Supplementary Material for
Boundary Defense against Cyber Threat for Power System Operation

Ming Jin* Javad Lavaei† Somayeh Sojoudi‡ Ross Baldick §

This supplementary material includes formal theory and additional experimental details for the paper "Boundary Defense against Cyber Threat for Power System Operation." The manuscript is organized as follows. We first discuss the preliminaries in Section 1, including notations, power system modeling, the linear basis of representation, and the measurement model considered in the study. We introduce the two-step pipeline of state estimation in Section 2, where we discuss the algorithms with and without the second-order cone constraints and their connection to robust statistics. Section 3 introduces the boundary defense mechanism, including the main results for boundary defense (Lemmas 5 and 12), implications of local property for global property (Lemmas 9 and 15), and performance guarantees for estimation accuracy and bad data detection (Theorems 10, 11, 17 and 18). Experimental details and additional figures are shown in Section 4. The proofs of the main theorems are delegated to Section 5.

1 Preliminaries

1.1 Notations

Vectors are shown by bold letters, and matrices are shown by bold and capital letters. Let x_i denote the i-th element of vector $\boldsymbol{x}$. We use $\mathbb{R}$ and $\mathbb{C}$ as the sets of real and complex numbers, and $\mathbb{S}^n$ and $\mathbb{H}^n$ to represent the spaces of $n \times n$ real symmetric matrices and $n \times n$ complex Hermitian matrices, respectively. A set of indices $\{1, 2, ..., m\}$ is denoted by $[m]$. 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 i. The notations $\angle x$ and $|x|$ indicate the angle and magnitude of a complex scalar. For a convex function $g(\boldsymbol{x})$, we use $\nabla g(\boldsymbol{x})$ and $\partial g(\boldsymbol{x})$ to denote its gradient and subgradient at $\boldsymbol{x}$, respectively. 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. Let $\boldsymbol{I}^{(n)}$ denote the identity matrix of dimension n, but sometimes for simplicity, we omit the superscript whenever the dimension is clear from the context. The notations $\|\boldsymbol{x}\|_0$, $\|\boldsymbol{x}\|_1$, $\|\boldsymbol{x}\|_2$ and $\|\boldsymbol{x}\|_\infty$ show the cardinality, 1-norm, 2-form and ∞-norm of $\boldsymbol{x}$. We use $\|\cdot\|_\infty$ to denote the matrix infinity norm (i.e., the maximum absolute column sum of the matrix). Note that the notations p and q are used for active power and reactive power, respectively.

*Department of Industrial Engineering and Operation Research, University of California Berkeley, CA 94720, USA
†Department of Industrial Engineering and Operation Research, University of California Berkeley, CA 94720, USA
‡Department of Electrical Engineering and Computer Sciences, University of California Berkeley, CA 94720, USA
§Department of Electrical and Computer Engineering, University of Texas at Austin, TX 78712, USA

1.2 Power system modeling

We model the electric grid as a graph $\mathcal{G} \coloneqq \{\mathcal{N}, \mathcal{L}\}$, where $\mathcal{N} \coloneqq [n_b]$ and $\mathcal{L} \coloneqq [n_l]$ represent its sets of buses and branches. Each branch $\ell \in \mathcal{L}$ that connects bus f and bus t is characterized by the branch admittance $y_\ell = g_\ell + \mathrm{i} b_\ell$ and the shunt admittance $y_\ell^{\text{sh}} = g_\ell^{\text{sh}} + \mathrm{i} b_\ell^{\text{sh}}$, where g_ℓ (resp., g_ℓ^{sh}) and b_ℓ (resp., b_ℓ^{sh}) denote the (shunt) conductance and susceptance, respectively. Typically, $g_\ell^{\text{sh}} \ll b_\ell^{\text{sh}}$, so it is set to zero in the subsequent description. In addition, to avoid duplicate definition, each line $\ell = (i, j)$ is defined with a direction from bus i (i.e., *from* end, given by $f(\ell) = i$) to bus j (i.e., *to* end, given by $t(\ell) = j$). We also use $\{i, j\}_\ell$ or simply $\{i, j\}$ to denote a line ℓ that connects nodes i and j.

The power system state is described by the complex voltage at each bus $\boldsymbol{v} = \begin{bmatrix} v_1, ..., v_{n_b} \end{bmatrix}^\top \in \mathbb{C}^{n_b}$, where $v_k \in \mathbb{C}$ is the complex voltage at bus $k \in \mathcal{N}$ with magnitude $|v_k|$ and phase $\theta_k \coloneqq \angle v_k$. Given the complex voltages, by Ohm's law, the complex current injected into line $\{k, j\}_\ell$ at bus k is given by:

$$i_{kj} = y_\ell (v_k - v_j) + \frac{\mathrm{i}}{2} b_\ell^{\text{sh}} v_k.$$

By defining $\theta_{kj} \coloneqq \theta_k - \theta_j$, one can write the power flow from bus k to bus j as

$$\begin{aligned} p_{kj}^{(\ell)} &= |v_k|^2 g_\ell - |v_k||v_j|(g_\ell \cos\theta_{kj} - b_\ell \sin\theta_{kj}), \\ q_{kj}^{(\ell)} &= -|v_k|^2 (b_\ell + \tfrac{1}{2} b_\ell^{\text{sh}}) + |v_k||v_j|(b_\ell \cos\theta_{kj} - g_\ell \sin\theta_{kj}), \end{aligned}$$

and active (reactive) power injections at bust f as

$$p_k = \sum_{\{k,j\}_\ell} p_{kj}^{(\ell)}, \qquad q_k = \sum_{\{k,j\}_\ell} q_{kj}^{(\ell)}. \tag{1}$$

The above formulas are based on polar coordinates of complex voltages, where measurements are nonlinear functions of voltage magnitudes and phases. Another popular representation is based on rectangular coordinates of complex numbers, where measurements are expressed as quadratic functions of the real and imaginary parts of voltages (see [4, Chap. 1] for more details). We use "PV bus" and "PQ bus" to denote buses with real power injection and voltage magnitudes, and buses with real and reactive power injection measurements, respectively.

1.3 Linear basis of representation

We discuss a new basis of representation proposed in [9], where measurements can be expressed as *linear combinations* of the quantities derived form bus voltages. Specifically, for a given system $\mathcal{G}$, we introduce two groups of variables:

1. voltage magnitude square, $x_k^{\text{mg}} \coloneqq |v_k|^2$, for each bus $k \in \mathcal{N}$, and
2. real and imaginary parts of complex products, denoted as $x_\ell^{\text{re}} \coloneqq \Re(v_i v_j^*)$ and $x_\ell^{\text{im}} \coloneqq \Im(v_i v_j^*)$, respectively, for each line $\ell = (i, j)$. Note that there is only one set of variables $\{x_\ell^{\text{re}}, x_\ell^{\text{im}}\}$ for each line.

Using this representation, we can derive various types of power and voltage measurements as follows:

- *Voltage magnitude square.* The voltage square magnitude square at bus $k \in \mathcal{N}$ is simply x_k^{mg} by definition;

- *Branch power flows.* For each line $\ell = (i, j)$, the real and reactive power flows from bus i to bus j and in the reverse direction are given by:

$$\begin{aligned} p_{ij}^{(\ell)} &= g_\ell x_i^{\text{mg}} - g_\ell x_\ell^{\text{re}} - b_\ell x_\ell^{\text{im}} \\ q_{ij}^{(\ell)} &= -(b_\ell + \tfrac{1}{2} b_\ell^{\text{sh}}) x_i^{\text{mg}} + b_\ell x_\ell^{\text{re}} - g_\ell x_\ell^{\text{im}} \\ p_{ji}^{(\ell)} &= g_\ell x_j^{\text{mg}} - g_\ell x_\ell^{\text{re}} + b_\ell x_\ell^{\text{im}} \\ q_{ji}^{(\ell)} &= -(b_\ell + \tfrac{1}{2} b_\ell^{\text{sh}}) x_j^{\text{mg}} + b_\ell x_\ell^{\text{re}} + g_\ell x_\ell^{\text{im}} \end{aligned}$$

- *Nodal power injection.* The power injection at bus node k consists of real and reactive powers, i.e. $p_k + \mathrm{i} q_k$, where:

$$\begin{aligned} p_k &= \sum_{k \in \ell} g_\ell x_k^{\text{mg}} - \sum_{k \in \ell} g_\ell x_\ell^{\text{re}} - (\sum_{f(\ell)=k} b_\ell - \sum_{t(\ell)=k} b_\ell) x_\ell^{\text{im}} \\ q_k &= -(\sum_{k \in \ell} b_\ell + \tfrac{1}{2} b_\ell^{\text{sh}}) x_k^{\text{mg}} + \sum_{k \in \ell} b_\ell x_\ell^{\text{re}} - (\sum_{f(\ell)=k} g_\ell - \sum_{t(\ell)=k} g_\ell) x_\ell^{\text{im}}, \end{aligned}$$

  where $\sum_{k \in \ell}$ is the sum over all lines $\ell \in \mathcal{L}$ that are connected to k, $\sum_{f(\ell)=k}$ is the sum over all lines ℓ where $f(\ell) = k$, and similarly, $\sum_{t(\ell)=k}$ is the sum over all lines ℓ where $t(\ell) = k$. Equivalently, we can use (1) to combine the branch power flows defined above.

Thus, each customary measurement in power systems that belongs to one of the above *measurement types* can be represented by a linear function[1]:

$$m_i(\boldsymbol{x}) = \boldsymbol{a}_i^\top \boldsymbol{x}_\natural, \tag{2}$$

where $\boldsymbol{a}_i \in \mathbb{R}^{n_x}$ is the vector for the i-th noiseless measurement and $\boldsymbol{x}_\natural = (\{x_k^{\text{mg}}\}_{k \in \mathcal{N}}, \{x_\ell^{\text{im}}, x_\ell^{\text{im}}\}_{\ell \in \mathcal{L}})$ is the regression vector. By collecting all the sensor measurements in a vector $\boldsymbol{m} \in \mathbb{R}^{n_m}$, we have

$$\boldsymbol{m} = \boldsymbol{A} \boldsymbol{x}_\natural, \tag{3}$$

where $\boldsymbol{A} \in \mathbb{R}^{n_m \times n_x}$ is the sensing matrix with rows $\boldsymbol{a}_i^\top$ for $i \in [n_m]$.

1.4 Measurement model

To perform SE, the supervisory control and data acquisition (SCADA) system collects measurements about power flows and complex voltages at key locations instrumented with sensors. This process is subject to both ubiquitous sensor noise and randomly occurring sensor faults. We consider the measurement model as follows:

$$\boldsymbol{y} = \boldsymbol{A} \boldsymbol{x}_\natural + \mathbf{w}_\natural + \boldsymbol{b}_\natural, \tag{4}$$

where $\boldsymbol{A} \in \mathbb{R}^{n_m \times n_x}$ and $\boldsymbol{x}_\natural \in \mathbb{R}^{n_x}$ are the sensing matrix and the true regression vector in (3), $\mathbf{w}_\natural \in \mathbb{R}^{n_m}$ denotes random noise, and $\mathbf{b}_\natural \in \mathbb{R}^m$ is the bad data error that accounts for sensor failures or adversarial noise [8]. Note that $\boldsymbol{x}$ serves as an intermediate parameter and the end goal is to find $\boldsymbol{v}$.

Because the sensor data are of different types and their corresponding measurements could be of different scales, we introduce the following condition.

[1] It is straightforward to include linear PMU measurements in our analysis as well using the relation $\tan \theta_{ij} = x_\ell^{\text{im}} / x_\ell^{\text{re}}$ for each line $\ell = (i, j)$. Thus, as long as we have two adjacent PMU measurements, we can use the phase difference to construct a linear measurement equation $x_\ell^{\text{im}} - \tan \theta_{ij} x_\ell^{\text{re}} = 0$.

Definition 1 (Measurement normalization convention)**.** *Each row of $\boldsymbol{A}$ corresponding to a voltage magnitude measurement is normalized by the degree of connection of the node k, $\|\boldsymbol{a}_i\|_2^2 = \deg(k)$, and 1 otherwise $\|\boldsymbol{a}_i\|_2^2 = 1$, where $\boldsymbol{a}_i$ is the i-th row of $\boldsymbol{A}$. The only exception is when the line vulnerability (c.f., Def. 7) is calculated, when all the measurements are normalized by 1.*

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.

2 Two-step pipeline of state estimation

This section describes the two-step state estimation method. For the first step, we discuss algorithms in two categories, which differ by whether or not the second-order cone constraints are incorporated (the case without second-order cone constraints is proposed in [10]). Within each category, we also propose two slight variations, which differ by whether the term of squared loss is included. For the second step, we propose two approaches based on quadratic programming. In the main paper, we mainly discussed the case in which we have both sparse corruption and dense noise. However, the case with sparse corruption but no dense noise is important and also simpler in terms of analysis. Therefore, we include them here as well.

2.1 Step 1: Estimation of $\boldsymbol{x}_\natural$

In the first step, the goal is to estimate $\boldsymbol{x}_\natural$ from a set of noisy and corrupted measurements $\boldsymbol{y}$. We consider two cases separately. In the first case, the dense noise is negligible, i.e., $\boldsymbol{w}_\natural = \boldsymbol{0}$, and we only need to consider the sparse measurement corruption $\boldsymbol{b}$.

Case 1: Sparse corruption but no dense noise (i.e., $\boldsymbol{w} = \boldsymbol{0}$)

In this case, the measurements are given by $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_\natural + \boldsymbol{b}_\natural$. To estimate $\boldsymbol{x}_\natural$, we solve the following program:

$$\min_{\boldsymbol{b}\in\mathbb{R}^{n_m},\boldsymbol{x}\in\mathbb{R}^{n_x}} \|\boldsymbol{b}\|_1, \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} = \boldsymbol{y}. \tag{S$^{(1)}$: ℓ_1}$$

Briefly, under some mild conditions on observability and robusteness to be specified in Section 3, we can faithfully recover $\boldsymbol{b}_\natural$ from the above program. As a consequence, $\boldsymbol{x}_\natural$ can be obtained by performing regression using the remaining good data.

For this case, we can also incorporate second-order cone (SOC) constraints:

$$\min_{\boldsymbol{b}\in\mathbb{R}^{n_m},\boldsymbol{x}\in\mathbb{R}^{n_x}} \|\boldsymbol{b}\|_1, \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} = \boldsymbol{y}, \quad \boldsymbol{x} \in \mathcal{K}, \tag{S$^{(1)}$: ℓ_1-$\mathcal{K}$}$$

where

$$\mathcal{K} = \left\{ \boldsymbol{x} \in \mathbb{R}^{n_x} \middle| \begin{bmatrix} x_i^{\text{mg}} & x_\ell^{\text{re}} + jx_\ell^{\text{im}} \\ x_\ell^{\text{re}} - jx_\ell^{\text{im}} & x_j^{\text{mg}} \end{bmatrix} \succeq 0, \quad \forall \ell := (i,j) \in \mathcal{L} \right\}. \tag{5}$$

Let $\sigma(x)$ denote the index of the variable x (e.g., $x_i^{\text{mg}}, x_\ell^{\text{re}}, x_\ell^{\text{im}}$) in the vector $\boldsymbol{x}$. For instance, $\sigma(x_i^{\text{mg}})$ denotes the index of x_i^{mg} in $\boldsymbol{x}$. The SOC constraint can be equivalently written as:

$$\boldsymbol{c}_\ell^\top \boldsymbol{x} \geq \|\boldsymbol{D}_\ell \boldsymbol{x}\|_2 \quad \Leftrightarrow \quad \begin{bmatrix} \boldsymbol{D}_\ell \\ \boldsymbol{c}_\ell^\top \end{bmatrix} \boldsymbol{x} \in \mathcal{C}_5, \tag{6}$$

where $\boldsymbol{c}_\ell \in \mathbb{R}^{n_x}$ has its $\sigma(x_i^{\text{mg}})$ and $\sigma(x_j^{\text{mg}})$ entries to be $\frac{1}{\sqrt{2}}$ and 0 elsewhere, and $\boldsymbol{D}_\ell \in \mathbb{R}^{4\times n_x}$ has its $(1, \sigma(x_i^{\text{mg}}))$ and $(2, \sigma(x_j^{\text{mg}}))$ entries to be $\frac{1}{\sqrt{2}}$ and its $(3, \sigma(x_\ell^{\text{re}}))$ and $(4, \sigma(x_\ell^{\text{im}}))$ entries to be 1, and 0 elsewhere, and $\mathcal{C}_5$ denotes the second-order cone of dimension 5.

The problem ($\text{S}^{(1)}$: ℓ_1-$\mathcal{K}$) can be reformulated as:

$$\min_{\boldsymbol{b}\in\mathbb{R}^{n_m},\boldsymbol{x}\in\mathbb{R}^{n_x}} \|\boldsymbol{b}\|_1, \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}=\boldsymbol{y}, \begin{bmatrix}\boldsymbol{D}_\ell \\ \boldsymbol{c}_\ell^\top\end{bmatrix}\boldsymbol{x} \in \mathcal{C}_5, \forall \ell \in \mathcal{L} \tag{7}$$

using standard SOCP notations. The Lagrangian is given by:

$$L\left(\boldsymbol{x},\boldsymbol{b},\{\nu_\ell,\boldsymbol{\mu}_\ell\}_{\ell\in\mathcal{L}},\boldsymbol{h}\right) = \|\boldsymbol{b}\|_1 + \boldsymbol{h}^\top\left(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\right) - \sum_{\ell\in\mathcal{L}}\left(\nu_\ell \boldsymbol{c}_\ell^\top\boldsymbol{x} + \boldsymbol{\mu}_\ell\boldsymbol{D}_\ell\boldsymbol{x}\right)$$

The Karush-Kuhn-Tucker (KKT) conditions are given by:

$$\begin{aligned}
&\text{(primal feasibility)} && \boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}=\boldsymbol{y}, \quad \boldsymbol{c}_\ell^\top\boldsymbol{x} \geq \|\boldsymbol{D}_\ell\boldsymbol{x}\|_2, \quad \forall \ell\in\mathcal{L} && (8)\\
&\text{(dual feasibility)} && \nu_\ell \geq \|\boldsymbol{\mu}_\ell\|_2, \quad \forall \ell\in\mathcal{L} && (9)\\
&\text{(stationarity)} && -\sum_{\ell\in\mathcal{L}}(\nu_\ell\boldsymbol{c}_\ell + \boldsymbol{D}_\ell^\top\boldsymbol{\mu}_\ell) = \boldsymbol{A}^\top\boldsymbol{h}, \quad \boldsymbol{h}\in\partial\|\boldsymbol{b}\|_1 && (10)\\
&\text{(complementary slackness)} && \nu_\ell\boldsymbol{c}_\ell^\top\boldsymbol{x} + \boldsymbol{\mu}_\ell^\top\boldsymbol{D}_\ell\boldsymbol{x} = \boldsymbol{0}, \quad \forall\ell\in\mathcal{L}. && (11)
\end{aligned}$$

Therefore, the dual program of ($\text{S}^{(1)}$: ℓ_1-$\mathcal{K}$) is given by:

$$\begin{aligned}
\max_{\boldsymbol{h}\in\mathbb{R}^{n_m},\{\nu_\ell,\boldsymbol{\mu}_\ell\}_{\ell\in\mathcal{L}}} \quad & \boldsymbol{h}^\top\boldsymbol{y} && (12a)\\
\text{subject to} \quad & -\sum_{\ell\in\mathcal{L}}(\nu_\ell\boldsymbol{c}_\ell + \boldsymbol{D}_\ell^\top\boldsymbol{\mu}_\ell) = \boldsymbol{A}^\top\boldsymbol{h} && (12b)\\
& \|\boldsymbol{h}\|_\infty \leq 1 && (12c)\\
& \nu_\ell \geq \|\boldsymbol{\mu}_\ell\|_2, \quad \forall\ell\in\mathcal{L} && (12d)
\end{aligned}$$

Case 2: Sparse corruption and dense noise

In this case, the dense noise cannot be ignored, and the measurements are given by (2). We perform the estimation by solving the following mixed-objective optimization:

$$\min_{\boldsymbol{b}\in\mathbb{R}^{nm},\boldsymbol{x}\in\mathbb{R}^{nx}} \tfrac{1}{2n_m}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2^2 + \lambda\|\boldsymbol{b}\|_1, \tag{$\text{S}^{(1)}$: $\ell_2\ell_1$}$$

where $\lambda > 0$ is the regularization coefficient. Due to the existence of dense noise, it is no longer possible to exactly recover the true $\boldsymbol{x}_\natural$; however, if the magnitude of each dense noise is small, then we can still have strong statistical bounds on the estimation error. The optimization ($\text{S}^{(1)}$: $\ell_2\ell_1$) is equivalent to minimizing the Huber loss as discussed in Section 2.2.

