188, 2000.

[39] A. Birchfield, T. Xu, K. Gegner, K. Shetye, and T. Overbye, "Grid structural characteristics as validation criteria for synthetic networks," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3258–3265, 2017.

[40] F. Alizadeh and D. Goldfarb, "Second-order cone programming," *Mathematical Programming*, vol. 95, no. 1, pp. 3–51, 2003.

APPENDIX

A. Proof of Proposition 1

For the first inequality, notice that the inner minimization of (16) can be written as

$$\begin{aligned} \min_{\alpha\in\mathbb{R},\boldsymbol{h}\in\mathbb{R}^{n_{\checkmark}^{i\to j}}} \quad & \|\boldsymbol{h}\|_\infty \\ \text{s. t.} \quad & \boldsymbol{A}^\top_{\mathcal{M}^{i\to j}_{\mathrm{bd}\checkmark},\mathcal{X}^{i\to j}_{\mathrm{bd}}}\boldsymbol{h} + \boldsymbol{A}^\top_{\mathcal{M}^{i\to j}_{\mathrm{bd}\times},\mathcal{X}^{i\to j}_{\mathrm{bd}}}\boldsymbol{\xi} = \boldsymbol{0}. \end{aligned}$$

Since for any $\boldsymbol{\xi}$, the vector $\hat{\boldsymbol{h}}(\boldsymbol{\xi}) = -\boldsymbol{A}^{\top+}_{\mathcal{M}^{i\to j}_{\mathrm{bd}\checkmark},\mathcal{X}^{i\to j}_{\mathrm{bd}}}\boldsymbol{A}^\top_{\mathcal{M}^{i\to j}_{\mathrm{bd}\times},\mathcal{X}^{i\to j}_{\mathrm{bd}}}\boldsymbol{\xi}$ is a feasible point, and

$$\max_{\boldsymbol{\xi}\in\{-1,+1\}^{n_\times^{i\to j}}} \|\hat{\boldsymbol{h}}(\boldsymbol{\xi})\|_\infty = \rho(\mathcal{M}^{i\to j}_{\mathrm{bd}\times}),$$

we have proved the first inequality.

For the second inequality, for any given $\boldsymbol{\xi}$, let $\hat{\boldsymbol{h}}$ be the optimal solution of the inner minimization of (16) with the property that $\|\hat{\boldsymbol{h}}\|_\infty \le \alpha_{i\to j}$. Then, the tuple $(\alpha^{\mathrm{SOCP}}_{i\to j} = \alpha_{i\to j}, \boldsymbol{\omega} = \boldsymbol{0}, \boldsymbol{h} = \hat{\boldsymbol{h}})$ is a feasible solution for the inner

minimization of (17). Therefore, for any given $\boldsymbol{\xi}$, the optimal solution of (16) is always included in the feasible set of the inner optimization of (17), and we have $\alpha_{i\to j}^{\mathrm{SOCP}}(\boldsymbol{x}) \le \alpha_{i\to j}$.

B. Statement of Lemma 2

Lemma 2. *If the gMI condition is satisfied, then for any $\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bi}}} \in [-1,1]^{|\mathcal{M}_{\mathrm{bi}}|}$, there exist $\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{sf}}\cup\mathcal{M}_{\mathrm{bo}}}$ and $\{\hat{\nu}_\ell, \hat{\boldsymbol{u}}_\ell\}_{\ell\in\mathcal{L}_{\mathrm{at}\cap\mathrm{bi}}\cup\mathcal{L}_{\mathrm{bd}}\cup\mathcal{L}_{\mathrm{sf}}}$ with the properties that $\|\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{sf}}\cup\mathcal{M}_{\mathrm{bo}}}\|_\infty \le 1-\gamma$ and*

$$\boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}}\cup\mathcal{M}_{\mathrm{bo}}}^{\circ\top}\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{sf}}\cup\mathcal{M}_{\mathrm{bo}}} + \boldsymbol{A}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bi}}} + \sum_{\ell\in\mathcal{L}_{\mathrm{at}\cap\mathrm{bi}}\cup\mathcal{L}_{\mathrm{bd}}\cup\mathcal{L}_{\mathrm{sf}}} \hat{\nu}_\ell \boldsymbol{c}_\ell^{\circ} + \boldsymbol{D}_\ell^{\circ\top}\hat{\boldsymbol{u}}_\ell = \boldsymbol{0}. \tag{20}$$

First, we provide the following result, which simplifies the proof of Lemma 2.

Lemma 3. *The gMI $\alpha_{i\to j}^{\mathrm{SOCP}}(\boldsymbol{x})$ coincides with the optimal objective value of the following minimax program:*

$$\max_{\tilde{\boldsymbol{\xi}}\in[-1,+1]^{n_\times^{i\to j}}} \quad \min_{\tilde\alpha\in\mathbb{R},\boldsymbol{\nu}\in\mathbb{R}^{n_{\mathcal{L}}^{i\to j}},\tilde{\boldsymbol{h}}\in\mathbb{R}^{n_\checkmark^{i\to j}}} \quad \tilde\alpha \tag{21a}$$

$$\text{s. t.} \quad \boldsymbol{A}_{\mathcal{M}_{\mathrm{bd}\checkmark}^{i\to j},\mathcal{X}_{\mathrm{bd}}^{i\to j}}^{\top}\tilde{\boldsymbol{h}} + \boldsymbol{A}_{\mathcal{M}_{\mathrm{bd}\times}^{i\to j},\mathcal{X}_{\mathrm{bd}}^{i\to j}}^{\top}\tilde{\boldsymbol{\xi}} + \sum_{\ell\in\mathcal{L}_{\mathrm{bd}}^{i\to j}} \nu_\ell \boldsymbol{c}_\ell + \boldsymbol{D}_\ell^{\top}\boldsymbol{u}_\ell = \boldsymbol{0} \tag{21b}$$

$$\nu_\ell \ge \|\boldsymbol{u}_\ell\|_2, \qquad \forall \ell\in\mathcal{L}_{\mathrm{bd}}^{i\to j} \tag{21c}$$

$$\nu_\ell \boldsymbol{c}_\ell^{\top}\boldsymbol{x} + \boldsymbol{u}_\ell^{\top}\boldsymbol{D}_\ell\boldsymbol{x} = 0, \qquad \forall \ell\in\mathcal{L}_{\mathrm{bd}}^{i\to j} \tag{21d}$$

$$\|\tilde{\boldsymbol{h}}\|_\infty \le \tilde\alpha \tag{21e}$$

Proof of Lemma 3. The equivalence between optimizing over $[-1,+1]^{n_\times^{i\to j}}$ and $\{-1,+1\}^{n_\times^{i\to j}}$ is due to the convexity of the feasibility region given $\boldsymbol{x}\in\mathcal{K}$ and $\tilde{\boldsymbol{\xi}}$. Since $\boldsymbol{x}$ satisfies the primal feasibility, which can be expressed as in (6), a standard result (c.f., [40, Lemma 15]) in analogy to linear programming indicates that (21d) is equivalent to:

$$\nu_\ell \boldsymbol{D}_\ell\boldsymbol{x} + \boldsymbol{c}_\ell^{\top}\boldsymbol{x}\boldsymbol{u}_\ell = \boldsymbol{0}, \qquad \forall \ell\in\mathcal{L}_{\mathrm{bd}}^{i\to j},$$

which indicates that $\nu_\ell = \omega_\ell \boldsymbol{c}_\ell^{\top}\boldsymbol{x}$ and $\boldsymbol{u}_\ell = -\omega_\ell \boldsymbol{D}_\ell\boldsymbol{x}$ for $\omega_\ell \ge 0$ and $\ell\in\mathcal{L}_{\mathrm{bd}}^{i\to j}$. It can be verified that this also satisfies the SOCs (21c). By the definition of $\boldsymbol{T}_\ell = \boldsymbol{c}_\ell\boldsymbol{c}_\ell^{\top} - \boldsymbol{D}_\ell^{\top}\boldsymbol{D}_\ell$, the equivalence to (17) is established. □

Proof of Lemma 4. First, we show that a sufficient condition for the existence of $\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{sf}}\cup\mathcal{M}_{\mathrm{bo}}} = \begin{bmatrix}\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{sf}}}^{\top} & \hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bo}}}^{\top}\end{bmatrix}^{\top}$ such that $\|\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{sf}}\cup\mathcal{M}_{\mathrm{bo}}}\|_\infty \le 1-\gamma$ and (20) are satisfied is that for any $\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bi}}}$ there exists an $\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bo}}}$ and $\{\hat{\nu}_\ell, \hat{\boldsymbol{u}}_\ell\}_{\ell\in\mathcal{L}_{\mathrm{at}\cap\mathrm{bi}}}$ such that $\|\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bo}}}\|_\infty \le 1-\gamma$ and

$$\boldsymbol{A}_{\mathcal{M}_{\mathrm{bo}},\mathcal{X}_{\mathrm{bd}}}^{\top}\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bo}}} + \boldsymbol{A}_{\mathcal{M}_{\mathrm{bi}},\mathcal{X}_{\mathrm{bd}}}^{\top}\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bi}}} + \sum_{\ell\in\mathcal{L}_{\mathrm{at}\cap\mathrm{bi}}} \nu_\ell \boldsymbol{c}_\ell + \boldsymbol{D}_\ell^{\top}\boldsymbol{u}_\ell = \boldsymbol{0}. \tag{22}$$

This is immediate by simply choosing $\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{sf}}\cup\mathcal{M}_{\mathrm{bo}}} = \begin{bmatrix}\boldsymbol{0}^{\top} & \hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bo}}}^{\top}\end{bmatrix}^{\top}$ and $\{\hat{\nu}_\ell = 0, \hat{\boldsymbol{u}}_\ell = \boldsymbol{0}\}_{\ell\in\mathcal{L}_{\mathrm{bd}}\cup\mathcal{L}_{\mathrm{sf}}}$. In what follows, we prove (22) by induction.

The induction rule is as follows: we start by arbitrarily choosing one line $\{i,j\}\in\mathcal{L}_{\mathrm{at}\cap\mathrm{bi}}$ to initialize $\mathcal{L}_{\mathrm{at}\cap\mathrm{bi}}^{(1)}$, where $i\in\mathcal{N}_{\mathrm{at}}$ and $j\in\mathcal{N}_{\mathrm{bi}}$, and initialize the measurement set $\mathcal{M}_{\mathrm{bo}}^{(1)} \coloneqq \mathcal{M}_{\mathrm{bd}\checkmark}^{i\to j}$, $\mathcal{M}_{\mathrm{bi}}^{(1)} \coloneqq \mathcal{M}_{\mathrm{bd}\times}^{i\to j}$ and the variable set $\mathcal{X}_{\mathrm{bd}}^{(1)} \coloneqq \mathcal{X}_{\mathrm{bd}}^{i\to j}$. For each step k, we add a new line $\{f,t\}\in\mathcal{L}_{\mathrm{at}\cap\mathrm{bi}}$ to $\mathcal{L}_{\mathrm{at}\cap\mathrm{bi}}^{(k)}$ and the associated measurements and variables to $\mathcal{M}_{\mathrm{bo}}^{(k)}$, $\mathcal{M}_{\mathrm{bi}}^{(k)}$ and $\mathcal{X}_{\mathrm{bd}}^{(k)}$, respectively. We also construct $\boldsymbol{c}_\ell^{(k)}$ and $\boldsymbol{D}_\ell^{(k)}$ with entries and columns corresponding to $\mathcal{X}_{\mathrm{bd}}^{(k)}$ for all $\ell\in\mathcal{L}_{\mathrm{at}\cap\mathrm{bi}}^{(k)}$, respectively. In each step, we check whether there exist $\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bo}}^{(k)}}$ and $\{\hat{\nu}_\ell^{(k)}, \hat{\boldsymbol{u}}_\ell^{(k)}\}_{\ell\in\mathcal{L}_{\mathrm{at}\cap\mathrm{bi}}^{(k)}}$ such that $\|\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bo}}^{(k)}}\|_\infty \le 1-\gamma$ and