We can also incorporate second-order cone constraints:

$$\min_{\boldsymbol{b}\in\mathbb{R}^{n_m},\boldsymbol{x}\in\mathbb{R}^{n_x}} \tfrac{1}{2n_m}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2^2 + \lambda\|\boldsymbol{b}\|_1, \quad \text{subject to} \quad \boldsymbol{x}\in\mathcal{K}, \tag{$\text{S}^{(1)}$: $\ell_2\ell_1$-$\mathcal{K}$}$$

where $\mathcal{K}$ is defined in (5). The Lagrangian of ($\text{S}^{(1)}$: $\ell_2\ell_1$-$\mathcal{K}$) is given by:

$$L\left(\boldsymbol{x},\boldsymbol{b},\{\boldsymbol{\mu}_\ell\}_{\ell\in\mathcal{L}},\{\nu_\ell\}_{\ell\in\mathcal{L}},\boldsymbol{h}\right) = \tfrac{1}{2n_m}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2^2 + \lambda\|\boldsymbol{b}\|_1 - \sum_{\ell\in\mathcal{L}}\left(\nu_\ell\boldsymbol{c}_\ell^\top\boldsymbol{x} + \boldsymbol{\mu}_\ell\boldsymbol{D}_\ell\boldsymbol{x}\right)$$

The KKT conditions are given by:

$$\begin{aligned}
&\text{(primal feasibility)} && \boldsymbol{c}_\ell^\top \boldsymbol{x} \geq \|\boldsymbol{D}_\ell \boldsymbol{x}\|_2\,, \quad \forall \ell \in \mathcal{L} && (13)\\
&\text{(dual feasibility)} && \nu_\ell \geq \|\boldsymbol{\mu}_\ell\|_2, \quad \forall \ell \in \mathcal{L} && (14)\\
&\text{(stationarity)} && \frac{1}{n_m}\boldsymbol{A}^\top(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}) + \sum_{\ell\in\mathcal{L}}(\nu_\ell \boldsymbol{c}_\ell + \boldsymbol{D}_\ell^\top \boldsymbol{\mu}_\ell) = \boldsymbol{0} && (15)\\
& && \frac{1}{n_m}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}) = \lambda \boldsymbol{h}, \quad \boldsymbol{h} \in \partial\|\boldsymbol{b}\|_1 && (16)\\
&\text{(complementary slackness)} && \nu_\ell \boldsymbol{c}_\ell^\top \boldsymbol{x} + \boldsymbol{\mu}_\ell^\top \boldsymbol{D}_\ell \boldsymbol{x} = \boldsymbol{0}, \quad \forall \ell \in \mathcal{L}. && (17)
\end{aligned}$$

The KKT conditions are important for the analysis in Section 3.

2.2 Connection with robust statistics for bad data detection

The so-called bad data rejection and state estimation form an important part of power systems supervisory control and data acquisition. There are traditional statistical approaches to bad data rejection that involve iteratively eliminating the measurements with the largest residual that are obtained from a least squares estimation (see [17, Section 9.6]). Such a smooth quadratic objective can, however, mask bad data by "spreading" the error around the system. An alternative approach developed in [2] is to use an ℓ_1 objective, which can identify multiple bad data directly. However, the resulting estimate does not average out the effect of dense, independent measurement errors.

The so-called Huber loss that is quadratic for small measurement residuals but constant or linear for large measurement residuals has been explored in [12, 3, 18]. The quadratic-linear loss function is convex, continuous and differentiable at the transition between the quadratic and linear part, and is given by [7]:

$$f_{\text{Huber}}(r;\psi) = \begin{cases} \frac{1}{2}r^2 & |r| \leq \psi \\ \psi(|r| - \frac{1}{2}\psi) & |r| > \psi \end{cases}, \tag{18}$$

where ψ is the hyper-parameter controlling the transition point between the ℓ_2 and ℓ_1 loss functions.

There is an interesting connection between ($\text{S}^{(1)}$: $\ell_2\ell_1$) and the Huber loss. To see this, we can view the optimization over $\boldsymbol{b}$ and $\boldsymbol{x}$ in ($\text{S}^{(1)}$: $\ell_2\ell_1$) as an inner optimization with $\boldsymbol{b}$ for a given $\boldsymbol{x}$, and an outer optimization with $\boldsymbol{x}$. The inner optimization is composed of a series of smaller optimization problems

$$\min_{b_i} \tfrac{1}{2n_m}(y_i - \boldsymbol{a}_i^\top \boldsymbol{x} - b_i)^2 + \psi|b_i|, \tag{19}$$

for $i \in [n_m]$, which has the optimal solution

$$b_i^* = \text{sign}(y_i - \boldsymbol{a}_i^\top \boldsymbol{x}) \max\left(0, \left|y_i - \boldsymbol{a}_i^\top \boldsymbol{x}\right| - \psi\right), \tag{20}$$

where $\text{sign}(y)$ is the sign of y. Now, by defining $r_i \coloneqq y_i - \boldsymbol{a}_i^\top \boldsymbol{x}$, we substitute the solution into the outer optimization to obtain

$$\tfrac{1}{n_m} \sum_{i\in[n_m]} \tfrac{1}{2}(r_i - \text{sign}(r_i)\max(0, |r_i| - \psi))^2 + \psi|\max(0, |r_i| - \psi)|. \tag{21}$$

Hence, it can be seen that the above expression is equal to the Huber loss:

$$\frac{1}{n_m}\sum_{i\in[n_m]} f_{\text{Huber}}(y_i - \boldsymbol{a}_i^\top \boldsymbol{x};\psi). \tag{22}$$

Despite the wide usage of Huber loss in power system estimation, the existing studies in the literature are mostly empirical. The approach proposed here allows for strong mathematical results that go well beyond the promising empirical results.

2.3 Step 2: Recovery of $\boldsymbol{v}$

The goal of the second step is to recover the underlying system voltage $\boldsymbol{v}$ from the estimation $\hat{\boldsymbol{x}}$ obtained in Step 1. First, we transform $\hat{\boldsymbol{x}}$ into estimations of voltage magnitudes and phase differences:

- The voltage magnitude at each bus $k \in \mathcal{N}$ can be obtained by $|\hat{v}_k| = \sqrt{\hat{x}_k^{\text{mg}}}$;
- The phase difference along each line $\ell = (i,j)$ is given by $\hat{\theta}_{ij} = \arctan \hat{x}_\ell^{\text{im}}/\hat{x}_\ell^{\text{re}}$.

To obtain the estimations of phases at each bus, we propose two methds. The first method is to solve the least-squares problem

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^{n_b}} \sum_{\ell=(i,j)} (\theta_i - \theta_j - \hat{\theta}_{ij})^2, \tag{S$^{(2)}$: ℓ_2}$$

which has a closed-form solution: let $\boldsymbol{\theta}_\Delta$ be a collection of $\hat{\theta}_{ij}$, and $\boldsymbol{L} \in \mathbb{R}^{n_l\times n_b}$ be a sparse matrix with $L(\ell,i) \coloneqq 1$ and $L(\ell,j) \coloneqq -1$ for each line $\ell = (i,j)$ and zero elsewhere. Then, the solution for (S$^{(2)}$: ℓ_2) is given by:

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{L}^\top \boldsymbol{L})^{-1}\boldsymbol{L}^\top \boldsymbol{\theta}_\Delta. \tag{23}$$

The second approach is to solve a mixed-objective problem, similar to the first step:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^{n_b}} \tfrac{1}{n_l}\sum_{\ell=(i,j)} (\theta_i - \theta_j - \hat{\theta}_{ij})^2 + \lambda_2 \sum_{\ell=(i,j)} |\theta_i - \theta_j - \hat{\theta}_{ij}|. \tag{S$^{(2)}$: $\ell_2\ell_1$}$$

In this case, there is no longer a closed-form solution available, but the advantage is that it is robust to large errors in the phase difference estimation, in case the first step method does not fully detect the bad data in the measurements.

Finally, we can reconstruct $\hat{\boldsymbol{v}}$ via the formula:

$$\hat{v}_k = |\hat{v}_k| e^{\mathrm{i}\hat{\theta}_k}, \qquad k \in \mathcal{N}. \tag{24}$$

If the regression vector from Step 1 is exact, i.e., $\hat{\boldsymbol{x}} = \boldsymbol{x}_\natural$, then we can use (S$^{(2)}$: ℓ_2) to accurately recover the system state $\hat{\boldsymbol{v}} = \boldsymbol{v}$. Even if the $\hat{\boldsymbol{x}}$ is not exact, the second stage estimator (S$^{(2)}$: $\ell_2\ell_1$) has nice properties to control the estimation error, and therefore any potential error in $\hat{\theta}_{ij}$ does not propagate along the branches.

3 Boundary defense mechanism

In this section, we give a detailed discussion of the proposed boundary defense mechanism. For a given attack scenario, we define a natural partition of the network into the attacked, inner and outer boundaries, and safe regions. For the rest of the analysis, we denote $\boldsymbol{x}_\sharp$ and $\boldsymbol{b}_\sharp$ as the ground truth for state $\boldsymbol{x}$ and bad data $\boldsymbol{b}$, as defined in (4). We also include a nomenclature table below to help manage the notations.

Nomenclature

General notations

$\mathcal{B}_{\rm at}$ Attacked region, i.e., the subgraph induced by $\mathcal{N}_{\rm at}$
$\mathcal{B}_{\rm bd}$ Boundary region, i.e., the subgraph induced by $\mathcal{N}_{\rm bd}$
$\mathcal{B}_{\rm bi}$ Inner boundary, i.e., the subgraph induced by $\mathcal{N}_{\rm bi}$
$\mathcal{B}_{\rm bo}$ Outer boundary, i.e., the subgraph induced by $\mathcal{N}_{\rm bo}$
$\mathcal{B}$ Union of the attacked and inner boundary regions, i.e., $\mathcal{B} = \mathcal{B}_{\rm at} \cup \mathcal{B}_{\rm bi}$
$\mathcal{B}_{\rm sf}$ Safe region, i.e., the subgraph induced by $\mathcal{N}_{\rm sf}$
$\mathcal{L}_{\rm at\cap bi}$ Set of lines that bridge nodes between $\mathcal{N}_{\rm at}$ and $\mathcal{N}_{\rm bi}$
$\mathcal{L}_{\rm at}$ Set of lines in the subgraph $\mathcal{B}_{\rm at}$
$\mathcal{L}_{\rm bd}$ Set of lines in the subgraph $\mathcal{B}_{\rm bd}$
$\mathcal{L}_{\rm bi\cap bo}$ Set of lines that bridge nodes between $\mathcal{N}_{\rm bi}$ and $\mathcal{N}_{\rm bo}$
$\mathcal{M}_{\rm at}$ Attacked measurements, includes measurement (if any) on nodes $\mathcal{N}_{\rm at}$ and lines $\mathcal{L}_{\rm at}$
$\mathcal{M}_{\rm bd}$ Boundary measurements, i.e., $\mathcal{M}_{\rm bi} \cup \mathcal{M}_{\rm bo}$
$\mathcal{M}_{\rm bi}$ Inner boundary measurements, includes measurement (if any) on nodes $\mathcal{N}_{\rm bi}$ and lines $\mathcal{L}_{\rm at\cap bi}$
$\mathcal{M}_{\rm bo}$ Outer boundary measurements, includes measurement (if any) on nodes $\mathcal{N}_{\rm bd}$ and lines $\mathcal{L}_{\rm bd}$
$\mathcal{M}$ Set of all measurements
$\mathcal{M}_{\rm sf}$ Safe measurements, $\mathcal{M} \setminus (\mathcal{M}_{\rm at} \cup \mathcal{M}_{\rm bd})$
$\mathcal{N}_{\rm at}$ Set of nodes under attack, the set of lines induced by $\mathcal{N}_{\rm at}$, and the set of lines connecting them
$\mathcal{N}_{\rm bd}$ Nodes in the boundary region $\mathcal{N}_{\rm bi} \cup \mathcal{N}_{\rm bo}$
$\mathcal{N}_{\rm bi}$ Set of nodes adjacent to the attacked region $\{i \in \mathcal{N} \setminus \mathcal{N}_{\rm at} \mid \exists j \in \mathcal{N}_{\rm at},\ \text{s.t. } \{i,j\} \in \mathcal{L}\}$
$\mathcal{N}_{\rm bo}$ Set of nodes adjacent to the inner boundary $\{i \in \mathcal{N} \setminus (\mathcal{N}_{\rm at} \cup \mathcal{N}_{\rm bi}) \mid \exists j \in \mathcal{N}_{\rm bi},\ \text{s.t. } \{i,j\} \in \mathcal{L}\}$
$\mathcal{N}_{\rm sf}$ Set of nodes that are not attacked or on the boundary $\mathcal{N} \setminus (\mathcal{N}_{\rm at} \cup \mathcal{N}_{\rm bd})$
$\mathcal{X}_{\rm at}$ Attacked variables, $\boldsymbol{x}_{\rm at} \in \mathcal{X}_{\rm at}$ includes $x_i^{\rm mg}$ for nodes in $\mathcal{N}_{\rm at}$ and $(x_\ell^{\rm re}, x_\ell^{\rm im})$ for lines in $\mathcal{L}_{\rm at} \cup \mathcal{L}_{\rm at\cap bi}$
$\mathcal{X}_{\rm bd}$ Boundary variables, $\boldsymbol{x}_{\rm bd} \in \mathcal{X}_{\rm bd}$ includes $x_i^{\rm mg}$ for nodes in $\mathcal{N}_{\rm bd}$ and $(x_\ell^{\rm re}, x_\ell^{\rm im})$ for lines in $\mathcal{L}_{\rm bd}$
$\boldsymbol{x}_\sharp$, $\boldsymbol{b}_\sharp$ Ground truth state vector and bad data defined in (4), respectively
$\mathcal{X}$ Set of all variables
$\mathcal{X}_{\rm sf}$ Safe variables, $\boldsymbol{x}_{\rm sf} \in \mathcal{X}_{\rm sf}$ includes all other variables $\mathcal{X} \setminus (\mathcal{X}_{\rm at} \cup \mathcal{X}_{\rm bd})$

Graphical mutual incoherence

$\mathcal{B}_{\rm at}^{i\to j}$ One-bus attack set that consists of i
$\mathcal{B}_{\rm bi}^{i\to j}$ One-bus inner boundary that consists of j
$\mathcal{B}_{\rm bo}^{i\to j}$ Set of buses (other than i) that are directly connected to j as the outer boundary
$\mathcal{L}_{\rm bd}^{i\to j}$ The union of line $\ell = (i,j)$ and the set of lines that bridge $\mathcal{B}_{\rm bi}^{i\to j}$ and $\mathcal{B}_{\rm bo}^{i\to j}$
$\mathcal{M}_{\rm bd}^{i\to j}$ Boundary measurements, i.e., $\mathcal{M}_{\rm bd\checkmark}^{i\to j} \cup \mathcal{M}_{\rm bd\times}^{i\to j}$
$\mathcal{M}_{\rm bd\times}^{i\to j}$ Set of measurements that depend on both $\mathcal{X}_{\rm bd}^{i\to j}$ and $\{x_\ell^{\rm re}, x_\ell^{\rm im}\}$ of the attacked line $\ell = (i,j)$
$\mathcal{M}_{\rm bd\checkmark}^{i\to j}$ Set of measurements that depend only on the boundary variables $\mathcal{X}_{\rm bd}^{i\to j}$
$\mathcal{X}_{\rm bd}^{i\to j}$ Boundary variables, include $\{x_k^{\rm mg}\}_{k\in\mathcal{B}_{\rm bi}^{i\to j}\cup\mathcal{B}_{\rm bo}^{i\to j}}$ and $\{x_\ell^{\rm re}, x_\ell^{\rm im}\}$ for the set of lines $\ell \in \mathcal{L}$ that connect the inner boundary $\mathcal{B}_{\rm bi}^{i\to j} = \{j\}$ to nodes in the outer boundary $\mathcal{B}_{\rm bo}^{i\to j}$

Graphical mutual incoherence for tree decomposition

$(\mathcal{T}, \mathcal{W})$ Tree decomposition, where $\mathcal{T}$ is a tree and $\mathcal{W} \coloneqq \{W_t \mid t \in \mathcal{N}(\mathcal{T})\}$ is the set of "bags"
$\mathcal{L}_{\rm ad}$ Union of $\mathcal{L}_{\rm ad}(\mathcal{W}_t^{\rm lk})$ for all link bags $\mathcal{W}_t^{\rm lk}$
$\mathcal{L}_{\rm ad}(\mathcal{W}_t^{\rm lk})$ Set of edges that connect adhesion nodes in $\mathcal{W}_t^{\rm lk}$ with infected nodes

$\mathcal{L}(i;\mathcal{G})$ Set of nodes (bags) in $\mathcal{G}$ that are connected to node (bag) i

$\mathcal{L}_{\rm if}$ Set of lines induced by the union of infected bags

$\mathcal{L}_{\rm lk}$ Set of lines induced by the union of link bags

$\mathcal{L}_{\rm sf}$ Set of lines induced by the union of safe bags

$\mathcal{M}_{\rm ad}$ Set of adhesion measurements, includes nodal power injections on nodes in $\mathcal{N}_{\rm ad}$ and line measurements on $\mathcal{L}_{\rm ad}$