$$\boldsymbol{A}_{\mathcal{M}_{\mathrm{bo}}^{(k)},\mathcal{X}_{\mathrm{bd}}^{(k)}}^{\top}\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bo}}^{(k)}} + \boldsymbol{A}_{\mathcal{M}_{\mathrm{bi}}^{(k)},\mathcal{X}_{\mathrm{bd}}^{(k)}}^{\top}\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bi}}^{(k)}} + \sum_{\ell\in\mathcal{L}_{\mathrm{at}\cap\mathrm{bi}}^{(k)}} \hat{\nu}_\ell^{(k)}\boldsymbol{c}_\ell^{(k)} + \boldsymbol{D}_\ell^{(k)\top}\hat{\boldsymbol{u}}_\ell^{(k)} = \boldsymbol{0}. \tag{23}$$

The base case for $k=1$ follows directly from the condition that $\alpha_{i\to j}^{\mathrm{SOCP}} \le 1-\gamma$. For any $k\ge 1$, let $\{f,t\}\in\mathcal{L}_{\mathrm{at}\cap\mathrm{bi}}$ denote the line to be added and consider two possible cases:

1) the new line does not share any nodes with the lines that have been already added; or
2) the new line shares the attacked node f with one (or more) of the lines already added

For each case, there are also three events that may occur:

a) one or more of the nodes in $\mathcal{N}_{\mathrm{bo}}^{f\to t}$ are connected to one or more of the nodes in the inner boundaries of lines that have already been added
b) one or more of the nodes in the outer boundary of the lines that have already been added are connected to t
c) none of the above

To prove the claim, we need to consider all the combinations between the two cases and the three events. Fortunately, they can be reduced to two typical scenarios, where the proofs can be directly applied. We consider these scenarios now.

The first typical scenario applies to combinations 1c and 2c, where $\mathcal{M}_{\mathrm{bo}}^{(k)} = \mathcal{M}_{\mathrm{bo}}^{(k-1)} \cup \mathcal{M}_{\mathrm{bd}\checkmark}^{f\to t}$, $\mathcal{M}_{\mathrm{bi}}^{(k)} = \mathcal{M}_{\mathrm{bi}}^{(k-1)} \cup \mathcal{M}_{\mathrm{bd}\times}^{f\to t}$, $\mathcal{X}_{\mathrm{bd}}^{(k)} = \mathcal{X}_{\mathrm{bd}}^{(k-1)} \cup \mathcal{X}_{\mathrm{bd}}^{f\to t}$, $\mathcal{M}_{\mathrm{bo}}^{(k-1)} \cap \mathcal{M}_{\mathrm{bd}\checkmark}^{f\to t} = \emptyset$, $\mathcal{M}_{\mathrm{bi}}^{(k-1)} \cap \mathcal{M}_{\mathrm{bd}\times}^{f\to t} = \emptyset$, and $\mathcal{X}_{\mathrm{bd}}^{(k-1)} \cap \mathcal{X}_{\mathrm{bd}}^{f\to t} = \emptyset$. Therefore, for any given $\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bi}}^{(k)}} = \begin{bmatrix}\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bi}}^{(k-1)}}^{\top} & \tilde{\boldsymbol{\xi}}^{\top}\end{bmatrix}^{\top}$ with $\|\tilde{\boldsymbol{\xi}}\|_\infty \le 1$, we can always find $\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bo}}^{(k)}} = \begin{bmatrix}\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bo}}^{(k-1)}}^{\top} & \tilde{\boldsymbol{h}}^{\top}\end{bmatrix}^{\top}$ and $\{\hat{\nu}_\ell^{(k)}, \hat{\boldsymbol{u}}_\ell^{(k)}\}_{\ell\in\mathcal{L}_{\mathrm{at}\cap\mathrm{bi}}^{(k)}}$ that satisfy (22), where $\tilde{\boldsymbol{h}}$ and $\{\hat{\nu}_\ell^{(k)}, \hat{\boldsymbol{u}}_\ell^{(k)}\}_{\ell\in\mathcal{L}_{\mathrm{at}\cap\mathrm{bi}}^{(k)}}$ are given by (21).

The second scenario applies to the remaining combinations. Let $\tilde{\mathcal{N}}_{\mathrm{bo}}$ be the set of nodes in the outer boundary shared by the new line in $\mathcal{B}_{\mathrm{bo}}^{f\to t}$ and those of the lines that have been added. Then, we have $\mathcal{M}_{\mathrm{bo}}^{(k)} = \mathcal{M}_{\mathrm{bo}}^{(k-1)} \cup \mathcal{M}_{\mathrm{bd}\checkmark}^{f\to t}$, $\mathcal{M}_{\mathrm{bi}}^{(k)} = \mathcal{M}_{\mathrm{bi}}^{(k-1)} \cup \mathcal{M}_{\mathrm{bd}\times}^{f\to t}$, $\mathcal{X}_{\mathrm{bd}}^{(k)} = \mathcal{X}_{\mathrm{bd}}^{(k-1)} \cup \mathcal{X}_{\mathrm{bd}}^{f\to t}$, where $\mathcal{M}_{\mathrm{bo}}^{(k-1)} \cap \mathcal{M}_{\mathrm{bd}\checkmark}^{f\to t}$ is the set of measurements that only depend on nodal variables of $\tilde{\mathcal{N}}_{\mathrm{bo}}$, $\mathcal{M}_{\mathrm{bi}}^{(k-1)} \cap \mathcal{M}_{\mathrm{bd}\times}^{f\to t} = \emptyset$, and $\mathcal{X}_{\mathrm{bd}}^{(k-1)} \cap \mathcal{X}_{\mathrm{bd}}^{f\to t}$ is the set of nodal variables of $\tilde{\mathcal{N}}_{\mathrm{bo}}$. For any given $\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bi}}^{(k)}}$ and $\tilde{\boldsymbol{\xi}}$, we can always find $\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bo}}^{(k)}}$ and $\tilde{\boldsymbol{h}}$, where $\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bo}}^{(k)}}$ is given by (23) and $\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bo}}^{(k)}}$ is given by (16). Let $\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bo}}^{(k)}}$ be further divided into the parts corresponding to the voltage magnitude measurements

(if available) of nodes in $\tilde{\mathcal{N}}_{\rm bo}$ (i.e. $\left[\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bo}^{(k)}}\right]_{\tilde{\mathcal{N}}_{\rm bo}}$) and the rest (i.e. $\left[\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bo}^{(k)}}\right]_{\tilde{\mathcal{N}}_{\rm bo}^c}$); similarly, let $\hat{\boldsymbol{h}}$ be further divided into $\left[\hat{\boldsymbol{h}}\right]_{\tilde{\mathcal{N}}_{\rm bo}}$ and the rest $\left[\hat{\boldsymbol{h}}\right]_{\tilde{\mathcal{N}}_{\rm bo}^c}$. Then, by setting $\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bo}^{(k+1)}} = \left[\left[\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bo}^{(k)}}\right]_{\tilde{\mathcal{N}}_{\rm bo}^c}^\top \quad \frac{1}{\deg(\tilde{\mathcal{N}}_{\rm bo})}\circ\left(\left[\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bo}^{(k)}}\right]_{\tilde{\mathcal{N}}_{\rm bo}}+\left[\hat{\boldsymbol{h}}\right]_{\tilde{\mathcal{N}}_{\rm bo}}\right)^\top \quad \left[\hat{\boldsymbol{h}}\right]_{\tilde{\mathcal{N}}_{\rm bo}^c}^\top\right]^\top$ where $\deg(\tilde{\mathcal{N}}_{\rm bo})$ is the connectivity degree for each node in $\tilde{\mathcal{N}}_{\rm bo}$, and $\circ$ indicates the Hadamard (element-wise) product, we can satisfy (23) for any given $\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}^{(k+1)}} = \left[\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}^{(k)}}^\top \quad \hat{\boldsymbol{\xi}}^\top\right]^\top$. Moreover, by construction, we have $\|\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bo}^{(k+1)}}\|_\infty \le 1-\gamma$ for all k. This completes the induction proof.

C. Proof of Theorem 1

The following lemma indicates a connection between Huber's loss and a mixed ℓ_1,ℓ_2 loss.

Lemma 4. *Let $\hat{\boldsymbol{x}}_1$ be the solution to* (3) *and let $(\hat{\boldsymbol{x}}_2,\hat{\boldsymbol{b}}_2)$ be the solution to the following problem:*

$$\min_{\boldsymbol{b}\in\mathbb{R}^{n_m},\boldsymbol{x}\in\mathbb{R}^{n_x}} \quad \tfrac{1}{2n_m}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2^2+\lambda\|\boldsymbol{b}\|_1 \tag{24a}$$

$$\text{s.t.} \quad \boldsymbol{c}_\ell^\top\boldsymbol{x}\ge\|\boldsymbol{D}_\ell\boldsymbol{x}\|_2\,, \qquad \forall \ell\in\mathcal{L}. \tag{24b}$$

Then, we have $\hat{\boldsymbol{x}}_1=\hat{\boldsymbol{x}}_2$, and the i-th component of $\hat{\boldsymbol{b}}_2$ is given by:

$$[\hat{b}_2]_i = \operatorname{sign}(y_i-\boldsymbol{a}_i^\top\hat{\boldsymbol{x}}_2)\max\left(0,\left|y_i-\boldsymbol{a}_i^\top\hat{\boldsymbol{x}}_2\right|-\lambda\right),$$

where $\operatorname{sign}(y)$ *denotes the sign of y.*

Proof. Given a feasible $\boldsymbol{x}$, the inner optimization can be decomposed into a series of smaller optimization problems

$$\min_{b_i}\tfrac{1}{2n_m}(y_i-\boldsymbol{a}_i^\top\boldsymbol{x}-b_i)^2+\lambda|b_i|, \tag{25}$$

for $i\in[n_m]$, which has a closed-form solution

$$b_i^* = \operatorname{sign}(y_i-\boldsymbol{a}_i^\top\boldsymbol{x})\max\left(0,\left|y_i-\boldsymbol{a}_i^\top\boldsymbol{x}\right|-\lambda\right), \tag{26}$$

Now, we substitute (26) into the outer minimization to see the equivalence to minimization of a Huber's loss. Furthermore, for the solution $\hat{\boldsymbol{x}}_2$, the optimal $\boldsymbol{b}$ of (24) is given by (26). □