$\mathcal{M}_{\rm bd}$ Set of boundary measurements, i.e., $\mathcal{M}_{\rm bd} \coloneqq \mathcal{M}_{\rm ad} \cup \mathcal{M}_{\rm ol}$

$\mathcal{M}_{\rm if}$ Set of infected measurements, includes measurements on lines induced by nodes in $\mathcal{W}^{\rm if}$ and on nodes in $\mathcal{W}^{\rm if}$ except for voltage magnitude measurements on nodes in $\mathcal{N}_{\rm ad}$

$\mathcal{M}_{\rm ol}$ Set of outer link measurements, includes voltage magnitude on nodes in $\mathcal{W}^{\rm lk}$ and line measurements induced by nodes in $\mathcal{W}^{\rm lk}$

$\mathcal{M}_{\rm sf}$ Set of safe measurements, measurements that are not boundary or infected measurements

$\mathcal{N}_{\rm ad}$ Union of the sets of adhesion nodes $\mathcal{N}_{\rm ad}(\mathcal{W}_t^{\rm lk}, \mathcal{W}_t^{\rm if})$ for all $\mathcal{W}_t^{\rm lk}$ and $\mathcal{W}_t^{\rm if}$

$\mathcal{N}_{\rm ad}(\mathcal{W}_t^{\rm lk}, \mathcal{W}_t^{\rm if})$ Set of adhesion nodes, i.e., $\mathcal{W}_t^{\rm lk} \cap \mathcal{W}_t^{\rm if} \subseteq \mathcal{N}_{\rm ad}(\mathcal{W}_t^{\rm lk}, \mathcal{W}_t^{\rm if})$ nodes shared between a link bag $\mathcal{W}_t^{\rm lk}$ and an infected bag $\mathcal{W}_t^{\rm if}$

$\mathcal{N}(\mathcal{G})$ Vertex set of graph $\mathcal{G}$

$\mathcal{N}_{\rm ol}(\mathcal{W}_t^{\rm lk}, \mathcal{W}_t^{\rm if})$ Set of outer link nodes, i.e., $\mathcal{W}_t^{\rm lk} \setminus \mathcal{W}_t^{\rm if} \subseteq \mathcal{N}_{\rm ol}$ nodes in $\mathcal{W}_t^{\rm lk}$ that are not adhesion nodes

$\mathcal{W}^{\rm if}$ Set of infected bags, i.e., $\mathcal{W}_t^{\rm if} \in \{\mathcal{W}_t \mid \mathcal{W}_t \cap \mathcal{N}_{\rm at} \neq \emptyset\}$ bags that contain attacked nodes

$\mathcal{W}^{\rm lk}$ Set of link bags, i.e., $\mathcal{W}_t^{\rm lk} \in \mathcal{W}^{\rm lk} = \{\mathcal{W}_t \mid \mathcal{W}_t \cap \mathcal{N}_{\rm at} = \emptyset, \exists\ \mathcal{W}_{t'}^{\rm if}, \mathcal{W}_t \in \mathcal{L}(\mathcal{W}_{t'}^{\rm if}; \mathcal{T})\}$ bags that are connected to an infected bag

$\mathcal{W}_t^{\rm sf}$ Set of safe bags, i.e., bags other than the infected or link bags

$\mathcal{X}_{\rm if}$ Set of infected variables, includes all variables on lines induced by $\mathcal{W}^{\rm if}$ and on nodes in $\mathcal{W}^{\rm if}$ except for adhesion nodes $\mathcal{N}_{\rm ad}$

$\mathcal{X}_{\rm lk}$ Set of link variables, includes variables on nodes in $\mathcal{W}^{\rm lk}$ and the lines induced by them

$\mathcal{X}_{\rm sf}$ Set of safe variables, includes variables that are either infected nor link variables

Definition 2 (Attacked, boundary, and safe regions)**.** *Let $\mathcal{N}_{\rm at}$ be the set of nodes under attack and the "attacked region" $\mathcal{B}_{\rm at} \coloneqq \{\mathcal{N}_{\rm at}, \mathcal{L}_{\rm at}\}$ be the induced subgraph. Let the "inner boundary" be the set of nodes adjacent to the attacked region $\mathcal{N}_{\rm bi} \coloneqq \{i \in \mathcal{N} \setminus \mathcal{N}_{\rm at} \mid \exists j \in \mathcal{N}_{\rm at},\ \text{s.t. } \{i,j\} \in \mathcal{L}\}$ and the induced graph be denoted as $\mathcal{B}_{\rm bi}$, and the "outer boundary" be the set of nodes adjacent to the inner boundary region $\mathcal{N}_{\rm bo} \coloneqq \{i \in \mathcal{N} \setminus (\mathcal{N}_{\rm at} \cup \mathcal{N}_{\rm bi}) \mid \exists j \in \mathcal{N}_{\rm bi},\ \text{s.t. } \{i,j\} \in \mathcal{L}\}$ and the induced graph be denoted as $\mathcal{B}_{\rm bo}$. Let $\mathcal{N}_{\rm bd} \coloneqq \mathcal{N}_{\rm bi} \cup \mathcal{N}_{\rm bo}$ be nodes in the "boundary region" and $\mathcal{B}_{\rm bd} \coloneqq \{\mathcal{N}_{\rm bd}, \mathcal{L}_{\rm bd}\}$ be the induced subgraph. We also denote the set of lines that bridge nodes between $\mathcal{N}_{\rm at}$ and $\mathcal{N}_{\rm bi}$ as $\mathcal{L}_{\rm at\cap bi}$, and the set of lines that bridge nodes between $\mathcal{N}_{\rm bi}$ and $\mathcal{N}_{\rm bo}$ as $\mathcal{L}_{\rm bi\cap bo}$. Lastly, let $\mathcal{N}_{\rm sf} \coloneqq \mathcal{N} \setminus (\mathcal{N}_{\rm at} \cup \mathcal{N}_{\rm bd})$ be the rest of the nodes and the "safe region" $\mathcal{B}_{\rm sf} \coloneqq \{\mathcal{N}_{\rm sf}, \mathcal{L}_{\rm sf}\}$ be the induced subgraph.*

When there is an attack on a local region, a subset of the local measurements are compromised. We use $\mathcal{B} = \mathcal{B}_{\rm at} \cup \mathcal{B}_{\rm bi}$ to delineate the smallest subgraph to cover this region. For the simplicity of the analysis, we assume that there are no lines connecting two inner boundary nodes in $\mathcal{B}_{\rm bi}$, and that no two nodes in $\mathcal{B}_{\rm at}$ are connected to the same node in $\mathcal{B}_{\rm bi}$ (one can always enlarge the region $\mathcal{B}$ to satisfy these conditions). Furthermore, we make the assumption that no measurements on the nodes (e.g., voltage magnitudes and nodal injections) or on the lines (e.g., power branch flows) within the boundary region $\mathcal{B}_{\rm bd}$ are attacked. The partition set notations in Def. 2 are illustrated in Fig. 1. With the set partition notions ready, we introduce a partition of the measurements and variables.

Definition 3 (Attacked, boundary and safe variables and measurements)**.** *The set of "attacked variables" $\mathcal{X}_{\rm at}$ includes variables on nodes in $\mathcal{N}_{\rm at}$ and lines in $\mathcal{L}_{\rm at} \cup \mathcal{L}_{\rm at\cap bi}$. The set of "boundary variables" $\mathcal{X}_{\rm bd}$*

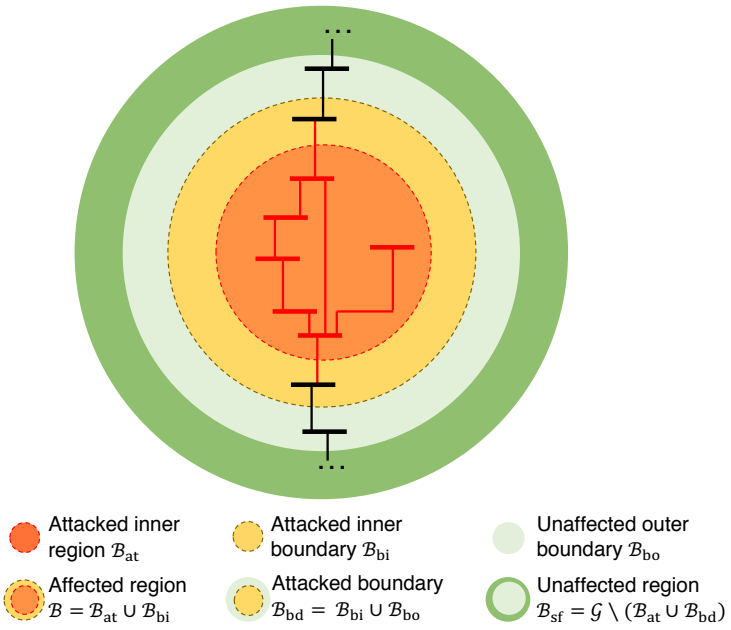

Figure 1: The illustrations of the partition set concepts introduced for the case of zonal attacks. Lines or buses whose measurements are under attack are shown in red.

includes variables on nodes in $\mathcal{N}_{\text{bd}}$ and lines in $\mathcal{L}_{\text{bd}}$. The set of "safe variables" $\mathcal{X}_{\text{sf}}$ includes all other variables. The set of "attacked measurements" $\mathcal{M}_{\text{at}}$ includes measurements on nodes in $\mathcal{B}_{\text{at}}$ and lines in $\mathcal{L}_{\text{at}}$. The set of "inner boundary measurements" $\mathcal{M}_{\text{bi}}$ includes nodal power injections in $\mathcal{B}_{\text{bi}}$ and line measurements in $\mathcal{L}_{\text{at}\cap\text{bi}}$, and the set of "outer boundary measurements" $\mathcal{M}_{\text{bo}}$ includes voltage magnitude and line measurements in $\mathcal{B}_{\text{bd}}$. Together, they form the "boundary measurements" $\mathcal{M}_{\text{bd}} \coloneqq \mathcal{M}_{\text{bi}} \cup \mathcal{M}_{\text{bo}}$. The rest of the measurements $\mathcal{M}_{\text{sf}}$ are "safe measurements."

By definition, the sets $\mathcal{M}_{\text{sf}}$, $\mathcal{M}_{\text{bo}}$, $\mathcal{M}_{\text{bi}}$, $\mathcal{M}_{\text{at}}$ form a partition of $[n_m]$, and the sets $\mathcal{X}_{\text{sf}}$, $\mathcal{X}_{\text{bd}}$, and $\mathcal{X}_{\text{at}}$ form a partition of $[n_x]$. Thus, we can rearrange and partition the matrix $\boldsymbol{A}$ as follows:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{\mathcal{M}_{\text{sf}},\mathcal{X}_{\text{sf}}} & \boldsymbol{A}_{\mathcal{M}_{\text{sf}},\mathcal{X}_{\text{bd}}} & \boldsymbol{A}_{\mathcal{M}_{\text{sf}},\mathcal{X}_{\text{at}}} \\ \boldsymbol{A}_{\mathcal{M}_{\text{bo}},\mathcal{X}_{\text{sf}}} & \boldsymbol{A}_{\mathcal{M}_{\text{bo}},\mathcal{X}_{\text{bd}}} & \boldsymbol{A}_{\mathcal{M}_{\text{bo}},\mathcal{X}_{\text{at}}} \\ \boldsymbol{A}_{\mathcal{M}_{\text{bi}},\mathcal{X}_{\text{sf}}} & \boldsymbol{A}_{\mathcal{M}_{\text{bi}},\mathcal{X}_{\text{bd}}} & \boldsymbol{A}_{\mathcal{M}_{\text{bi}},\mathcal{X}_{\text{at}}} \\ \boldsymbol{A}_{\mathcal{M}_{\text{at}},\mathcal{X}_{\text{sf}}} & \boldsymbol{A}_{\mathcal{M}_{\text{at}},\mathcal{X}_{\text{bd}}} & \boldsymbol{A}_{\mathcal{M}_{\text{at}},\mathcal{X}_{\text{at}}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{\mathcal{M}_{\text{sf}},\mathcal{X}_{\text{sf}}} & \boldsymbol{A}_{\mathcal{M}_{\text{sf}},\mathcal{X}_{\text{bd}}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\text{bo}},\mathcal{X}_{\text{bd}}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\text{bi}},\mathcal{X}_{\text{bd}}} & \boldsymbol{A}_{\mathcal{M}_{\text{bi}},\mathcal{X}_{\text{at}}} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\text{at}},\mathcal{X}_{\text{at}}} \end{bmatrix}. \tag{25}$$

There is no loss of generality in arranging $\boldsymbol{A}$ as above, which is simply for the purpose of presentation. Let $\boldsymbol{I}^{(n_m)}_{\mathcal{M}_{\text{at}}}$, $\boldsymbol{I}^{(n_m)}_{\mathcal{M}_{\text{bi}}}$, $\boldsymbol{I}^{(n_m)}_{\mathcal{M}_{\text{bo}}}$, and $\boldsymbol{I}^{(n_m)}_{\mathcal{M}_{\text{sf}}}$ be matrices that consist of the $\mathcal{M}_{\text{at}}$, $\mathcal{M}_{\text{bi}}$, $\mathcal{M}_{\text{bo}}$, and $\mathcal{M}_{\text{sf}}$ rows from the identity matrix of size n_m, respectively, and $\boldsymbol{I}^{(n_x)}_{\mathcal{X}_{\text{at}}}$, $\boldsymbol{I}^{(n_x)}_{\mathcal{X}_{\text{bd}}}$ and $\boldsymbol{I}^{(n_x)}_{\mathcal{X}_{\text{sf}}}$ be the matrices that consist of the $\mathcal{X}_{\text{at}}$, $\mathcal{X}_{\text{bd}}$ and $\mathcal{X}_{\text{sf}}$ rows from the identity matrix of size n_x. Then, we can obtain each subblock that accounts for a set of measurements (e.g. $\mathcal{M}_{\text{sf}}$) and variables (e.g. $\mathcal{X}_{\text{sf}}$) using the equation $\boldsymbol{A}_{\mathcal{M}_{\text{sf}},\mathcal{X}_{\text{sf}}} = \boldsymbol{I}^{(n_m)}_{\mathcal{M}_{\text{sf}}} \boldsymbol{A} \boldsymbol{I}^{(n_x)\top}_{\mathcal{X}_{\text{sf}}}$ without having to specify a particular order sequence of measurements $\boldsymbol{y}$ or variables $\boldsymbol{x}$,

We introduce the following properties to characterize the sensing matrix $\boldsymbol{A}$.

Definition 4 (Lower eigenvalue)**.** *Let* $\boldsymbol{Q}_{\mathcal{M}_{\rm bd},\mathcal{X}_{\rm bd}} \coloneqq \begin{bmatrix} \boldsymbol{A}_{\mathcal{M}_{\rm bd},\mathcal{X}_{\rm bd}} & \boldsymbol{I}_{\mathcal{M}_{\rm bi}}^{(|\mathcal{M}_{\rm bd}|)\top} \end{bmatrix}$*, where* $\boldsymbol{I}_{\mathcal{M}_{\rm bi}}^{(|\mathcal{M}_{\rm bd}|)}$ *consists of* $\mathcal{M}_{\rm bi}$ *rows of the size–*$|\mathcal{M}_{\rm bd}|$ *identity matrix. Then, the lower eigenvalue* $C_{\min}$ *is the lower bound:*

$$\min\left\{\lambda_{\min}\left(\boldsymbol{Q}_{\mathcal{M}_{\rm bd},\mathcal{X}_{\rm bd}}^{\top}\boldsymbol{Q}_{\mathcal{M}_{\rm bd},\mathcal{X}_{\rm bd}}\right),\lambda_{\min}\left(\boldsymbol{A}_{\mathcal{M}_{\rm bo},\mathcal{X}_{\rm bd}}^{\top}\boldsymbol{A}_{\mathcal{M}_{\rm bo},\mathcal{X}_{\rm bd}}\right),\lambda_{\min}\left(\boldsymbol{A}_{\mathcal{M}_{\rm sf},\mathcal{X}_{\rm sf}}^{\top}\boldsymbol{A}_{\mathcal{M}_{\rm sf},\mathcal{X}_{\rm sf}}\right)\right\}\geq C_{\min}. \tag{26}$$

The value $C_{\min}$ characterizes the influence of bad data on the identifiability of $\boldsymbol{x}_{\natural}$ outside the attacked region. If $C_{\min}$ is strictly positive and one can accurately detect the support of bad data on the boundary, then it is possible to obtain a satisfactory estimation of $\boldsymbol{x}_{\natural}$ outside the attacked region.

3.1 Graphical mutual incoherence and boundary defense for LP/QP

Our goal is to find the attacked region by detecting a sufficiently large number of measurements within $\mathcal{M}_{\rm at}$ while avoiding making false positive detection for measurements belonging to the unaffected region. In other words, if $\hat{\mathcal{J}} \coloneqq \operatorname{supp}(\hat{\boldsymbol{b}})$ denotes the support of the estimated bad data, then it is desirable to have $\hat{\mathcal{J}} \subseteq \mathcal{M}_{\rm at} \cup \mathcal{M}_{\rm bi}$ (here, we allow both false positives and false negatives within the attacked region). The following lemma establishes a key result for the estimation by LP/QP.