Due to the above connection, hereafter we will analyze the equivalent problem (24). For an arbitrary set of attacked measurements $\mathcal{M}_{\rm at}$, the corresponding boundary $\mathcal{M}_{\rm bd}:=\mathcal{M}_{\rm bi}\cup\mathcal{M}_{\rm bo}$ and safe measurements $\mathcal{M}_{\rm sf}$, we design a primal-dual witness (PDW) process as follows:

1) Set $\hat{\boldsymbol{b}}_{\mathcal{M}_{\rm sf}}=\boldsymbol{0}$ and $\hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bo}}=\boldsymbol{0}$;

2) Determine $\hat{\boldsymbol{x}} = \left[\hat{\boldsymbol{x}}_{\rm sf}^\top \quad \hat{\boldsymbol{x}}_{\rm bd}^\top \quad \hat{\boldsymbol{x}}_{\rm at}^\top\right]^\top$ and $\hat{\boldsymbol{b}} = \left[\boldsymbol{0}^\top \quad \boldsymbol{0}^\top \quad \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}}^\top \quad \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm at}}^\top\right]^\top$ by solving the following program:

$$\min_{\boldsymbol{b}_{\mathcal{M}_{\rm bi}},\boldsymbol{b}_{\mathcal{M}_{\rm at}},\boldsymbol{x}} \quad \frac{1}{2n_m}\left\|\begin{bmatrix}\boldsymbol{y}_{\mathcal{M}_{\rm sf}}\\ \boldsymbol{y}_{\mathcal{M}_{\rm bo}}\\ \boldsymbol{y}_{\mathcal{M}_{\rm bi}}\\ \boldsymbol{y}_{\mathcal{M}_{\rm at}}\end{bmatrix}-\boldsymbol{A}\begin{bmatrix}\boldsymbol{x}_{\rm sf}\\ \boldsymbol{x}_{\rm bd}\\ \boldsymbol{x}_{\rm at}\end{bmatrix}-\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\\ \boldsymbol{b}_{\mathcal{M}_{\rm bi}}\\ \boldsymbol{b}_{\mathcal{M}_{\rm at}}\end{bmatrix}\right\|_2^2 +\lambda\left\|\begin{bmatrix}\boldsymbol{b}_{\mathcal{M}_{\rm bi}}\\ \boldsymbol{b}_{\mathcal{M}_{\rm at}}\end{bmatrix}\right\|_1, \tag{27a}$$

$$\text{s.t.} \quad \boldsymbol{c}_\ell^\top\boldsymbol{x}\ge\|\boldsymbol{D}_\ell\boldsymbol{x}\|_2\,, \qquad \forall\ell\in\mathcal{L}, \tag{27b}$$

and $\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}}\in\partial\|\hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}}\|_1$ and $\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm at}}\in\partial\|\hat{\boldsymbol{b}}_{\mathcal{M}_{\rm at}}\|_1$ satisfying the optimality conditions

$$-\frac{1}{n_m}(\boldsymbol{y}_{\mathcal{M}_{\rm at}}-\boldsymbol{A}_{\mathcal{M}_{\rm at},\mathcal{X}_{\rm at}}\hat{\boldsymbol{x}}_{\rm at}-\hat{\boldsymbol{b}}_{\mathcal{M}_{\rm at}})+\lambda\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm at}}=\boldsymbol{0}, \tag{28a}$$

$$-\frac{1}{n_m}\Big(\boldsymbol{y}_{\mathcal{M}_{\rm bi}}-\boldsymbol{A}_{\mathcal{M}_{\rm bi},\mathcal{X}_{\rm bd}}\hat{\boldsymbol{x}}_{\rm bd}-\boldsymbol{A}_{\mathcal{M}_{\rm bi},\mathcal{X}_{\rm at}}\hat{\boldsymbol{x}}_{\rm at} -\hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}}\Big)+\lambda\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}}=\boldsymbol{0}. \tag{28b}$$

3) Solve $(\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}},\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bo}})$ via the zero-subgradient equation:

$$-\frac{1}{n_m}\left(\boldsymbol{y}-\boldsymbol{A}\hat{\boldsymbol{x}}-\hat{\boldsymbol{b}}\right)+\lambda\hat{\boldsymbol{h}}=\boldsymbol{0}, \tag{29}$$

where $\hat{\boldsymbol{x}} = \left[\hat{\boldsymbol{x}}_{\mathcal{B}_{\rm sf}}^\top \quad \hat{\boldsymbol{x}}_{\mathcal{B}_{\rm bd}}^\top \quad \hat{\boldsymbol{x}}_{\mathcal{B}_{\rm at}}^\top\right]^\top$ and $\hat{\boldsymbol{b}} = \left[\boldsymbol{0}^\top \quad \boldsymbol{0}^\top \quad \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}}^\top \quad \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm at}}^\top\right]^\top$ are the solutions obtained in (27), and $\hat{\boldsymbol{h}} = \left[\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}}^\top \quad \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bo}}^\top \quad \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}}^\top \quad \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm at}}^\top\right]^\top$ where $(\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}},\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm at}})$ are given in (28). We check whether strict feasibility conditions $\|\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}}\|_\infty<1$ and $\|\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bo}}\|_\infty<1$ hold.

The next lemma relates the PDW procedure to the solution properties of (24).

Lemma 5. *If the PDW procedure succeeds, then $(\hat{\boldsymbol{x}},\hat{\boldsymbol{b}})$ that is optimal for* (27) *is also optimal for* (24)*. Furthermore, for any optimal solution $(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{b}})$, it holds that* $\operatorname{supp}(\tilde{\boldsymbol{b}})\subseteq\mathcal{M}_{\rm bi}\cup\mathcal{M}_{\rm at}$.