Lemma 5 (Boundary defense stops error propagation)**.** *Suppose that there is no dense measurement noise (i.e.,* $\boldsymbol{w} = \boldsymbol{0}$*), and the bad data are confined within* $\mathcal{M}_{\rm at}$*, i.e.,* $\operatorname{supp}(\boldsymbol{b}_{\natural}) \subseteq \mathcal{M}_{\rm at}$*. Also, suppose that* $\boldsymbol{A}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bd},\mathcal{X}_{\rm sf}\cup\mathcal{X}_{\rm bd}}$ *has full column rank. If for an arbitrary* $\boldsymbol{b}_{\star\mathcal{M}_{\rm bd}}$ *defined for measurements in* $\mathcal{M}_{\rm bd}$ *with support limited to the inner boundary, i.e.,* $\operatorname{supp}(\boldsymbol{b}_{\star\mathcal{M}_{\rm bd}}) \subseteq \mathcal{M}_{\rm bi}$*, the solution* $\hat{\boldsymbol{x}}_{\rm bd} \in \mathcal{X}_{\rm bd}$ *to the program*

$$\min_{\boldsymbol{x}_{\rm bd}} \|\boldsymbol{z}_{\mathcal{M}_{\rm bd}} - \boldsymbol{A}_{\mathcal{M}_{\rm bd},\mathcal{X}_{\rm bd}}\boldsymbol{x}_{\rm bd}\|_1 \tag{27}$$

is unique and satisfies the properties $\hat{\boldsymbol{x}}_{\rm bd} = \boldsymbol{x}_{\natural{\rm bd}}$*, where* $\boldsymbol{z}_{\mathcal{M}_{\rm bd}} = \boldsymbol{A}_{\mathcal{M}_{\rm bd},\mathcal{X}_{\rm bd}}\boldsymbol{x}_{\natural{\rm bd}} + \boldsymbol{b}_{\star\mathcal{M}_{\rm bd}}$*, then the solution* $\hat{\boldsymbol{x}}$ *to* (S$^{(1)}$: ℓ_1) *satisfies the properties* $\hat{\boldsymbol{x}}_{\rm bd} = \boldsymbol{x}_{\natural{\rm bd}}$ *and* $\hat{\boldsymbol{x}}_{\rm sf} = \boldsymbol{x}_{\natural{\rm sf}}$*.*

To sketch the proof, since by assumption the unique optimal solution for the measurement-sensing matrix pair $(\boldsymbol{y}_{\mathcal{M}_{\rm sf}}, \boldsymbol{A}_{\mathcal{M}_{\rm sf},\mathcal{X}_{\rm sf}\cup\mathcal{X}_{\rm bd}})$ given $\boldsymbol{x}_{\natural{\rm bd}}$ recovers the ground truth $\boldsymbol{x}_{\natural{\rm sf}}$, we aim at showing that the unique optimal solution of $(\boldsymbol{y}_{\mathcal{M}_{\rm bd}\cup\mathcal{M}_{\rm at}}, \boldsymbol{A}_{\mathcal{M}_{\rm bd}\cup\mathcal{M}_{\rm at},\mathcal{X}_{\rm bd}\cup\mathcal{X}_{\rm at}})$ corresponding to the boundary state coincides with $\boldsymbol{x}_{\natural{\rm bd}}$, which completes the proof because this set of measurements is independent of the states $\boldsymbol{x}_{\rm sf}$. This achieves a de facto coupling of the "weakly coupled" system due to the overlapping regions corresponding to measurements $\boldsymbol{y}_{\mathcal{M}_{\rm bd}}$.

Proof. There are two ways to prove the statement. The first one relies on logical reasoning that is intuitive, while the second approach is based on KKT conditions that can be easily generalized to measurements with dense noise. We start with the first approach, which partitions the loss function in (S$^{(1)}$: ℓ_1) into the sum of three terms:

$$\begin{aligned} f_1(\boldsymbol{x}_{\rm sf},\boldsymbol{x}_{\rm bd}) &= \|\boldsymbol{y}_{\mathcal{M}_{\rm sf}} - \boldsymbol{A}_{\mathcal{M}_{\rm sf},\mathcal{X}_{\rm sf}}\boldsymbol{x}_{\rm sf} - \boldsymbol{A}_{\mathcal{M}_{\rm sf},\mathcal{X}_{\rm bd}}\boldsymbol{x}_{\rm bd}\|_1;\\ f_2(\boldsymbol{x}_{\rm bd},\boldsymbol{x}_{\rm at}) &= \|\boldsymbol{y}_{\mathcal{M}_{\rm bd}} - \boldsymbol{A}_{\mathcal{M}_{\rm bd},\mathcal{X}_{\rm bd}}\boldsymbol{x}_{\rm bd} - \boldsymbol{A}_{\mathcal{M}_{\rm bd},\mathcal{X}_{\rm at}}\boldsymbol{x}_{\rm at}\|_1;\\ f_3(\boldsymbol{x}_{\rm at}) &= \|\boldsymbol{y}_{\mathcal{M}_{\rm at}} - \boldsymbol{A}_{\mathcal{M}_{\rm at},\mathcal{X}_{\rm at}}\boldsymbol{x}_{\rm at}\|_1. \end{aligned}$$

Let $\boldsymbol{z}_{\mathcal{M}_{\rm bd}} = \boldsymbol{y}_{\mathcal{M}_{\rm bd}} - \boldsymbol{A}_{\mathcal{M}_{\rm bd},\mathcal{X}_{\rm at}}\boldsymbol{x}_{\rm at} = \boldsymbol{A}_{\mathcal{M}_{\rm bd},\mathcal{X}_{\rm bd}}\boldsymbol{x}_{\natural{\rm bd}} - \boldsymbol{A}_{\mathcal{M}_{\rm bd},\mathcal{X}_{\rm at}}(\boldsymbol{x}_{\natural{\rm at}} - \boldsymbol{x}_{\rm at})$, and by the structure of $\boldsymbol{A}_{\mathcal{M}_{\rm bd},\mathcal{X}_{\rm at}}$ shown in (25), we have $\operatorname{supp}\left(\boldsymbol{A}_{\mathcal{M}_{\rm bd},\mathcal{X}_{\rm at}}(\boldsymbol{x}_{\rm at} - \boldsymbol{x}_{\natural{\rm at}})\right) \subseteq \mathcal{M}_{\rm at}$. Hence, we have that the unique optimal of $f_2(\boldsymbol{x}_{\rm bd},\boldsymbol{x}_{\rm at})$ satisfies $\hat{\boldsymbol{x}}_{\rm bd} = \boldsymbol{x}_{\natural{\rm bd}}$ for any given $\boldsymbol{x}_{\rm at}$. Since there are no bad data for $\boldsymbol{y}_{\mathcal{M}_{\rm sf}}$

and $\boldsymbol{y}_{\mathcal{M}_{\rm bd}}$ and moreover $\boldsymbol{A}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bd},\mathcal{X}_{\rm sf}\cup\mathcal{X}_{\rm bd}}$ has full column rank, the unique minimum of $f_1(\boldsymbol{x}_{\rm sf},\boldsymbol{x}_{\rm bd})$ is $(\hat{\boldsymbol{x}}_{\rm sf},\hat{\boldsymbol{x}}_{\rm bd})=(\boldsymbol{x}_{\natural{\rm sf}},\boldsymbol{x}_{\natural{\rm bd}})$. Therefore, for any given $\boldsymbol{x}_{\rm at}$, the unique optimal of $f_1(\boldsymbol{x}_{\rm sf},\boldsymbol{x}_{\rm bd})+f_2(\boldsymbol{x}_{\rm bd},\boldsymbol{x}_{\rm at})$ is $(\hat{\boldsymbol{x}}_{\rm sf},\hat{\boldsymbol{x}}_{\rm bd})=(\boldsymbol{x}_{\natural{\rm sf}},\boldsymbol{x}_{\natural{\rm bd}})$. Since $f_3(\boldsymbol{x}_{\rm at})$ does not depend on $(\boldsymbol{x}_{\rm sf},\boldsymbol{x}_{\rm bd})$, the unique optimal solution of ($\mathrm{S}^{(1)}$: ℓ_1) recovers the true solution.

The second approach is as follows. We can write the dual program of ($\mathrm{S}^{(1)}$: ℓ_1) as:

$$\max_{\boldsymbol{h}\in\mathbb{R}^{nm}} \boldsymbol{h}^\top\boldsymbol{y},\quad \text{subject to}\quad \boldsymbol{A}^\top\boldsymbol{h}=\boldsymbol{0},\quad \|\boldsymbol{h}\|_\infty\le 1. \tag{$\mathrm{S}^{(1)}$: ℓ_1-dual}$$

To show that $\left(\hat{\boldsymbol{x}}=\begin{bmatrix}\boldsymbol{x}_{\natural{\rm sf}}^\top & \boldsymbol{x}_{\natural{\rm bd}}^\top & \hat{\boldsymbol{x}}_{\rm at}\end{bmatrix}^\top, \hat{\boldsymbol{b}}=\begin{bmatrix}\boldsymbol{0}^\top & \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bd}}^\top & \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm at}}^\top\end{bmatrix}^\top\right)$ is the optimal solution of ($\mathrm{S}^{(1)}$: ℓ_1), we simply need to find a dual certificate $\boldsymbol{h}_\star=\begin{bmatrix}\boldsymbol{h}_{\mathcal{M}_{\rm sf}}^\top & \boldsymbol{h}_{\mathcal{M}_{\rm bd}}^\top & \boldsymbol{h}_{\mathcal{M}_{\rm at}}^\top\end{bmatrix}^\top$ that satisfies the KKT conditions:

$$\text{(dual feasibility)}\quad \boldsymbol{A}^\top\boldsymbol{h}_\star=\boldsymbol{0}, \tag{28}$$

$$\text{(stationarity)}\quad \boldsymbol{h}_\star\in\partial\|\hat{\boldsymbol{b}}\|_1. \tag{29}$$

Since by the reasoning above, $\begin{bmatrix}\boldsymbol{x}_{\natural{\rm sf}}^\top & \boldsymbol{x}_{\natural{\rm bd}}^\top\end{bmatrix}^\top$ is the unique optimal of the objective $f_1(\boldsymbol{x}_{\rm sf},\boldsymbol{x}_{\rm bd})+f_2(\boldsymbol{x}_{\rm bd},\boldsymbol{x}_{\rm at})$, it corresponds to a dual certificate $\begin{bmatrix}\boldsymbol{h}_{\mathcal{M}_{\rm sf}}^\top & \boldsymbol{h}_{\mathcal{M}_{\rm bd}}^\top\end{bmatrix}^\top$ such that

$$\boldsymbol{A}_{\mathcal{M}_{\rm sf},\mathcal{X}_{\rm sf}\cup\mathcal{X}_{\rm bd}}^\top\boldsymbol{h}_{\mathcal{M}_{\rm sf}}+\boldsymbol{A}_{\mathcal{M}_{\rm bd},\mathcal{X}_{\rm sf}\cup\mathcal{X}_{\rm bd}}^\top\boldsymbol{h}_{\mathcal{M}_{\rm bd}}=\boldsymbol{0}, \tag{30}$$

$$\|\boldsymbol{h}_{\mathcal{M}_{\rm sf}}\|_\infty\le 1,\|\boldsymbol{h}_{\mathcal{M}_{\rm bd}}\|_\infty\le 1. \tag{31}$$

Similarly, by the optimality of $\hat{\boldsymbol{x}}_{\rm at}$ for $f_3(\boldsymbol{x}_{\rm at})$, we can find a dual certificate such that:

$$\boldsymbol{A}_{\mathcal{M}_{\rm at},\mathcal{X}_{\rm at}}^\top\boldsymbol{h}_{\mathcal{M}_{\rm at}}=\boldsymbol{0},\quad \boldsymbol{h}_{\mathcal{M}_{\rm at}}\in\partial\|\hat{\boldsymbol{b}}_{\mathcal{M}_{\rm at}}\|_1. \tag{32}$$

Thus, by the structure of $\boldsymbol{A}$, the construction $\boldsymbol{h}_\star=\begin{bmatrix}\boldsymbol{h}_{\mathcal{M}_{\rm sf}}^\top & \boldsymbol{h}_{\mathcal{M}_{\rm bd}}^\top & \boldsymbol{h}_{\mathcal{M}_{\rm at}}^\top\end{bmatrix}^\top$ yields a dual certificate.

□

A key condition in Lemma 5 is the recovery of the boundary variables in the presence of arbitrary bad data that occur in the attacked region. This condition needs to be checked for every possible attack scenario, which is not useful to understand the graphical mutual incoherence in the general case. Instead, we propose the notion of graphical mutual incoherence in the main text, which provides a sufficient condition in this context. The technical definition is as follows.

Definition 6 (Local boundary variables and measurements)**.** *For each line ℓ that connects nodes i and j, let us distinguish the directions $i\to j$ and $j\to i$. For the direction $i\to j$, let i denote the node under attack and j be the node within the inner defense boundary. Accordingly, let $\mathcal{B}_{\rm bo}^{i\to j}$ denote the set of buses (other than i) that are directly connected to j as the outer boundary, $\mathcal{B}_{\rm bi}^{i\to j}=\{j\}$ be the one-bus inner boundary, and $\mathcal{B}_{\rm at}^{i\to j}=\{i\}$ be the one-bus attack set. Let $\mathcal{L}_{\rm bd}^{i\to j}$ represent the union of line ℓ and the set of lines that bridge $\mathcal{B}_{\rm bi}^{i\to j}$ and $\mathcal{B}_{\rm bo}^{i\to j}$. Define the "boundary variables" $\mathcal{X}_{\rm bd}^{i\to j}$ as the collection of voltage magnitudes $\{x_k^{\rm mg}\}_{k\in\mathcal{B}_{\rm bi}^{i\to j}\cup\mathcal{B}_{\rm bo}^{i\to j}}$ and variables $\{x_\eta^{\rm re},x_\eta^{\rm im}\}$ for the set of lines $\eta\in\mathcal{L}$ that connect the inner boundary j to nodes in the outer boundary $\mathcal{B}_{\rm bo}^{i\to j}$. Define the "boundary measurements" $\mathcal{M}_{\rm bd}^{i\to j}=\mathcal{M}_{{\rm bd}\checkmark}^{i\to j}\cup\mathcal{M}_{{\rm bd}\times}^{i\to j}$ as the collection of measurements 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 measurements that depend on both $\mathcal{X}_{\rm bd}^{i\to j}$ and variables $\{x_\ell^{\rm re},x_\ell^{\rm im}\}$ of the attacked line ℓ, denoted by $\mathcal{M}_{{\rm bd}\times}^{i\to j}$. The above terms can be similarly defined for the direction $j\to i$ by replacing $i\to j$ to $j\to i$ in the notations. Thus, for each line, we will have two sets of boundary variables and measurements.*

With the above notations, we can formally describe the graphical mutual incoherence.

Definition 7 (Graphical mutual incoherence)**.** *For each line $\{i,j\}_\ell \in \mathcal{L}$, define the graphical mutual incoherence $\alpha_{i\to j}$ along the direction $i \to j$ as the optimal objective value of the following minimax program:*

$$\alpha_{i\to j} = \max_{\boldsymbol{\xi}\in\{-1,+1\}^{n_\times^{i\to j}}} \min_{\alpha\in\mathbb{R},\boldsymbol{h}\in\mathbb{R}^{n_\checkmark^{i\to j}}} \alpha \tag{33a}$$

$$\text{subject to} \quad \boldsymbol{A}^\top_{\mathcal{M}^{i\to j}_{\mathrm{bd}\checkmark},\mathcal{X}^{i\to j}_{\mathrm{bd}}}\boldsymbol{h} + \boldsymbol{A}^\top_{\mathcal{M}^{i\to j}_{\mathrm{bd}\times},\mathcal{X}^{i\to j}_{\mathrm{bd}}}\boldsymbol{\xi} = \boldsymbol{0} \tag{33b}$$

$$\|\boldsymbol{h}\|_\infty \le \alpha, \tag{33c}$$

where $n_\checkmark^{i\to j} = |\mathcal{M}^{i\to j}_{\mathrm{bd}\checkmark}|$ and $n_\times^{i\to j} = |\mathcal{M}^{i\to j}_{\mathrm{bd}\times}|$ are the number of measurements in $\mathcal{M}^{i\to j}_{\mathrm{bd}\checkmark}$ and $\mathcal{M}^{i\to j}_{\mathrm{bd}\times}$, respectively, and $\mathcal{X}^{i\to j}_{\mathrm{bd}}$, $\mathcal{M}^{i\to j}_{\mathrm{bd}\checkmark}$ and $\mathcal{M}^{i\to j}_{\mathrm{bd}\times}$ are the boundary variables and measurement indices introduced in Def. 6. Similarly, we can define the backward graphical mutual incoherence $\alpha_{j\to i}$ by replacing $i\to j$ to $j\to i$ in (33)*. We adopt the measurement normalization convention in Def. 1.*

The name "mutual incoherence" originates from the compressed sensing literature [6, 14, 19, 16]. The closest condition proposed in the literature 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., $\boldsymbol{A}_{\mathcal{J}^c}$) [9, 10]. However, the graphical mutual incoherence (gMI) condition 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 of our condition on each instance of the bad data support $\mathcal{J}$. On the contrary, the conditions that exist in the study are all designed for global recovery, so they do not apply to the boundary defense mechanism. Also, even though the condition requires solving a minimax problem that is NP-hard in general, the computation becomes much more tractable because we limit the variables to those associated with a line. On the contrary, the conditions in the literature cannot be easily verified for large-scale systems. Moreover, we show that the condition in the literature is often more conservative than the graphical mutual incoherence.

Note that for the simple case where there are no lines between any two nodes in $\mathcal{N}_{\mathcal{B}_{\mathrm{bi}}}$, we can extend the above definition to treat each node in $\mathcal{N}_{\mathcal{B}_{\mathrm{bi}}}$ separately. Due to the localized nature, this condition is much weaker than the mutual incoherence condition introduced in [9].

Proposition 8 (Global mutual incoherence is more conservative than graphical mutual incoherence)**.** *For each line ℓ and the corresponding partitions of measurements $\mathcal{M}^{i\to j}_{\mathrm{bd}\checkmark}$, $\mathcal{M}^{i\to j}_{\mathrm{bd}\times}$ and variables $\mathcal{X}^{i\to j}_{\mathrm{bd}}$, let*

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

be the mutual incoherence metric defined in [9, 10], where $\boldsymbol{A}^+_{\mathcal{J}} = (\boldsymbol{A}^\top_{\mathcal{J}}\boldsymbol{A}_{\mathcal{J}})^{-1}\boldsymbol{A}^\top_{\mathcal{J}}$ denotes the pseudo-inverse. Then, it holds that $\rho(\mathcal{M}^{i\to j}_{\mathrm{bd}\times}) \ge \alpha_{i\to j}$.

Proof. Notice that the graphical mutual incoherence can be written as

$$\alpha_{i\to j} = \max_{\boldsymbol{\xi}\in\{-1,+1\}^{n_\times^{i\to j}}} \min_{\alpha\in\mathbb{R},\boldsymbol{h}\in\mathbb{R}^{n_\checkmark^{i\to j}}} \|\boldsymbol{h}\|_\infty \tag{34a}$$

$$\text{subject to} \quad \boldsymbol{A}^\top_{\mathcal{M}^{i\to j}_{\mathrm{bd}\checkmark},\mathcal{X}^{i\to j}_{\mathrm{bd}}}\boldsymbol{h} + \boldsymbol{A}^\top_{\mathcal{M}^{i\to j}_{\mathrm{bd}\times},\mathcal{X}^{i\to j}_{\mathrm{bd}}}\boldsymbol{\xi} = \boldsymbol{0}. \tag{34b}$$

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 for the inner optimization, and

$$\max_{\boldsymbol{\xi}\in\{-1,+1\}^{n^{i\to j}_{\times}}} \|\hat{\boldsymbol{h}}(\boldsymbol{\xi})\|_\infty = \rho(\mathcal{M}^{i\to j}_{\mathrm{bd}\times}), \tag{35}$$

the proof is immediately concluded. □

A key step in establishing the validity of the boundary defense mechanism is to ensure that local defense is sufficient to guard against attacks when solving the problem globally.

Lemma 9 (Local property implies global property)**.** *Given $\mathcal{B}_{\mathrm{at}}, \mathcal{B}_{\mathrm{bi}}, \mathcal{B}_{\mathrm{bo}}$, and $\mathcal{B}_{\mathrm{sf}}$ and the associated set partitioning (c.f., Def. 3), let $\boldsymbol{A}^\circ = \begin{bmatrix} \boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}},\mathcal{X}_{\mathrm{sf}}} & \boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}},\mathcal{X}_{\mathrm{bd}}} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\mathrm{bo}},\mathcal{X}_{\mathrm{bd}}} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\mathrm{bi}},\mathcal{X}_{\mathrm{bd}}} \end{bmatrix}$. If $\alpha_{i\to j} \le 1-\gamma$ and $\gamma > 0$ for all $\{i,j\} \in \mathcal{L}_{\mathrm{at}\cap\mathrm{bi}}$ such that $i \in \mathcal{B}_{\mathrm{at}}$ and $j \in \mathcal{B}_{\mathrm{bi}}$, then for any $\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bi}}} \in [-1,1]^{|\mathcal{M}_{\mathrm{bi}}|}$, there exists an $\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{sf}}\cup\mathcal{M}_{\mathrm{bo}}}$ with the properties $\|\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{sf}}\cup\mathcal{M}_{\mathrm{bo}}}\|_\infty \le 1-\gamma$ and*

$$\boldsymbol{A}^{\circ\top}_{\mathcal{M}_{\mathrm{sf}}\cup\mathcal{M}_{\mathrm{bo}}} \hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{sf}}\cup\mathcal{M}_{\mathrm{bo}}} + \boldsymbol{A}^{\circ\top}_{\mathcal{M}_{\mathrm{bi}}} \hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bi}}} = \boldsymbol{0}. \tag{36}$$

Proof. First, we show that a sufficient condition for the existence of $\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{sf}}\cup\mathcal{M}_{\mathrm{bo}}} = \begin{bmatrix} \hat{\boldsymbol{h}}^\top_{\mathcal{M}_{\mathrm{sf}}} & \hat{\boldsymbol{h}}^\top_{\mathcal{M}_{\mathrm{bo}}} \end{bmatrix}^\top$ such that $\|\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{sf}}\cup\mathcal{M}_{\mathrm{bo}}}\|_\infty \le 1-\gamma$ and (36) is satisfied is that for any $\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bi}}}$, there exists an $\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bo}}}$ such that $\|\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bo}}}\|_\infty \le 1-\gamma$ and

$$\boldsymbol{A}^{\top}_{\mathcal{M}_{\mathrm{bo}},\mathcal{X}_{\mathrm{bd}}} \hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bo}}} + \boldsymbol{A}^{\top}_{\mathcal{M}_{\mathrm{bi}},\mathcal{X}_{\mathrm{bd}}} \hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bi}}} = \boldsymbol{0}. \tag{37}$$

This is immediate by simply choosing $\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{sf}}\cup\mathcal{M}_{\mathrm{bo}}} = \begin{bmatrix} \boldsymbol{0}^\top & \hat{\boldsymbol{h}}^\top_{\mathcal{M}_{\mathrm{bo}}} \end{bmatrix}^\top$. In what follows, we prove (37) by induction. The induction rule is as follows: we start by arbitrarily choosing one line $\{i,j\} \in \mathcal{L}_{\mathrm{at}\cap\mathrm{bi}}$, where $i \in \mathcal{B}_{\mathrm{at}}$ and $j \in \mathcal{B}_{\mathrm{bi}}$, and initialize the measurement set $\mathcal{M}^{(1)}_{\mathrm{bo}} \coloneqq \mathcal{M}^{i\to j}_{\mathrm{bd}\checkmark}$, $\mathcal{M}^{(1)}_{\mathrm{bi}} \coloneqq \mathcal{M}^{i\to j}_{\mathrm{bd}\times}$ and the variable set $\mathcal{X}^{(1)}_{\mathrm{bd}} \coloneqq \mathcal{X}^{i\to j}_{\mathrm{bd}}$. For each step k, we add a new line $\{f,t\} \in \mathcal{L}_{\mathrm{at}\cap\mathrm{bi}}$ and the associated measurements and variables to $\mathcal{M}^{(k)}_{\mathrm{bo}}$, $\mathcal{M}^{(k)}_{\mathrm{bi}}$ and $\mathcal{X}^{(k)}_{\mathrm{bd}}$, respectively. After the inclusion of all the lines in $\mathcal{L}_{\mathrm{at}\cap\mathrm{bi}}$, we should obtain the set $\mathcal{M}_{\mathrm{bo}}$, $\mathcal{M}_{\mathrm{bi}}$ and $\mathcal{X}_{\mathrm{bd}}$. In each step, we check whether there exists a vector $\hat{\boldsymbol{h}}_{\mathcal{M}^{(k)}_{\mathrm{bo}}}$ such that $\|\hat{\boldsymbol{h}}_{\mathcal{M}^{(k)}_{\mathrm{bo}}}\|_\infty \le 1-\gamma$ and

$$\boldsymbol{A}^{\top}_{\mathcal{M}^{(k)}_{\mathrm{bo}},\mathcal{X}^{(k)}_{\mathrm{bd}}} \hat{\boldsymbol{h}}_{\mathcal{M}^{(k)}_{\mathrm{bo}}} + \boldsymbol{A}^{\top}_{\mathcal{M}^{(k)}_{\mathrm{bi}},\mathcal{X}^{(k)}_{\mathrm{bd}}} \hat{\boldsymbol{h}}_{\mathcal{M}^{(k)}_{\mathrm{bi}}} = \boldsymbol{0}. \tag{38}$$

The base case for $k=1$ follows directly from the condition that $\alpha_{i\to j} \le 1-\gamma$. For any $k \ge 1$, let $\{f,t\} \in \mathcal{L}_{\mathrm{at}\cap\mathrm{bi}}$ denote the line to be added, where $f \in \mathcal{M}_{\mathrm{at}}$ and $t \in \mathcal{M}_{\mathrm{bi}}$. There are 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 attack node f with one (or more) of the lines already added (note that by definition, the new line cannot share the inner boundary node t 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 $\mathcal{B}^{i\to j}_{\mathrm{bo}}$ are connected to one or more of the nodes in the inner boundaries of lines that have already been added; and/or **b)** one or more of the nodes in the outer boundary of the lines that have already been added are connected to t; or **c)** none of the above (note that by definition, there are no lines within the inner boundary region). We need to consider all the combinations between the three cases and the three events to show that (38) holds in all scenarios. Fortunately, all the combinations can be reduced to two typical scenarios, where the proofs can be directly applied. We consider these scenarios now.

The first scenario applies to Cases 1c and 2c, where $\mathcal{M}_{\rm bo}^{(k+1)} = \mathcal{M}_{\rm bo}^{(k)} \cup \mathcal{M}_{{\rm bd}\checkmark}^{f\to t}$, $\mathcal{M}_{\rm bi}^{(k+1)} = \mathcal{M}_{\rm bi}^{(k)} \cup \mathcal{M}_{{\rm bd}\times}^{f\to t}$, $\mathcal{X}_{\rm bd}^{(k+1)} = \mathcal{X}_{\rm bd}^{(k)} \cup \mathcal{X}_{\rm bd}^{f\to t}$, $\mathcal{M}_{\rm bo}^{(k)} \cap \mathcal{M}_{{\rm bd}\checkmark}^{f\to t} = \emptyset$, $\mathcal{M}_{\rm bi}^{(k)} \cap \mathcal{M}_{{\rm bd}\times}^{f\to t} = \emptyset$, and $\mathcal{X}_{\rm bd}^{(k)} \cap \mathcal{X}_{\rm bd}^{f\to t} = \emptyset$. Therefore, for any given $\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}^{(k+1)}} = \left[\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}^{(k)}}^\top \quad \hat{\boldsymbol{\xi}}^\top\right]^\top$ with $\|\hat{\boldsymbol{\xi}}\|_\infty \le 1$, we can always find $\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bo}^{(k+1)}} = \left[\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bo}^{(k)}}^\top \quad \hat{\boldsymbol{h}}^\top\right]^\top$, where $\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bo}^{(k)}}$ is given by (38) and $\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bo}^{(k)}}$ is given by (33), and $\|\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bo}^{(k+1)}}\|_\infty \le 1-\gamma$ by definition.

The second scenario applies to Cases 1a, 1b, 2a and 2b. Let $\tilde{\mathcal{N}}_{\rm bo}$ be the set of nodes in the outer boundary shared by the new line $\mathcal{B}_{\rm bo}^{f\to t}$ and those of the lines that have been added. Then, we have $\mathcal{M}_{\rm bo}^{(k+1)} = \mathcal{M}_{\rm bo}^{(k)} \cup \mathcal{M}_{{\rm bd}\checkmark}^{f\to t}$, $\mathcal{M}_{\rm bi}^{(k+1)} = \mathcal{M}_{\rm bi}^{(k)} \cup \mathcal{M}_{{\rm bd}\times}^{f\to t}$, $\mathcal{X}_{\rm bd}^{(k+1)} = \mathcal{X}_{\rm bd}^{(k)} \cup \mathcal{X}_{\rm bd}^{f\to t}$, where $\mathcal{M}_{\rm bo}^{(k)} \cap \mathcal{M}_{{\rm bd}\checkmark}^{f\to t}$ is the set of voltage magnitude measurements of nodes in $\tilde{\mathcal{N}}_{\rm bo}$, $\mathcal{M}_{\rm bi}^{(k)} \cap \mathcal{M}_{{\rm bd}\times}^{f\to t} = \emptyset$, and $\mathcal{X}_{\rm bd}^{(k)} \cap \mathcal{X}_{\rm bd}^{f\to t}$ is the set of voltage magnitude variables of nodes in $\tilde{\mathcal{N}}_{\rm bo}$. For any given $\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}^{(k)}}$ and $\hat{\boldsymbol{\xi}}^\top$, we can always find $\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bo}^{(k)}}$ and $\hat{\boldsymbol{h}}^\top$, where $\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bo}^{(k)}}$ is given by (38) and $\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bo}^{(k)}}$ is given by (33). Let $\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bo}^{(k)}}$ 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 (38) for any given $\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}^{(k+1)}} = \left[\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}^{(k)}}^\top \quad \hat{\boldsymbol{\xi}}^\top\right]^\top$ (note that the voltage magnitude measurement in the calculation of graphical mutual incoherence is normalized by 1, but it is weighted by the degree of each node in the actual estimation algorithm, c.f., Def. 1). 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. □

Lemma 9 implies that as long as all the graphical mutual incoherence metrics are bounded away from 1, we have a desirable property in terms of defending against bad data on the boundary. This is formalized in the following theorem.

Theorem 10. *Consider the measurements $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_\natural + \boldsymbol{b}_\natural$, where $\mathrm{supp}(\boldsymbol{b}_\natural) \subseteq \mathcal{M}_{\rm at}$. Suppose that for the given partitioning of the network as $\mathcal{B}_{\rm at}, \mathcal{B}_{\rm bi}, \mathcal{B}_{\rm bo}$, and $\mathcal{B}_{\rm sf}$, the following conditions hold:*

- *(Full column rank for the safe and boundary region) $\boldsymbol{A}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bd},\mathcal{X}_{\rm sf}\cup\mathcal{X}_{\rm bd}}$ and*

$$\boldsymbol{Q}_{\mathcal{M}_{\rm bd},\mathcal{X}_{\rm bd}} = \left[\boldsymbol{A}_{\mathcal{M}_{\rm bd},\mathcal{X}_{\rm bd}} \quad \boldsymbol{I}_{\mathcal{M}_{\rm bi}}^{(|\mathcal{M}_{\rm bd}|)\top}\right]$$

  have full column rank.

- *(Localized mutual incoherence) for all lines $\{i,j\} \in \mathcal{L}_{{\rm at}\cap{\rm bi}}$ that bridge the attacked region and the inner boundary, where $i \in \mathcal{B}_{\rm at}$, $j \in \mathcal{B}_{\rm bi}$, we have $\alpha_{i\to j} \le 1-\gamma$ for some $\gamma > 0$.*

Then, the solution to ($\mathrm{S}^{(1)}$: ℓ_1)*, denoted as $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{b}})$, uniquely recovers the true state outside the attacked region (i.e., $\hat{\boldsymbol{x}}_{\rm sf} = \boldsymbol{x}_{\natural{\rm sf}}$ and $\hat{\boldsymbol{x}}_{\rm bd} = \boldsymbol{x}_{\natural{\rm bd}}$). Furthermore, the state estimation by* ($\mathrm{S}^{(2)}$: ℓ_2) *recovers the true state for the unaffected region (i.e., $\hat{v}_k = v_k$ for $k \in \mathcal{B}_{\rm sf} \cup \mathcal{B}_{\rm bd}$).*

Proof. To prove the claim, we simply need to show that for an arbitrary $\boldsymbol{b}_\star$ with its support limited to the inner boundary $\mathrm{supp}(\boldsymbol{b}_\star) \subseteq \mathcal{M}_{\mathrm{bi}}$, the solution $\hat{\boldsymbol{x}}_{\mathrm{bd}} \in \mathcal{X}_{\mathrm{bd}}$ to the program

$$\min_{\boldsymbol{x}_{\mathrm{bd}}} \|\boldsymbol{z}_{\mathcal{M}_{\mathrm{bd}}} - \boldsymbol{A}_{\mathcal{M}_{\mathrm{bd}},\mathcal{X}_{\mathrm{bd}}} \boldsymbol{x}_{\mathrm{bd}}\|_1 \tag{39}$$

is unique and satisfies $\hat{\boldsymbol{x}}_{\mathrm{bd}} = \boldsymbol{x}_{\natural\mathrm{bd}}$, where $\boldsymbol{z}_{\mathcal{M}_{\mathrm{bd}}} = \boldsymbol{A}_{\mathcal{M}_{\mathrm{bd}},\mathcal{X}_{\mathrm{bd}}} \boldsymbol{x}_{\natural\mathrm{bd}} + \boldsymbol{b}_\star$. To show this, we obtain the dual program:

$$\max_{\boldsymbol{h}_{\mathcal{M}_{\mathrm{bd}}}} \boldsymbol{h}_{\mathcal{M}_{\mathrm{bd}}}^\top \boldsymbol{z}_{\mathcal{M}_{\mathrm{bd}}}, \quad \text{subject to} \quad \boldsymbol{A}_{\mathcal{M}_{\mathrm{bd}},\mathcal{X}_{\mathrm{bd}}}^\top \boldsymbol{h}_{\mathcal{M}_{\mathrm{bd}}} = \boldsymbol{0}, \|\boldsymbol{h}_{\mathcal{M}_{\mathrm{bd}}}\|_\infty \leq 1. \tag{40}$$

Our goal is to find a dual certificate $\boldsymbol{h}_{\star\mathcal{M}_{\mathrm{bd}}}$ that satisfies the KKT conditions:

$$\text{(dual feasibility)} \quad \boldsymbol{A}_{\mathcal{M}_{\mathrm{bd}},\mathcal{X}_{\mathrm{bd}}}^\top \boldsymbol{h}_{\star\mathcal{M}_{\mathrm{bd}}} = \boldsymbol{0}, \tag{41}$$

$$\text{(stationarity)} \quad \boldsymbol{h}_{\star\mathcal{M}_{\mathrm{bd}}} \in \partial \|\boldsymbol{b}_\star\|_1. \tag{42}$$

By the limited support assumption, we need to find a vector $\boldsymbol{h}_\star$ such that $\boldsymbol{h}_{\star\mathcal{M}_{\mathrm{bi}}} = \mathrm{sign}(\boldsymbol{b}_{\star\mathcal{M}_{\mathrm{bi}}})$ and $\|\boldsymbol{h}_{\star\mathcal{M}_{\mathrm{bo}}}\|_\infty \leq 1$. By the mutual incoherence condition and Lemma 9, we can always find $\boldsymbol{h}_{\star\mathcal{M}_{\mathrm{bo}}}$ that satisfies (41) for any given $\boldsymbol{h}_{\star\mathcal{M}_{\mathrm{bi}}}$ and $\|\boldsymbol{h}_{\star\mathcal{M}_{\mathrm{bo}}}\| \leq 1 - \gamma < 1$. Thus, this certifies the optimality of $(\boldsymbol{x}_{\natural\mathrm{bd}}, \boldsymbol{b}_\star)$ for (40).

To show that $(\boldsymbol{x}_{\natural\mathrm{bd}}, \boldsymbol{b}_\star)$ is the unique optimal solution, let $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{b}})$ be an arbitrary feasible point of (39) that is different from $(\boldsymbol{x}_{\natural\mathrm{bd}}, \boldsymbol{b}_\star)$. Due to the lower eigenvalue condition, the matrix $\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bd}},\mathcal{X}_{\mathrm{bd}}} \coloneqq \begin{bmatrix} \boldsymbol{A}_{\mathcal{M}_{\mathrm{bd}},\mathcal{X}_{\mathrm{bd}}} & \boldsymbol{I}_{\mathcal{M}_{\mathrm{bi}}}^{(|\mathcal{M}_{\mathrm{bd}}|)\top} \end{bmatrix}$ has full column rank. By letting $\tilde{\mathcal{J}} = \mathrm{supp}(\tilde{\boldsymbol{b}})$, the set $\tilde{\mathcal{J}}$ can not be equal to or be a subset of $\mathcal{M}_{\mathrm{bi}}$, because otherwise, from $\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bd}},\mathcal{X}_{\mathrm{bd}}} \begin{bmatrix} \boldsymbol{x}_{\natural\mathrm{bd}} \\ \boldsymbol{b}_\star \end{bmatrix} = \boldsymbol{Q}_{\mathcal{M}_{\mathrm{bd}},\mathcal{X}_{\mathrm{bd}}} \begin{bmatrix} \tilde{\boldsymbol{x}} \\ \tilde{\boldsymbol{b}} \end{bmatrix} = \boldsymbol{z}_{\mathcal{M}_{\mathrm{bd}}}$, we must have $\begin{bmatrix} \boldsymbol{x}_{\natural\mathrm{bd}} \\ \boldsymbol{b}_\star \end{bmatrix} = \begin{bmatrix} \tilde{\boldsymbol{x}} \\ \tilde{\boldsymbol{b}} \end{bmatrix}$, which is contradictory to the assumption. Let $\tilde{\mathcal{J}}_c = \tilde{\mathcal{J}} \setminus \mathcal{M}_{\mathrm{bi}}$; then,

$$\begin{aligned}
\|\boldsymbol{b}_\star\|_1 &= \boldsymbol{h}_{\star\mathcal{M}_{\mathrm{bd}}}^\top \boldsymbol{z}_{\mathcal{M}_{\mathrm{bd}}} && (43)\\
&= \boldsymbol{h}_{\star\mathcal{M}_{\mathrm{bd}}}^\top (\boldsymbol{A}_{\mathcal{M}_{\mathrm{bd}},\mathcal{X}_{\mathrm{bd}}} \tilde{\boldsymbol{x}} + \boldsymbol{I}_{\tilde{\mathcal{J}}_c}^\top \tilde{\boldsymbol{b}}_{\tilde{\mathcal{J}}_c} + \boldsymbol{I}_{\mathcal{M}_{\mathrm{bi}}}^\top \tilde{\boldsymbol{b}}_{\mathcal{M}_{\mathrm{bi}}}) && (44)\\
&= \boldsymbol{h}_{\star\tilde{\mathcal{J}}_c}^\top \tilde{\boldsymbol{b}}_{\tilde{\mathcal{J}}_c} + \boldsymbol{h}_{\star\mathcal{M}_{\mathrm{bi}}}^\top \tilde{\boldsymbol{b}}_{\mathcal{M}_{\mathrm{bi}}} && (45)\\
&\leq \|\boldsymbol{h}_{\star\tilde{\mathcal{J}}_c}\|_\infty \|\tilde{\boldsymbol{b}}_{\tilde{\mathcal{J}}_c}\|_1 + \|\boldsymbol{h}_{\star\mathcal{M}_{\mathrm{bi}}}\|_\infty \|\tilde{\boldsymbol{b}}_{\mathcal{M}_{\mathrm{bi}}}\|_1 && (46)\\
&< \|\tilde{\boldsymbol{b}}_{\tilde{\mathcal{J}}_c}\|_1 + \|\tilde{\boldsymbol{b}}_{\mathcal{M}_{\mathrm{bi}}}\|_1 && (47)\\
&= \|\tilde{\boldsymbol{b}}\|_1, && (48)
\end{aligned}$$

where (43) is due to the strong duality between (39) and (40), (44) is due to the primal feasibility of $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{b}})$, (45) is due to the dual feasibility condition (41), (46) is due to the Hölder inequality, and (47) is due to the strict feasibility of $\boldsymbol{h}_\star$. Thus, we have shown the uniqueness of the optimal solution $(\boldsymbol{x}_{\natural\mathrm{bd}}, \boldsymbol{b}_\star)$. Together with Lemma 9, we have proved the theorem.