Proof. It can be checked that if PDW succeeds, then the optimality conditions of (24) corresponding to $(\hat{\boldsymbol{x}},\hat{\boldsymbol{b}})$ are satisfied, which certify the optimality.

The subgradient $\hat{\boldsymbol{h}}$ satisfies $\|\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}}\|_\infty<1$, $\|\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bo}}\|_\infty<1$ and $\left\langle\hat{\boldsymbol{h}},\hat{\boldsymbol{b}}\right\rangle=\|\hat{\boldsymbol{b}}\|_1$. Now, let $(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{b}})$ be any other optimal, and let $F(\boldsymbol{x},\boldsymbol{b})=\frac{1}{2n_m}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2^2$; then,

$$F(\hat{\boldsymbol{x}},\hat{\boldsymbol{b}})+\lambda\left\langle\hat{\boldsymbol{h}},\hat{\boldsymbol{b}}\right\rangle=F(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{b}})+\lambda\|\tilde{\boldsymbol{b}}\|_1,$$

and hence,

$$F(\hat{\boldsymbol{x}},\hat{\boldsymbol{b}})+\lambda\left\langle\hat{\boldsymbol{h}},\hat{\boldsymbol{b}}-\tilde{\boldsymbol{b}}\right\rangle=F(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{b}})+\lambda\left(\|\tilde{\boldsymbol{b}}\|_1-\left\langle\hat{\boldsymbol{h}},\tilde{\boldsymbol{b}}\right\rangle\right).$$

By the stationarity condition of KKT, we have $\lambda\hat{\boldsymbol{h}}=-\nabla_b F(\hat{\boldsymbol{x}},\hat{\boldsymbol{b}})=\frac{1}{n_m}(\boldsymbol{y}-\boldsymbol{A}\hat{\boldsymbol{x}}-\hat{\boldsymbol{b}})$, which implies that

$$\begin{aligned}&F(\hat{\boldsymbol{x}},\hat{\boldsymbol{b}})-\left\langle\nabla_b F(\hat{\boldsymbol{x}},\hat{\boldsymbol{b}}),\hat{\boldsymbol{b}}-\tilde{\boldsymbol{b}}\right\rangle-F(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{b}})\\ &=\lambda\left(\|\tilde{\boldsymbol{b}}\|_1-\left\langle\hat{\boldsymbol{h}},\tilde{\boldsymbol{b}}\right\rangle\right)\le 0\end{aligned}$$

due to convexity. We thus have $\|\tilde{\boldsymbol{b}}\|_1\le\left\langle\hat{\boldsymbol{h}},\tilde{\boldsymbol{b}}\right\rangle$. Since by Holder's inequality, we also have $\left\langle\hat{\boldsymbol{h}},\tilde{\boldsymbol{b}}\right\rangle\le\|\hat{\boldsymbol{h}}\|_\infty\|\tilde{\boldsymbol{b}}\|_1$, and $\|\hat{\boldsymbol{h}}\|_\infty\le 1$, it holds that $\|\tilde{\boldsymbol{b}}\|_1=\left\langle\hat{\boldsymbol{h}},\tilde{\boldsymbol{b}}\right\rangle$. Since by the success of PDW, $\|\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}}\|_\infty<1$, $\|\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bo}}\|_\infty<1$, we have $\tilde{\boldsymbol{b}}_j=0$ for $j\in\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bo}$. □

In the following, we denote $\boldsymbol{x}^\circ$ as the subvector of $\boldsymbol{x}$ that removes the entries corresponding to $\boldsymbol{x}_{\rm at}$, $\boldsymbol{w}^\circ$ as the subvector of $\boldsymbol{w}$ that removes entries corresponding to $\mathcal{M}_{\rm at}$, and $\boldsymbol{I}^\circ$ as the identity matrix of size $n_m-|\mathcal{M}_{\rm at}|$.

Proof of Theorem 1

Part 1): By the construction of PDW, we have $\hat{\boldsymbol{b}}_{\mathcal{M}_{\rm sf}} = \boldsymbol{b}_{\natural\mathcal{M}_{\rm sf}} = \boldsymbol{0}$ and $\hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bo}} = \boldsymbol{b}_{\natural\mathcal{M}_{\rm bo}} = \boldsymbol{0}$. The optimal solution $\hat{\boldsymbol{x}}_{\rm at}$ and $\hat{\boldsymbol{b}}_{\mathcal{M}_{\rm at}}$ of (27) can take any value as long as the nonbinding SOC constraint condition is satisfied. Thus, for any given $\hat{\boldsymbol{x}}_{\rm at}$ and $\hat{\boldsymbol{b}}_{\mathcal{M}_{\rm at}}$, we can fix $\boldsymbol{x}_{\rm at}$ and $\boldsymbol{b}_{\mathcal{M}_{\rm at}}$ in (27) and solve the following smaller program:

$$\min_{\boldsymbol{b}_{\mathcal{M}_{\rm bi}},\boldsymbol{x}_{\rm sf},\boldsymbol{x}_{\rm bd}} \quad \frac{1}{2n_m}\Bigg\| \underbrace{\begin{bmatrix}\boldsymbol{y}_{\mathcal{M}_{\rm sf}}\\ \boldsymbol{y}_{\mathcal{M}_{\rm bo}}\\ \boldsymbol{z}_{\mathcal{M}_{\rm bi}}\end{bmatrix}}_{\boldsymbol{z}^\circ} - \boldsymbol{A}^\circ \underbrace{\begin{bmatrix}\boldsymbol{x}_{\rm sf}\\ \boldsymbol{x}_{\rm bd}\end{bmatrix}}_{\boldsymbol{x}^\circ} - \begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\\ \boldsymbol{b}_{\mathcal{M}_{\rm bi}}\end{bmatrix}\Bigg\|_2^2 + \lambda\left\|\boldsymbol{b}_{\mathcal{M}_{\rm bi}}\right\|_1, \tag{30a}$$