□

This result can be used to certify robustness under different attack scenarios. For example, if there is a topological error caused by line mis-specification, say $\ell = (i, j)$, we can treat the two ends of the line as the attacked nodes, i.e., $\mathcal{N}_{\mathrm{at}} = \{i, j\}$, treat the adjacent nodes to them as inner boundary $\mathcal{N}_{\mathrm{bi}}$, and treat the adjacent nodes to inner boundary as outer boundary $\mathcal{N}_{\mathrm{bo}}$. As long as the graphical mutual incoherence for the lines surrounding the attacked nodes are less than 1, one can identify this gross injection error and

thus the topological mistake. We can extend the analysis to the case where the measurements have both sparse bad data and dense noise. In this case, we need to solve a program that combines quadratic loss with absolute value loss. The guarantees now depend on the distribution of the dense noise.

Theorem 11 (Robust SE with ($\mathrm{S}^{(1)}$: $\ell_2\ell_1$))**.** *Consider the measurements $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_\natural + \boldsymbol{w}_\natural + \boldsymbol{b}_\natural$, where $\boldsymbol{w}_\natural$ has independent entries with zero mean and subgaussian parameter σ and $\mathrm{supp}(\boldsymbol{b}_\natural) \subseteq \mathcal{M}_{\mathrm{at}}$. Suppose that the rows of $\boldsymbol{A}$ are normalized (c.f., Def. 1), and the regularization parameter λ is chosen such that*

$$\lambda > \frac{2}{n_m \gamma}\sqrt{2\sigma^2 \log n_m}. \tag{49}$$

In addition, suppose that for the given partitioning of the network, i.e. $\mathcal{B}_{\mathrm{at}}, \mathcal{B}_{\mathrm{bi}}, \mathcal{B}_{\mathrm{bo}}$, and $\mathcal{B}_{\mathrm{sf}}$, the following conditions hold:

- *(Full column rankness) $\boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}}\cup\mathcal{M}_{\mathrm{bd}},\mathcal{X}_{\mathrm{sf}}\cup\mathcal{X}_{\mathrm{bd}}}$ and*

$$\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bd}},\mathcal{X}_{\mathrm{bd}}} = \begin{bmatrix} \boldsymbol{A}_{\mathcal{M}_{\mathrm{bd}},\mathcal{X}_{\mathrm{bd}}} & \boldsymbol{I}_{\mathcal{M}_{\mathrm{bi}}}^{(|\mathcal{M}_{\mathrm{bd}}|)\top} \end{bmatrix}$$

  have full column rank.

- *(Localized mutual incoherence) for all lines $\{i,j\} \in \mathcal{L}_{\mathrm{at}\cap\mathrm{bi}}$ that bridge the attacked region and the inner boundary, where $i \in \mathcal{B}_{\mathrm{at}}$, $j \in \mathcal{B}_{\mathrm{bi}}$, we have $\alpha_{i\to j} \leq 1-\gamma$ for some $\gamma > 0$.*

Then, the following properties hold for the solution to ($\mathrm{S}^{(1)}$: $\ell_2\ell_1$), denoted as $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{b}})$:

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

2. *(Large bad data detection) Let $\boldsymbol{A}^\circ := \begin{bmatrix} \boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}},\mathcal{X}_{\mathrm{sf}}} & \boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}},\mathcal{X}_{\mathrm{bd}}} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\mathrm{bo}},\mathcal{X}_{\mathrm{bd}}} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\mathrm{bi}},\mathcal{X}_{\mathrm{bd}}} \end{bmatrix}$ and $\boldsymbol{Q}^\circ_{\mathcal{M}_{\mathrm{bi}}} = \begin{bmatrix} \boldsymbol{A}^\circ & \boldsymbol{I}^\top_{\mathcal{M}_{\mathrm{bi}}} \end{bmatrix}$, and*

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

   be a threshold value, 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 error at the boundary. 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*

$$\|\boldsymbol{x}_{\natural,\mathcal{X}_{\mathrm{sf}}\cup\mathcal{X}_{\mathrm{bd}}} - \hat{\boldsymbol{x}}_{\mathcal{X}_{\mathrm{sf}}\cup\mathcal{X}_{\mathrm{bd}}}\|_2 \leq t\frac{\sqrt{|\mathcal{X}_{\mathrm{sf}}| + |\mathcal{X}_{\mathrm{bd}}| + |\mathcal{M}_{\mathrm{bi}}|}}{C_{min}} + n_m\lambda\|\boldsymbol{I}_x(\boldsymbol{Q}^{\circ\top}_{\mathcal{M}_{\mathrm{bi}}}\boldsymbol{Q}^\circ_{\mathcal{M}_{\mathrm{bi}}})^{-1}\boldsymbol{I}_b^\top\|_{\infty,2}$$

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

Despite the difference in measurement assumptions (i.e., existence of dense noise $\boldsymbol{w}$) and estimation algorithms (i.e., (S$^{(1)}$: ℓ_1) or (S$^{(1)}$: $\ell_2\ell_1$)), it is remarkable that the boundary defense conditions in Theorems 10 and 11 are coincident. In the case of negligible dense noise, a deterministic boundary defense is achieved. With the presence of dense noise, it is no longer possible to have deterministic guarantees; however, Theorem 11 indicates that with a proper selection of the penalty coefficient λ, one can avoid false detection of bad data in the unaffected region (part 1), detect bad data with magnitudes greater than a threshold in the attacked region (part 2), and achieve estimation within bounded error margin for states within the unaffected region. Furthermore, both the bad data threshold and the error bound decrease with stronger mutual incoherence condition and lower-eigenvalue condition. The proof of the theorem is provided in Section 5.1.

3.2 Boundary defense for second-order cone programming

In this section, we extend the analysis of boundary defense to the case where we perform state estimation with the additional second-order cone constraints.

Lemma 12 (Boundary defense stops error propagation with SOCP)**.** *Suppose that there is no dense measurement noise (i.e., $\boldsymbol{w} = \boldsymbol{0}$), and the bad data are confined within $\mathcal{M}_{\rm at}$, i.e., $\mathrm{supp}(\boldsymbol{b}_\natural) \subseteq \mathcal{M}_{\rm at}$. Let $\mathcal{K}_{\rm bd}$ and $\mathcal{K}_{\rm at}$ be the subsets of SOC constraints $\mathcal{K}$ restricted to variables $\boldsymbol{x}_{\rm bd}$ and $\boldsymbol{x}_{\rm at}$, respectively, and let*

$$\tilde{\mathcal{K}}_{\rm at}(\hat{\boldsymbol{x}}_{\rm bd}) = \Bigg\{ \boldsymbol{x}_{\rm at} \Bigg| \begin{bmatrix} x_i^{mg} & x_\ell^{re} + jx_\ell^{im} \\ x_\ell^{re} - jx_\ell^{im} & x_j^{mg} \end{bmatrix} \succeq 0, \quad \forall \ell := (i,j) \in \mathcal{L}_{\rm at} \cup \mathcal{L}_{\rm at\cap bi},\ \textit{where } x_i^{mg} = \hat{x}_i^{mg} \quad \forall i \in \mathcal{B}_{\rm bi} \Bigg\},$$

be the confined feasible set for $\boldsymbol{x}_{\rm at}$, which fixes the boundary variables $\hat{\boldsymbol{x}}_{\rm bd}$ in the SOCP constraints. Assume that for an arbitrary $\boldsymbol{b}_{\star\mathcal{M}_{\rm bd}}$ with its support limited to the inner boundary, i.e. $\mathrm{supp}(\boldsymbol{b}_{\star\mathcal{M}_{\rm bd}}) \subseteq \mathcal{M}_{\rm bi}$, the solution $\hat{\boldsymbol{x}}_{\rm bd} \in \mathcal{X}_{\rm bd}$ to the program

$$\min_{\boldsymbol{x}_{\rm bd} \in \mathcal{K}_{\rm bd}} \|\boldsymbol{z}_{\mathcal{M}_{\rm bd}} - \boldsymbol{A}_{\mathcal{M}_{\rm bd},\mathcal{X}_{\rm bd}} \boldsymbol{x}_{\rm bd}\|_1, \tag{50}$$

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

$$\min_{\boldsymbol{x}_{\rm at} \in \mathcal{K}_{\rm at}} \|\boldsymbol{y}_{\mathcal{M}_{\rm at}} - \boldsymbol{A}_{\mathcal{M}_{\rm at},\mathcal{X}_{\rm at}} \boldsymbol{x}_{\rm at}\|_1, \tag{51}$$

also satisfies that $\hat{\boldsymbol{x}}_{\rm at} \in \tilde{\mathcal{K}}_{\rm at}(\boldsymbol{x}_{\natural{\rm bd}})$. Then, the solution $\hat{\boldsymbol{x}}$ to (S$^{(1)}$: ℓ_1-$\mathcal{K}$) *satisfies $\hat{\boldsymbol{x}}_{\rm bd} = \boldsymbol{x}_{\natural{\rm bd}}$ and $\hat{\boldsymbol{x}}_{\rm sf} = \boldsymbol{x}_{\natural{\rm sf}}$.*

Proof. To show that $\left(\hat{\boldsymbol{x}} = \begin{bmatrix} \boldsymbol{x}_{\natural{\rm sf}}^\top & \boldsymbol{x}_{\natural{\rm bd}}^\top & \hat{\boldsymbol{x}}_{\rm at}^\top \end{bmatrix}^\top, \hat{\boldsymbol{b}} = \begin{bmatrix} \boldsymbol{0}^\top & \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bd}}^\top & \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm at}}^\top \end{bmatrix}^\top \right)$ is the optimal solution of (S$^{(1)}$: ℓ_1-$\mathcal{K}$), we simply need to find a dual certificate $\left(\boldsymbol{h}_\star = \begin{bmatrix} \boldsymbol{h}_{\mathcal{M}_{\rm sf}}^\top & \boldsymbol{h}_{\mathcal{M}_{\rm bd}}^\top & \boldsymbol{h}_{\mathcal{M}_{\rm at}}^\top \end{bmatrix}^\top, \{\nu_\ell, \boldsymbol{u}_\ell\}_{\ell \in \mathcal{L}} \right)$ that satisfies the KKT conditions:

(stationarity) $$\boldsymbol{h}_\star \in \partial \|\hat{\boldsymbol{b}}\|_1, \tag{52}$$

(dual feasibility) $$\boldsymbol{A}^\top \boldsymbol{h}_\star + \sum_{\ell \in \mathcal{L}} \left(\nu_\ell \boldsymbol{c}_\ell + \boldsymbol{D}_\ell^\top \boldsymbol{u}_\ell \right) = \boldsymbol{0}; \qquad \nu_\ell \geq \|\boldsymbol{u}_\ell\|_2, \quad \forall \ell \in \mathcal{L}, \tag{53}$$

(complementary slackness) $$\nu_\ell \boldsymbol{c}_\ell^\top \hat{\boldsymbol{x}} + \boldsymbol{u}_\ell^\top \boldsymbol{D}_\ell \hat{\boldsymbol{x}} = \boldsymbol{0}, \quad \forall \ell \in \mathcal{L}, \tag{54}$$

For a given $\boldsymbol{x}_{\rm bd} = \boldsymbol{x}_{\natural \rm bd}$, let $\hat{\boldsymbol{x}}_{\rm sf}$ be the optimal solution to

$$\min_{\boldsymbol{x}_{\rm sf} \in \mathcal{K}_{\rm sf}} \|\boldsymbol{y}_{\mathcal{M}_{\rm sf}} - \boldsymbol{A}_{\mathcal{M}_{\rm sf},\mathcal{X}_{\rm sf}} \boldsymbol{x}_{\rm sf} - \boldsymbol{A}_{\mathcal{M}_{\rm sf},\mathcal{X}_{\rm bd}} \boldsymbol{x}_{\natural \rm bd}\|_1,$$

where $\mathcal{K}_{\rm sf}$ is set of all SOCP constraints that involve at least one variable in $\mathcal{X}_{\rm sf}$. By the lower eigenvalue condition, $\hat{\boldsymbol{x}}_{\rm sf} = \boldsymbol{x}_{\natural \rm sf}$ is the unique optimal solution. Since for a given $\hat{\boldsymbol{x}}_{\rm at} \in \tilde{\mathcal{K}}_{\rm at}(\boldsymbol{x}_{\natural \rm bd})$, $\hat{\boldsymbol{x}}_{\rm bd} = \boldsymbol{x}_{\natural \rm bd}$ is the unique optimal of (50), we can conclude that $\begin{bmatrix} \boldsymbol{x}_{\natural \rm sf}^\top & \boldsymbol{x}_{\natural \rm bd}^\top \end{bmatrix}^\top$ is the unique optimal of

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

which corresponds to a dual certificate $\Big(\begin{bmatrix} \boldsymbol{h}_{\mathcal{M}_{\rm sf}}^\top & \boldsymbol{h}_{\mathcal{M}_{\rm bd}}^\top \end{bmatrix}^\top, \{\nu_\ell, \boldsymbol{u}_\ell\}_{\ell \in \mathcal{L}_{\rm sf} \cup \mathcal{L}_{\rm bd}} \Big)$ such that

$$\boldsymbol{A}_{\mathcal{M}_{\rm sf},\mathcal{X}_{\rm sf} \cup \mathcal{X}_{\rm bd}}^\top \boldsymbol{h}_{\mathcal{M}_{\rm sf}} + \boldsymbol{A}_{\mathcal{M}_{\rm bd},\mathcal{X}_{\rm sf} \cup \mathcal{X}_{\rm bd}}^\top \boldsymbol{h}_{\mathcal{M}_{\rm bd}} + \sum_{\ell \in \mathcal{L}_{\rm sf} \cup \mathcal{L}_{\rm bd}} \Big(\nu_\ell \boldsymbol{c}_\ell + \boldsymbol{D}_\ell^\top \boldsymbol{u}_\ell \Big) = \boldsymbol{0}, \tag{55a}$$

$$\nu_\ell \geq \|\boldsymbol{u}_\ell\|_2, \quad \forall \ell \in \mathcal{L}_{\rm sf} \cup \mathcal{L}_{\rm bd}, \tag{55b}$$

$$\nu_\ell \boldsymbol{c}_\ell^\top \hat{\boldsymbol{x}} + \boldsymbol{u}_\ell^\top \boldsymbol{D}_\ell \hat{\boldsymbol{x}} = \boldsymbol{0}, \quad \forall \ell \in \mathcal{L}_{\rm sf} \cup \mathcal{L}_{\rm bd}, \tag{55c}$$

$$\|\boldsymbol{h}_{\mathcal{M}_{\rm sf}}\|_\infty \leq 1, \|\boldsymbol{h}_{\mathcal{M}_{\rm bd}}\|_\infty \leq 1. \tag{55d}$$

Similarly, by the optimality of $\hat{\boldsymbol{x}}_{\rm at}$ for (51), we can find a dual certificate such that:

$$\boldsymbol{A}_{\mathcal{M}_{\rm at},\mathcal{X}_{\rm at}}^\top \boldsymbol{h}_{\mathcal{M}_{\rm at}} + \sum_{\ell \in \mathcal{L}_{\rm at}} \Big(\nu_\ell \boldsymbol{c}_\ell + \boldsymbol{D}_\ell^\top \boldsymbol{u}_\ell \Big) = \boldsymbol{0}, \quad \boldsymbol{h}_{\mathcal{M}_{\rm at}} \in \partial \|\hat{\boldsymbol{b}}_{\mathcal{M}_{\rm at}}\|_1.$$

Thus, by setting $\Big(\{\nu_\ell = 0, \boldsymbol{u}_\ell = \boldsymbol{0}\}_{\ell \in \mathcal{L}_{\rm at} \cap \rm bi} \Big)$, and note that $\mathcal{L} = \mathcal{L}_{\rm sf} \cup \mathcal{L}_{\rm bd} \cup \mathcal{L}_{\rm at} \cap \rm bi \cup \mathcal{L}_{\rm at}$, the construction $\big(\{\nu_\ell, \boldsymbol{u}_\ell\}_{\ell \in \mathcal{L}} \big)$ and $\boldsymbol{h}_\star = \begin{bmatrix} \boldsymbol{h}_{\mathcal{M}_{\rm sf}}^\top & \boldsymbol{h}_{\mathcal{M}_{\rm bd}}^\top & \boldsymbol{h}_{\mathcal{M}_{\rm at}}^\top \end{bmatrix}^\top$ yield a dual certificate. □

Now, we formally define the graphical mutual incoherence.

Definition 13 (Graphical mutual incoherence for SOCP)**.** *For each line $\{i,j\}_\ell \in \mathcal{L}$ and a given $\boldsymbol{x} \in \mathcal{K}$ that satisfies primal feasibility, define the graphical mutual incoherence $\alpha_{i \to j}^{\rm SOCP}$ along the direction $i \to j$ as the optimal value of the following minimax program:*

$$\alpha_{i \to j}^{\rm SOCP}(\boldsymbol{x}) = \max_{\boldsymbol{\xi} \in \{-1,+1\}^{n_\times^{i \to j}}} \min_{\alpha \in \mathbb{R}, \boldsymbol{\omega} \in \mathbb{R}^{n_{\mathcal{L}}^{i \to j}}, \boldsymbol{h} \in \mathbb{R}^{n_\checkmark^{i \to j}}} \alpha \tag{56a}$$

$$\text{subject to} \quad \boldsymbol{A}_{\mathcal{M}_{\rm bd\checkmark}^{i \to j}, \mathcal{X}_{\rm bd}^{i \to j}}^\top \boldsymbol{h} + \boldsymbol{A}_{\mathcal{M}_{\rm bd\times}^{i \to j}, \mathcal{X}_{\rm bd}^{i \to j}}^\top \boldsymbol{\xi} + \sum_{\ell \in \mathcal{L}_{\rm bd}^{i \to j}} \omega_\ell \boldsymbol{T}_\ell \boldsymbol{x} = \boldsymbol{0} \tag{56b}$$

$$\omega_\ell \geq 0, \qquad \forall \ell \in \mathcal{L}_{\rm bd}^{i \to j} \tag{56c}$$

$$\|\boldsymbol{h}\|_\infty \leq \alpha, \tag{56d}$$

where $n_\checkmark^{i \to j} = |\mathcal{M}_{\rm bd\checkmark}^{i \to j}|, n_\times^{i \to j} = |\mathcal{M}_{\rm bd\times}^{i \to j}|, n_{\mathcal{L}}^{i \to j} = |\mathcal{L}_{\rm bd}^{i \to j}|$ are the number of measurements/lines in $\mathcal{M}_{\rm bd\checkmark}^{i \to j}$, $\mathcal{M}_{\rm bd\times}^{i \to j}$ and $\mathcal{L}_{\rm bd}^{i \to j}$, respectively, and $\mathcal{X}_{\rm bd}^{i \to j}$, $\mathcal{M}_{\rm bd\checkmark}^{i \to j}$, $\mathcal{M}_{\rm bd\times}^{i \to j}$ and $\mathcal{L}_{\rm bd}^{i \to j}$ are defined in Def. 6. Also, we define $\boldsymbol{T}_\ell = \boldsymbol{c}_\ell \boldsymbol{c}_\ell^\top - \boldsymbol{D}_\ell^\top \boldsymbol{D}_\ell$, where $\boldsymbol{c}_\ell$ and $\boldsymbol{D}_\ell$ are given in (6)*. Similarly, we define the backward graphical mutual incoherence $\alpha_{j \to i}$ by replacing $i \to j$ to $j \to i$ in* (56)*. We adopt the measurement normalization convention in Def. 1.*

Lemma 14. *The graphical mutual incoherence $\alpha_{i\to j}^{\mathrm{SOCP}}(\boldsymbol{x})$ for a given $\boldsymbol{x} \in \mathcal{K}$ that satisfies the primal feasibility coincides with the optimal objective value of the following minimax program:*

$$\tilde{\alpha}_{i\to j}^{\mathrm{SOCP}}(\boldsymbol{x}) = \max_{\tilde{\boldsymbol{\xi}}\in[-1,+1]^{n_\times^{i\to j}}} \min_{\tilde{\alpha}\in\mathbb{R},\boldsymbol{\nu}\in\mathbb{R}^{n_{\mathcal{L}}^{i\to j}},\tilde{\boldsymbol{h}}\in\mathbb{R}^{n_\checkmark^{i\to j}}} \tilde{\alpha} \tag{57a}$$

$$\text{subject to} \quad \boldsymbol{A}_{\mathcal{M}_{\mathrm{bd}\checkmark}^{i\to j},\mathcal{X}_{\mathrm{bd}}^{i\to j}}^{\top}\tilde{\boldsymbol{h}} + \boldsymbol{A}_{\mathcal{M}_{\mathrm{bd}\times}^{i\to j},\mathcal{X}_{\mathrm{bd}}^{i\to j}}^{\top}\tilde{\boldsymbol{\xi}} + \sum_{\ell\in\mathcal{L}_{\mathrm{bd}}^{i\to j}} \nu_\ell \boldsymbol{c}_\ell + \boldsymbol{D}_\ell^{\top}\boldsymbol{u}_\ell = \boldsymbol{0} \tag{57b}$$

$$\nu_\ell \geq \|\boldsymbol{u}_\ell\|_2, \qquad \forall \ell \in \mathcal{L}_{\mathrm{bd}}^{i\to j} \tag{57c}$$

$$\nu_\ell \boldsymbol{c}_\ell^{\top}\boldsymbol{x} + \boldsymbol{u}_\ell^{\top}\boldsymbol{D}_\ell\boldsymbol{x} = \boldsymbol{0}, \qquad \forall \ell \in \mathcal{L}_{\mathrm{bd}}^{i\to j} \tag{57d}$$

$$\|\tilde{\boldsymbol{h}}\|_\infty \leq \tilde{\alpha}, \tag{57e}$$

with the same notations as in Def. 7, where $\boldsymbol{c}_\ell$ and $\boldsymbol{D}_\ell$ are define in (6)*.*

Proof. The equivalence between optimizing over $[-1,+1]^{n_\times^{i\to j}}$ and $\{-1,+1\}^{n_\times^{i\to j}}$ for the outer minimization can be reasoned as in (19) 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., [1, Lemma 15]) in analogy to linear programming indicates that (57d) 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 $\omega_\ell \geq 0$ and $\ell \in \mathcal{L}_{\mathrm{bd}}^{i\to j}$. It can be verified that this also satisfies the SOCP constraints (57c). By the definition of $\boldsymbol{T}_\ell = \boldsymbol{c}_\ell\boldsymbol{c}_\ell^{\top} - \boldsymbol{D}_\ell^{\top}\boldsymbol{D}_\ell$, the equivalence to (56) is established.

□

Lemma 15 (Local property implies global property for SOCP)**.** *Given $\mathcal{B}_{\mathrm{at}}, \mathcal{B}_{\mathrm{bi}}, \mathcal{B}_{\mathrm{bo}}$, and $\mathcal{B}_{\mathrm{sf}}$ and the associated set partitioning (c.f., Def. 3), let $\boldsymbol{A}^{\circ} = \begin{bmatrix} \boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}},\mathcal{X}_{\mathrm{sf}}} & \boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}},\mathcal{X}_{\mathrm{bd}}} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\mathrm{bo}},\mathcal{X}_{\mathrm{bd}}} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\mathrm{bi}},\mathcal{X}_{\mathrm{bd}}} \end{bmatrix}$, and $\boldsymbol{c}_\ell^{\circ}$ and $\boldsymbol{D}_\ell^{\circ}$ to be the subvector and submatrix of $\boldsymbol{c}_\ell$ and $\boldsymbol{D}_\ell$ indexed by $\mathcal{X}_{\mathrm{sf}} \cup \mathcal{X}_{\mathrm{bd}}$. If $\alpha_{i\to j}^{\mathrm{SOCP}} \leq 1-\gamma$ and $\gamma > 0$ for all $\{i,j\} \in \mathcal{L}_{\mathrm{at}\cap\mathrm{bi}}$ such that $i \in \mathcal{B}_{\mathrm{at}}$ and $j \in \mathcal{B}_{\mathrm{bi}}$, 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 $\|\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{sf}}\cup\mathcal{M}_{\mathrm{bo}}}\|_\infty \leq 1-\gamma$ and*

$$\boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}}\cup\mathcal{M}_{\mathrm{bo}}}^{\circ\top}\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{sf}}\cup\mathcal{M}_{\mathrm{bo}}} + \boldsymbol{A}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{bi}}} + \sum_{\ell\in\mathcal{L}_{\mathrm{at}\cap\mathrm{bi}}\cup\mathcal{L}_{\mathrm{bd}}\cup\mathcal{L}_{\mathrm{sf}}} \hat{\nu}_\ell \boldsymbol{c}_\ell^{\circ} + \boldsymbol{D}_\ell^{\circ\top}\hat{\boldsymbol{u}}_\ell = \boldsymbol{0}. \tag{58}$$

Proof. The proof is similar to the one for Lemma 9 and is omitted for brevity.

□

Proposition 16 (SOC constraint can improve graphical mutual incoherence)**.** *For any $\boldsymbol{x} \in \mathcal{K}$, it holds that*

$$\alpha_{i\to j}^{\mathrm{SOCP}}(\boldsymbol{x}) \leq \alpha_{i\to j}$$

Proof. For any given $\boldsymbol{\xi}$, let $\hat{\boldsymbol{h}}$ be the optimal solution of the inner minimizer in (33) with $\|\hat{\boldsymbol{h}}\|_\infty \leq \alpha_{i\to j}$. Then, the tuple $(\alpha_{i\to j}^{\mathrm{SOCP}} = \alpha_{i\to j}, \boldsymbol{\omega} = \boldsymbol{0}, \boldsymbol{h} = \hat{\boldsymbol{h}})$ is a feasible solution for (56), which proves that we always have $\alpha_{i\to j}^{\mathrm{SOCP}}(\boldsymbol{x}) \leq \alpha_{i\to j}$. □

The above proposition implies a key advantage of incorporating SOCP constraints—to improve robustness. This has also been empirically validated in our study as shown in the main text.

Theorem 17. *Consider the measurements $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_\natural + \boldsymbol{b}_\natural$, where $\mathrm{supp}(\boldsymbol{b}_\natural) \subseteq \mathcal{M}_{\mathrm{at}}$, and also a partitioning of the network as $\mathcal{B}_{\mathrm{at}}, \mathcal{B}_{\mathrm{bi}}, \mathcal{B}_{\mathrm{bo}}$, and $\mathcal{B}_{\mathrm{sf}}$. Let $\mathcal{K}_{\mathrm{bd}}$ and $\mathcal{K}_{\mathrm{at}}$ be the subsets of SOCP constraints $\mathcal{K}$ restricted to variables $\boldsymbol{x}_{\mathrm{bd}}$ and $\boldsymbol{x}_{\mathrm{at}}$, respectively, and let*

$$\tilde{\mathcal{K}}_{\mathrm{at}}(\hat{\boldsymbol{x}}_{\mathrm{bd}}) = \Bigg\{ \boldsymbol{x}_{\mathrm{at}} \Bigg| \begin{bmatrix} x_i^{mg} & x_\ell^{re} + jx_\ell^{im} \\ x_\ell^{re} - jx_\ell^{im} & x_j^{mg} \end{bmatrix} \succeq 0, \\ \forall \ell := (i,j) \in \mathcal{L}_{\mathrm{at}} \cup \mathcal{L}_{\mathrm{at}\cap\mathrm{bi}}, \text{ where } x_i^{mg} = \hat{x}_i^{mg} \quad \forall i \in \mathcal{B}_{\mathrm{bi}} \Bigg\},$$

be the confined feasible set for $\boldsymbol{x}_{\mathrm{at}}$, which fixes the boundary variables $\hat{\boldsymbol{x}}_{\mathrm{bd}}$ in the SOCP constraints. Suppose that the following conditions hold:

- *(Full column rank for the safe and boundary region) $\boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}}\cup\mathcal{M}_{\mathrm{bd}},\mathcal{X}_{\mathrm{sf}}\cup\mathcal{X}_{\mathrm{bd}}}$ and*

$$\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bd}},\mathcal{X}_{\mathrm{bd}}} = \begin{bmatrix} \boldsymbol{A}_{\mathcal{M}_{\mathrm{bd}},\mathcal{X}_{\mathrm{bd}}} & \boldsymbol{I}_{\mathcal{M}_{\mathrm{bi}}}^{(|\mathcal{M}_{\mathrm{bd}}|)\top} \end{bmatrix}$$

  have full column rank.

- *(Localized mutual incoherence) for all lines $\{i,j\} \in \mathcal{L}_{\mathrm{at}\cap\mathrm{bi}}$ that bridge the attacked region and the inner boundary, where $i \in \mathcal{B}_{\mathrm{at}}$, $j \in \mathcal{B}_{\mathrm{bi}}$, we have $\alpha_{i\to j}^{\mathrm{SOCP}} \leq 1-\gamma$ for some $\gamma > 0$.*

- *(Nonbinding SOCP constraints in the boundary) the solution for the attacked states satisfies $\hat{\boldsymbol{x}}_{\mathrm{at}} \in \tilde{\mathcal{K}}_{\mathrm{at}}(\boldsymbol{x}_{\natural\mathrm{bd}})$.*

Then, the solution to ($\mathrm{S}^{(1)}$: ℓ_1-$\mathcal{K}$)*, denoted as $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{b}})$, uniquely recovers the true state outside the attacked region (i.e., $\hat{\boldsymbol{x}}_{\mathrm{sf}} = \boldsymbol{x}_{\natural\mathrm{sf}}$ and $\hat{\boldsymbol{x}}_{\mathrm{bd}} = \boldsymbol{x}_{\natural\mathrm{bd}}$). Furthermore, the state estimation by* ($\mathrm{S}^{(2)}$: ℓ_2) *recovers the true state for the unaffected region (i.e., $\hat{v}_k = v_k$ for $k \in \mathcal{B}_{\mathrm{sf}} \cup \mathcal{B}_{\mathrm{bd}}$).*

Proof. To prove the claim, we simply need to show that for an arbitrary $\boldsymbol{b}_\star$ with its support limited to the inner boundary $\mathrm{supp}(\boldsymbol{b}_\star) \subseteq \mathcal{M}_{\mathrm{bi}}$, the solution $\hat{\boldsymbol{x}}_{\mathrm{bd}} \in \mathcal{X}_{\mathrm{bd}}$ to the program

$$\min_{\boldsymbol{x}_{\mathrm{bd}} \in \mathcal{K}_{\mathrm{bd}}, \boldsymbol{b}} \|\boldsymbol{b}\|_1, \quad \text{subject to} \quad \boldsymbol{A}_{\mathcal{M}_{\mathrm{bd}},\mathcal{X}_{\mathrm{bd}}} \boldsymbol{x}_{\mathrm{bd}} + \boldsymbol{b} = \boldsymbol{z}_{\mathcal{M}_{\mathrm{bd}}} \tag{59}$$

is unique and satisfies $\hat{\boldsymbol{x}}_{\mathrm{bd}} = \boldsymbol{x}_{\natural\mathrm{bd}}$, where $\boldsymbol{z}_{\mathcal{M}_{\mathrm{bd}}} = \boldsymbol{A}_{\mathcal{M}_{\mathrm{bd}},\mathcal{X}_{\mathrm{bd}}} \boldsymbol{x}_{\natural\mathrm{bd}} + \boldsymbol{b}_\star$. To show this, we obtain the dual program:

$$\begin{aligned} \min_{\boldsymbol{h}_{\mathcal{M}_{\mathrm{bd}}}, \{\nu_\ell, \boldsymbol{\mu}_\ell\}_{\ell \in \mathcal{L}_{\mathrm{bd}}}} \quad & \boldsymbol{h}_{\mathcal{M}_{\mathrm{bd}}}^\top \boldsymbol{z}_{\mathcal{M}_{\mathrm{bd}}} && \text{(60a)} \\ \text{subject to} \quad & \boldsymbol{A}_{\mathcal{M}_{\mathrm{bd}},\mathcal{X}_{\mathrm{bd}}}^\top \boldsymbol{h}_{\mathcal{M}_{\mathrm{bd}}} + \sum_{\ell \in \mathcal{L}_{\mathrm{at}\cap\mathrm{bi}} \cup \mathcal{L}_{\mathrm{bd}}} (\nu_\ell \boldsymbol{c}_\ell + \boldsymbol{D}_\ell^\top \boldsymbol{\mu}_\ell) = \boldsymbol{0} && \text{(60b)} \\ & \|\boldsymbol{h}_{\mathcal{M}_{\mathrm{bd}}}\|_\infty \leq 1 && \text{(60c)} \\ & \nu_\ell \geq \|\boldsymbol{\mu}_\ell\|_2, \forall \ell \in \mathcal{L}_{\mathrm{bd}}, && \text{(60d)} \end{aligned}$$

Our goal is to find a dual certificate $\boldsymbol{h}_{\star\mathcal{M}_{\rm bd}}$ and $\{\lambda_{\star\ell}, \boldsymbol{\mu}_{\star\ell}\}_{\mathcal{L}_{\rm bd}}$ that satisfies the KKT conditions:

$$\text{(dual feasibility)} \qquad \lambda_{\star\ell} \geq \|\boldsymbol{\mu}_{\star\ell}\|_2, \quad \forall \ell \in \mathcal{L}_{\rm bd} \tag{61}$$

$$\text{(stationarity)} \qquad \boldsymbol{A}^\top_{\mathcal{M}_{\rm bd},\mathcal{X}_{\rm bd}} \boldsymbol{h}_{\star\mathcal{M}_{\rm bd}} + \sum_{\ell \in \mathcal{L}_{\rm at\cap bi} \cup \mathcal{L}_{\rm bd}} (\lambda_{\star\ell} \boldsymbol{c}_\ell + \boldsymbol{D}_\ell^\top \boldsymbol{\mu}_{\star\ell}) = \boldsymbol{0}, \tag{62}$$

$$\boldsymbol{h}_{\star\mathcal{M}_{\rm bd}} \in \partial \|\boldsymbol{b}_\star\|_1 \tag{63}$$

$$\text{(complementary slackness)} \qquad \lambda_{\star\ell} \boldsymbol{c}_\ell^\top \boldsymbol{x}_\star + \boldsymbol{\mu}_{\star\ell}^\top \boldsymbol{D}_\ell \boldsymbol{x}_\star = \boldsymbol{0}, \quad \forall \ell \in \mathcal{L}_{\rm bd}. \tag{64}$$

where $\boldsymbol{x}_\star = \begin{bmatrix} \boldsymbol{x}_{\natural \rm sf}^\top & \boldsymbol{x}_{\natural \rm bd}^\top & \hat{\boldsymbol{x}}_{\rm at}^\top \end{bmatrix}^\top$. By the limited support assumption, we need to find a vector $\boldsymbol{h}_\star$ such that $\boldsymbol{h}_{\star\mathcal{M}_{\rm bi}} = \operatorname{sign}(\boldsymbol{b}_{\star\mathcal{M}_{\rm bi}})$ and $\|\boldsymbol{h}_{\star\mathcal{M}_{\rm bo}}\|_\infty \leq 1$. By the graphical mutual incoherence condition and Lemma 15, we can always find $\boldsymbol{h}_{\star\mathcal{M}_{\rm bo}}$ and $\{\lambda_{\star\ell}, \boldsymbol{\mu}_{\star\ell}\}_{\mathcal{L}_{\rm bd}}$ that satisfy the KKT conditions for a given $\boldsymbol{h}_{\star\mathcal{M}_{\rm bi}}$, such that $\|\boldsymbol{h}_{\star\mathcal{M}_{\rm bo}}\| \leq 1 - \gamma < 1$. Thus, this certifies the optimality of $(\boldsymbol{x}_{\natural \rm bd}, \boldsymbol{b}_\star)$ for (40). Clearly, under the nonbinding SOCP constraints assumption, $(\boldsymbol{x}_{\natural \rm bd}, \boldsymbol{b}_\star)$ is feasible. Following the uniqueness argument of Theorem 10, we conclude the proof. □

We can extend the analysis to the case where the measurements have both sparse bad data and dense noise. In this case, we need to solve a second-order cone program that combines quadratic loss with absolute value loss, in addition to the SOCP constraints.