$$\text{s. t.} \quad \boldsymbol{c}_\ell^\top \boldsymbol{x} \geq \left\|\boldsymbol{D}_\ell \boldsymbol{x}\right\|_2, \qquad \forall \ell \in \mathcal{L}\setminus\mathcal{L}_{\rm at}, \tag{30b}$$

where $\boldsymbol{z}_{\mathcal{M}_{\rm bi}} = \boldsymbol{y}_{\mathcal{M}_{\rm bi}} - \boldsymbol{A}_{\mathcal{M}_{\rm bi},\mathcal{X}_{\rm at}}\hat{\boldsymbol{x}}_{\rm at} = \boldsymbol{A}_{\mathcal{M}_{\rm bi},\mathcal{X}_{\rm bd}}\boldsymbol{x}_{\natural{\rm bd}} + \tilde{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}}$ and $\tilde{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}} = \boldsymbol{A}_{\mathcal{M}_{\rm bi},\mathcal{X}_{\rm at}}(\boldsymbol{x}_{\natural{\rm at}} - \hat{\boldsymbol{x}}_{\rm at})$. Thus, we have $\boldsymbol{z}^\circ = \boldsymbol{A}^\circ \boldsymbol{x}_\natural^\circ + \boldsymbol{w}_\natural^\circ + \boldsymbol{I}_{\mathcal{M}_{\rm bi}}^{\circ\top}\tilde{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}}$. The solution $(\hat{\boldsymbol{x}}_{\rm sf}, \hat{\boldsymbol{x}}_{\rm bd}, \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}})$ of (30) is unique due to the lower eigenvalue condition. By the KKT condition, (28) is satisfied. We combine (28) and (29) and partition the relation into equations indexed by $\mathcal{M}_{\rm bi}$:

$$\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}} = \frac{1}{n_m\lambda}\boldsymbol{I}_{\mathcal{M}_{\rm bi}}^\circ\left(\begin{bmatrix}\boldsymbol{A}^\circ & \boldsymbol{I}_{\mathcal{M}_{\rm bi}}^{\circ\top}\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_\natural^\circ - \hat{\boldsymbol{x}}^\circ\\ \tilde{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}} - \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}}\end{bmatrix} + \boldsymbol{w}_\natural^\circ\right), \tag{31}$$

as well as those indexed by $\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bo}$, which can be solved for $\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bo}} = \begin{bmatrix}\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}}^\top & \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bo}}^\top\end{bmatrix}^\top$:

$$\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bo}} = \frac{1}{n_m\lambda}\boldsymbol{I}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bo}}^\circ\left(\boldsymbol{A}^\circ(\boldsymbol{x}_\natural^\circ - \hat{\boldsymbol{x}}^\circ) + \boldsymbol{w}_\natural^\circ\right). \tag{32}$$

Since $\hat{\boldsymbol{x}}^\circ$ is the optimal solution of (30), it satisfies the optimality condition:

$$\frac{1}{n_m}\boldsymbol{A}^{\circ\top}\left(\begin{bmatrix}\boldsymbol{A}^\circ & \boldsymbol{I}_{\mathcal{M}_{\rm bi}}^{\circ\top}\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_\natural^\circ - \hat{\boldsymbol{x}}^\circ\\ \tilde{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}} - \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}}\end{bmatrix} + \boldsymbol{w}_\natural^\circ\right) + \sum_{\ell\in\mathcal{L}_{\rm at\cap bi}\cup\mathcal{L}_{\rm bd}\cup\mathcal{L}_{\rm sf}} \hat{\nu}_\ell \boldsymbol{c}_\ell^\circ + \boldsymbol{D}_\ell^{\circ\top}\hat{\boldsymbol{u}}_\ell = \boldsymbol{0} \tag{33}$$

Combining (31), (32) and (33) and after some elementary operations, it yields that

$$\lambda\boldsymbol{A}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bo}}^{\circ\top}\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bo}} + \lambda\boldsymbol{A}_{\mathcal{M}_{\rm bi}}^{\circ\top}\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}} + \sum_{\ell\in\mathcal{L}_{\rm at\cap bi}\cup\mathcal{L}_{\rm bd}\cup\mathcal{L}_{\rm sf}} \hat{\nu}_\ell \boldsymbol{c}_\ell^\circ + \boldsymbol{D}_\ell^{\circ\top}\hat{\boldsymbol{u}}_\ell = \boldsymbol{0}. \tag{34}$$

By Lemma 2, for any $\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}} \in \partial\|\hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}}\|_1$, there always exist $\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bo}}$ and $\{\hat{\nu}_\ell, \hat{\boldsymbol{u}}_\ell\}_{\mathcal{L}_{\rm at\cap bi}\cup\mathcal{L}_{\rm bd}\cup\mathcal{L}_{\rm sf}}$ such that $\|\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bo}}\|_\infty < 1$. Thus, the strict feasibility condition of PDW is satisfied deterministically. Since PDW is successful, we can conclude the first part based on Lemma 5.

Part 2): Let $\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^\circ = \begin{bmatrix}\boldsymbol{A}^\circ & \boldsymbol{I}_{\mathcal{M}_{\rm bi}}^{\circ\top}\end{bmatrix}$ and $\hat{\boldsymbol{h}}^\circ = \begin{bmatrix}\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bo}}^\top & \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}}^\top\end{bmatrix}^\top$. By the lower eigenvalue condition, we can solve (31), (33) and (34):

$$\begin{aligned}\boldsymbol{\Delta} &:= \begin{bmatrix}\boldsymbol{x}_\natural^\circ - \hat{\boldsymbol{x}}^\circ\\ \tilde{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}} - \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}}\end{bmatrix}\\ &= -\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ+}\boldsymbol{w}_\natural^\circ + n_m\lambda(\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ\top}\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^\circ)^{-1}\begin{bmatrix}\boldsymbol{A}^{\circ\top}\hat{\boldsymbol{h}}^\circ\\ \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}}\end{bmatrix}\\ &= -\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ+}\boldsymbol{w}_\natural^\circ + n_m\lambda(\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ\top}\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^\circ)^{-1}\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ\top}\hat{\boldsymbol{h}},\end{aligned} \tag{35}$$

Then, we can bound the estimation error $\boldsymbol{\Delta}$ in (35). First, we bound the infinity norm of $\tilde{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}} - \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}} = \boldsymbol{I}_b\boldsymbol{\Delta}$. By triangle inequality,

$$\|\boldsymbol{I}_b\boldsymbol{\Delta}\|_\infty \leq \|\boldsymbol{I}_b\boldsymbol{Q}^{\circ+}\boldsymbol{w}_\natural^\circ\|_\infty + n_m\lambda\|\boldsymbol{I}_b\boldsymbol{Q}^{\circ+}\|_\infty. \tag{36}$$

Since the second term is deterministic, we will now bound the first term. By the normalized measurement condition and the lower eigenvalue condition, each entry of $\boldsymbol{Q}^{\circ+}\boldsymbol{w}_\natural^\circ$ is zero-mean sub-Gaussian with parameter at most

$$\sigma^2\|(\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ\top}\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^\circ)^{-1}\|_2 \leq \frac{\sigma^2}{C_{\min}}. \tag{37}$$

Thus, by the union bound, we have

$$\mathbb{P}\left(\|\boldsymbol{I}_b\boldsymbol{Q}^{\circ+}\boldsymbol{w}_\natural^\circ\|_\infty > t\right) \leq 2\exp\left(-\frac{C_{\min}t^2}{2\sigma^2} + \log|\mathcal{M}_{\rm bi}|\right). \tag{38}$$

Then, set $t = \frac{n_m\lambda}{2\sqrt{C_{\min}}}$, and note that by our choice of λ, we have $\frac{C_{\min}t^2}{2\sigma^2} > \log|\mathcal{M}_{\rm bi}|$. Thus, we conclude that

$$\begin{aligned}&\|\tilde{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}} - \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}}\|_\infty\\ &\leq n_m\lambda\left(\frac{1}{2\sqrt{C_{\min}}} + \|\boldsymbol{I}_b(\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ\top}\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^\circ)^{-1}\boldsymbol{I}_b^\top\|_\infty\right)\end{aligned}$$

with probability greater than $1 - 2\exp(-c_2 n_m^2\lambda^2)$. This indicates that all bad data entries greater than

$$g(\lambda) = n_m\lambda\left(\frac{1}{2\sqrt{C_{\min}}} + \|\boldsymbol{I}_b(\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ\top}\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^\circ)^{-1}\boldsymbol{I}_b^\top\|_\infty\right) \tag{39}$$

will be detected by $\hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}}$.

Part 3): From (35), we can upper bound the ℓ_2 norm of the signal error $\boldsymbol{x}_\natural^\circ - \hat{\boldsymbol{x}}^\circ = \boldsymbol{I}_x\boldsymbol{\Delta}$ by

$$\|\boldsymbol{I}_x\boldsymbol{Q}^{\circ+}\boldsymbol{w}_\natural^\circ\|_2 + n_m\lambda\|\boldsymbol{I}_x(\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ\top}\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^\circ)^{-1}\boldsymbol{I}_b^\top\|_{\infty,2}.$$

For the first term, by the application of standard sub-gaussian concentration,

$$\mathbb{P}\left(\|\boldsymbol{I}_x\boldsymbol{Q}^{\circ+}\boldsymbol{w}_\natural^\circ\|_2 > \|\boldsymbol{I}_x\boldsymbol{Q}^{\circ+}\|_F + t\|\boldsymbol{I}_x\boldsymbol{Q}^{\circ+}\|_2\right),$$

is upper bounded by $\exp\left(-\frac{c_1t^2}{\sigma^4}\right)$. First, we see that both $\|\boldsymbol{I}_x\boldsymbol{Q}^{\circ+}\|_F$ and $\|\boldsymbol{I}_x\boldsymbol{Q}^{\circ+}\|_2$ are bounded by $\|\boldsymbol{I}_x\|_2\|(\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ\top}\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^\circ)^{-1}\|_2\|\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ\top}\|_F \leq \frac{\sqrt{|\mathcal{X}_{\rm sf}|+|\mathcal{X}_{\rm bd}|+|\mathcal{M}_{\rm bi}|}}{C_{\min}}$ due to the lower eigenvalue condition and the normalized measurement condition. Moreover, the probability

$$\mathbb{P}\left(\|\boldsymbol{I}_x(\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ\top}\boldsymbol{w}_\natural^\circ\|_2 > t\frac{\sqrt{|\mathcal{X}_{\rm sf}| + |\mathcal{X}_{\rm bd}| + |\mathcal{M}_{\rm bi}|}}{C_{\min}}\right)$$

is upper bounded by $\exp\left(-\frac{c_1t^2}{\sigma^4}\right)$ for any positive t. Together, we conclude the proof.