Theorem 18 (Robust SE with ($\mathrm{S}^{(1)}$: $\ell_2\ell_1$-$\mathcal{K}$))**.** *Given the measurements $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_\natural + \boldsymbol{w}_\natural + \boldsymbol{b}_\natural$, where $\boldsymbol{w}_\natural$ has independent entries with zero mean and subgaussian parameter σ and $\operatorname{supp}(\boldsymbol{b}_\natural) \subseteq \mathcal{M}_{\rm at}$, consider a partitioning of the network as $\mathcal{B}_{\rm at}, \mathcal{B}_{\rm bi}, \mathcal{B}_{\rm bo}$, and $\mathcal{B}_{\rm sf}$. Let $\mathcal{K}_{\rm bd}$ and $\mathcal{K}_{\rm at}$ be the subsets of SOCP constraints $\mathcal{K}$ restricted to the variables $\boldsymbol{x}_{\rm bd}$ and $\boldsymbol{x}_{\rm at}$, respectively, and let*

$$\tilde{\mathcal{K}}_{\rm at}(\hat{\boldsymbol{x}}_{\rm bd}) = \Bigg\{ \boldsymbol{x}_{\rm at} \Bigg| \begin{bmatrix} x_i^{mg} & x_\ell^{re} + j x_\ell^{im} \\ x_\ell^{re} - j x_\ell^{im} & x_j^{mg} \end{bmatrix} \succeq 0, \quad \forall \ell := (i,j) \in \mathcal{L}_{\rm at} \cup \mathcal{L}_{\rm at\cap bi},\ \text{where } x_i^{mg} = \hat{x}_i^{mg} \quad \forall i \in \mathcal{B}_{\rm bi} \Bigg\},$$

be the confined feasible set for $\boldsymbol{x}_{\rm at}$, which fixes the boundary variables $\hat{\boldsymbol{x}}_{\rm bd}$ in the SOCP constraints. Suppose that the rows of $\boldsymbol{A}$ are normalized (c.f., Def. 1), and the regularization parameter λ is chosen such that

$$\lambda > \frac{2}{n_m \gamma} \sqrt{2\sigma^2 \log n_m}. \tag{65}$$

In addition, suppose the following conditions hold:

- *(Full column rank for the safe and boundary region) both $\boldsymbol{A}_{\mathcal{M}_{\rm sf} \cup \mathcal{M}_{\rm bd}, \mathcal{X}_{\rm sf} \cup \mathcal{X}_{\rm bd}}$ and*

  $$\boldsymbol{Q}_{\mathcal{M}_{\rm bd},\mathcal{X}_{\rm bd}} = \begin{bmatrix} \boldsymbol{A}_{\mathcal{M}_{\rm bd},\mathcal{X}_{\rm bd}} & \boldsymbol{I}_{\mathcal{M}_{\rm bi}}^{(|\mathcal{M}_{\rm bd}|)\top} \end{bmatrix}$$

  have full column rank.

- *(Localized mutual incoherence) for all lines $\{i,j\} \in \mathcal{L}_{\rm at\cap bi}$ that bridge the attacked region and the inner boundary, where $i \in \mathcal{B}_{\rm at}$, $j \in \mathcal{B}_{\rm bi}$, we have $\alpha_{i\to j}^{\rm SOCP} \leq 1 - \gamma$ for some $\gamma > 0$.*

- *(Nonbinding SOCP constraints in the boundary) the solution for the attacked states satisfies $\hat{\boldsymbol{x}}_{\rm at} \in \tilde{\mathcal{K}}_{\rm at}(\boldsymbol{x}_{\natural \rm bd})$.*

Then, the following properties hold for the solution to ($\mathrm{S}^{(1)}$: $\ell_2\ell_1$)*, denoted as* $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{b}})$*:*

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

2. *(Large bad data detection) Let* $\boldsymbol{A}^\circ := \begin{bmatrix} \boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}},\mathcal{X}_{\mathrm{sf}}} & \boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}},\mathcal{X}_{\mathrm{bd}}} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\mathrm{bo}},\mathcal{X}_{\mathrm{bd}}} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\mathrm{bi}},\mathcal{X}_{\mathrm{bd}}} \end{bmatrix}$ *and* $\boldsymbol{Q}^\circ_{\mathcal{M}_{\mathrm{bi}}} = \begin{bmatrix} \boldsymbol{A}^\circ & \boldsymbol{I}^\top_{\mathcal{M}_{\mathrm{bi}}} \end{bmatrix}$*, and*

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

be a threshold value, 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 error at the boundary. 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*

$$\|\boldsymbol{x}_{\natural \mathcal{X}_{\mathrm{sf}} \cup \mathcal{X}_{\mathrm{bd}}} - \hat{\boldsymbol{x}}_{\mathcal{X}_{\mathrm{sf}} \cup \mathcal{X}_{\mathrm{bd}}}\|_2 \leq t \frac{\sqrt{|\mathcal{X}_{\mathrm{sf}}| + |\mathcal{X}_{\mathrm{bd}}| + |\mathcal{M}_{\mathrm{bi}}|}}{C_{min}} + n_m \lambda \|\boldsymbol{I}_x (\boldsymbol{Q}^{\circ\top}_{\mathcal{M}_{\mathrm{bi}}} \boldsymbol{Q}^\circ_{\mathcal{M}_{\mathrm{bi}}})^{-1} \boldsymbol{Q}^{\circ\top}_{\mathcal{M}_{\mathrm{bi}}}\|_{\infty,2}$$

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

3.3 Scalable methods to calculate the graphical mutual incoherence

The minimax program (33) consists of a linear programming in the inner minimization and a discrete optimization in the outer maximization. For small-scale systems, the number of feasible points in the outer maximization is not too large. This is the case when we consider the graphical mutual incoherence on a line-by-line basis. But for large-scale problems when we consider a group of attacked lines, it is essential to develop more scalable numerical algorithms. We first show the following result.

Lemma 19. *The graphical mutual incoherence* $\alpha_{i\to j}$ *coincides with the optimal value of the following minimax program:*

$$\tilde{\alpha}_{i\to j} = \max_{\tilde{\boldsymbol{\xi}} \in [-1,+1]^{n_\times^{i\to j}}} \min_{\tilde{\alpha} \in \mathbb{R}, \tilde{\boldsymbol{h}} \in \mathbb{R}^{n_\checkmark^{i\to j}}} \tilde{\alpha} \tag{66a}$$

$$\text{subject to} \quad \boldsymbol{A}^\top_{\mathcal{M}^{i\to j}_{\mathrm{bd}\checkmark}, \mathcal{X}^{i\to j}_{\mathrm{bd}}} \tilde{\boldsymbol{h}} + \boldsymbol{A}^\top_{\mathcal{M}^{i\to j}_{\mathrm{bd}\times}, \mathcal{X}^{i\to j}_{\mathrm{bd}}} \tilde{\boldsymbol{\xi}} = \boldsymbol{0} \tag{66b}$$

$$\|\tilde{\boldsymbol{h}}\|_\infty \leq \tilde{\alpha}, \tag{66c}$$

with the same notations as in Def. 7. Note that the difference in (66) *is that the minimizer is over the hypercube* $[-1,+1]^{n_\times^{i\to j}}$ *rather than the simplex* $\{-1,+1\}^{n_\times^{i\to j}}$*.*

Proof. Since the feasible region of the outside maximizer in (66) is a superset of that in (33), we always have $\tilde{\alpha}_{i\to j} \geq \alpha_{i\to j}$. To show the other direction, we simply need to show that for any $\tilde{\boldsymbol{\xi}} \in [-1,+1]^{n_\times^{i\to j}}$, we can always find a feasible solution for the minimizer $\tilde{\boldsymbol{h}}$ such that $\|\tilde{\boldsymbol{h}}\|_\infty \leq \alpha_{i\to j}$. Since $\tilde{\boldsymbol{\xi}}$ belongs to a hypercube, which is convex, there always exists a set of non-negative coefficients β_k such that $\beta_k \geq 0$,

$\sum_k \beta_k = 1$ and $\tilde{\boldsymbol{\xi}} = \sum_k \beta_k \boldsymbol{\xi}_k$, where $\boldsymbol{\xi}_k \in \{-1, +1\}^{n_\times^{i\to j}}$. Since for each $\boldsymbol{\xi}_k$, there exists $\boldsymbol{h}_k$ such that it is feasible in (33) and $\|\boldsymbol{h}_k\|_\infty \le \alpha_{i\to j}$, by choosing $\tilde{\boldsymbol{h}} = \sum_k \beta_k \boldsymbol{h}_k$, we have:

$$\|\tilde{\boldsymbol{h}}\|_\infty \le \sum_k \beta_k \|\boldsymbol{h}_k\|_\infty \le \sum_k \beta_k \alpha_{i\to j} = \alpha_{i\to j},$$

which completes the proof. □

We can thereby reformulate the problem as a linear complimentarity problem as follows. The KKT conditions for the inner minimization of (66) are:

- (Primal feasibility) $\boldsymbol{A}^\top_{\mathcal{M}^{i\to j}_{\mathrm{bd}\checkmark}, \mathcal{X}^{i\to j}_{\mathrm{bd}}} \boldsymbol{h} + \boldsymbol{A}^\top_{\mathcal{M}^{i\to j}_{\mathrm{bd}\times}, \mathcal{X}^{i\to j}_{\mathrm{bd}}} \boldsymbol{\xi} = \boldsymbol{0}, \boldsymbol{q}_+ = \alpha\boldsymbol{1} - \boldsymbol{h}, \boldsymbol{q}_- = \alpha\boldsymbol{1} + \boldsymbol{h}, \boldsymbol{q}_+ \ge 0, \boldsymbol{q}_- \ge 0$;
- (Dual feasibility) $\boldsymbol{\mu}_+ \ge 0, \boldsymbol{\mu}_- \ge 0$;
- (Stationarity) The Lagrangian function

  $$\mathcal{L}(\alpha, \boldsymbol{h}, \boldsymbol{\mu}_+, \boldsymbol{\mu}_-, \boldsymbol{\lambda}) = \alpha + \boldsymbol{\lambda}^\top (\boldsymbol{A}^\top_{\mathcal{M}^{i\to j}_{\mathrm{bd}\checkmark}, \mathcal{X}^{i\to j}_{\mathrm{bd}}} \boldsymbol{h} + \boldsymbol{A}^\top_{\mathcal{M}^{i\to j}_{\mathrm{bd}\times}, \mathcal{X}^{i\to j}_{\mathrm{bd}}} \boldsymbol{\xi}) + \boldsymbol{\mu}_+^\top (\boldsymbol{h} - \alpha\boldsymbol{1}) + \boldsymbol{\mu}_-^\top (-\boldsymbol{h} - \alpha\boldsymbol{1})$$

  and stationarity conditions:

  $$\frac{\partial \mathcal{L}}{\partial \alpha} = 1 - \boldsymbol{\mu}_+^\top \boldsymbol{1} - \boldsymbol{\mu}_-^\top \boldsymbol{1} = 0$$
  $$\frac{\partial \mathcal{L}}{\partial \boldsymbol{h}} = \boldsymbol{A}_{\mathcal{M}^{i\to j}_{\mathrm{bd}\checkmark}, \mathcal{X}^{i\to j}_{\mathrm{bd}}} \boldsymbol{\lambda} + \boldsymbol{\mu}_+ - \boldsymbol{\mu}_- = \boldsymbol{0}$$

- (complementary slackness) $\boldsymbol{\mu}_+ \circ \boldsymbol{q}_+ = \boldsymbol{0}, \boldsymbol{\mu}_- \circ \boldsymbol{q}_- = \boldsymbol{0}$

Thus, we can write (66) as a linear complementarity problem:

$$\begin{aligned}
\tilde{\alpha}_{i\to j} = \max_{\boldsymbol{\xi} \in \mathbb{R}^{n_\times^{i\to j}}, \alpha \in \mathbb{R}, \boldsymbol{h} \in \mathbb{R}^{n_\checkmark^{i\to j}}} \quad & \alpha && \text{(67a)}\\
\text{subject to} \quad & -\boldsymbol{1} \le \boldsymbol{\xi} \le \boldsymbol{1} && \text{(67b)}\\
& \boldsymbol{A}^\top_{\mathcal{M}^{i\to j}_{\mathrm{bd}\checkmark}, \mathcal{X}^{i\to j}_{\mathrm{bd}}} \boldsymbol{h} + \boldsymbol{A}^\top_{\mathcal{M}^{i\to j}_{\mathrm{bd}\times}, \mathcal{X}^{i\to j}_{\mathrm{bd}}} \boldsymbol{\xi} = \boldsymbol{0} && \text{(67c)}\\
& \boldsymbol{q}_+ = \alpha\boldsymbol{1} - \boldsymbol{h} && \text{(67d)}\\
& \boldsymbol{q}_- = \alpha\boldsymbol{1} + \boldsymbol{h} && \text{(67e)}\\
& 1 - \boldsymbol{\mu}_+^\top \boldsymbol{1} - \boldsymbol{\mu}_-^\top \boldsymbol{1} = 0 && \text{(67f)}\\
& \boldsymbol{A}_{\mathcal{M}^{i\to j}_{\mathrm{bd}\checkmark}, \mathcal{X}^{i\to j}_{\mathrm{bd}}} \boldsymbol{\lambda} + \boldsymbol{\mu}_+ - \boldsymbol{\mu}_- = \boldsymbol{0} && \text{(67g)}\\
& \boldsymbol{\mu}_+ \circ \boldsymbol{q}_+ = \boldsymbol{\mu}_- \circ \boldsymbol{q}_- = \boldsymbol{0} && \text{(67h)}\\
, \quad & \boldsymbol{q}_+, \boldsymbol{q}_-, \boldsymbol{\mu}_+, \boldsymbol{\mu}_- \ge 0 && \text{(67i)}
\end{aligned}$$

This problem can be solved readily using off-the-shelf solvers such as PATH Solver [5] or YALMIP [11]. We can also use the big-M method to replace the complimentarity condition using a mixed-integer formulation, and solve the problem using standard packages such as Gurobi. In our experiments, we only focus on each line, so the improvement of computation is not significant. The advantage becomes more obvious when we scale the computation to multiple lines, such as the case for the tree decomposition.

3.4 Extension to tree decomposition

So far, we have been focusing on evaluating line vulnerabilities. In this section, we introduce a powerful extension of the vulnerability index to tree decomposition of a graph. This allows us to study the effect of sparsity on network robustness. We use $\mathcal{N}(\mathcal{G})$ to represent the vertices of graph $\mathcal{G}$, and $\mathcal{L}(i;\mathcal{G}) = \{j \in \mathcal{N} \mid \{i,j\} \in \mathcal{L}\}$ to represent the set of nodes in $\mathcal{G}$ that are connected to node i. First, we introduce the standard definition of tree decomposition and treewidth.

Definition 20 (Tree decomposition and treewidth)**.** *A tree decomposition of a graph $\mathcal{G} \coloneqq \{\mathcal{N}, \mathcal{L}\}$ is $(\mathcal{T}, \mathcal{W})$, where $\mathcal{T}$ is a tree and $\mathcal{W} \coloneqq \{W_t \mid t \in \mathcal{N}(\mathcal{T})\}$ is the set of "bags" W_t which satisfies the following properties*

1. *(Node coverage) $\cup_{t\in\mathcal{N}(\mathcal{T})}\mathcal{W}_t = \mathcal{N}(\mathcal{G})$, i.e., the union of the vertices of $\mathcal{T}$, referred to as "bags," is the set of nodes of $\mathcal{G}$;*
2. *(Edge coverage) For any $(i,j) \in \mathcal{L}$, there exists $t \in \mathcal{N}(\mathcal{T})$ such that $i, j \in \mathcal{W}_t$, i.e., each edge of $\mathcal{G}$ is in at least one of the "bags" of $\mathcal{T}$;*
3. *(Running intersection property) The subtree of $\mathcal{T}$ consisting of all "bags" containing $u \in \mathcal{N}$ is connected.*

Furthermore, the width of a tree decomposition is $\max(|\mathcal{W}_t| - 1 : t \in \mathcal{N}(\mathcal{T}))$. The treewidth of $\mathcal{G}$ is the minimum width of a tree decomposition of $\mathcal{G}$.

Clearly, a graph may have several different tree decompositions. The analysis below does not require any particular tree decompositions. However, the easiest tree decomposition is to lump all vertices into one bag, which does not reveal any robustness properties of the graph. In general, the smaller the width of the decomposition, the easier it is to certify robustness.

Definition 21 (Infected bags, link bags, safe bags)**.** *For a given set of attacked nodes $\mathcal{N}_{\rm at}$ and a tree decomposition $(\mathcal{T}, \mathcal{W})$, any bag that contains attacked nodes is referred to as an infected bag $\mathcal{W}_t^{\rm if} \in \mathcal{W}^{\rm if} = \{\mathcal{W}_t \mid \mathcal{W}_t \cap \mathcal{N}_{\rm at} \neq \emptyset\}$. Furthermore, the set of lines induced by the union of infected bags is denoted as $\mathcal{L}_{\rm if}$. The bags that are immediately connected to an infected bag are called link bags $\mathcal{W}_t^{\rm lk} \in \mathcal{W}^{\rm lk} = \{\mathcal{W}_t \mid \mathcal{W}_t \cap \mathcal{N}_{\rm at} = \emptyset, \exists\ \mathcal{W}_{t'}^{\rm if}, \mathcal{W}_t \in \mathcal{L}(\mathcal{W}_{t'}^{\rm if}; \mathcal{T})\}$, and the set of lines induced by the union of link bags is shown as $\mathcal{L}_{\rm lk}$. The rest of the bags are safe bags $\mathcal{W}_t^{\rm sf}$, and the set of lines induced by the union of safe bags is represented by $\mathcal{L}_{\rm sf}$. Nodes shared between a link bag $\mathcal{W}_t^{\rm lk}$ and an infected bag $\mathcal{W}_t^{\rm if}$ are called adhesion nodes $\mathcal{N}_{\rm ad}(\mathcal{W}_t^{\rm lk}, \mathcal{W}_t^{\rm if}) = \mathcal{W}_t^{\rm lk} \cap \mathcal{W}_t^{\rm if} \subseteq \mathcal{N}_{\rm ad}$, and the rest of the nodes in $\mathcal{W}_t^{\rm lk}$ are outer link nodes $\mathcal{N}_{\rm ol}(\mathcal{W}_t^{\rm lk}, \mathcal{W}_t^{\rm if}) = \mathcal{W}_t^{\rm lk} \setminus \mathcal{W}_t^{\rm if} \subseteq \mathcal{N}_{\rm ol}$. We denote the edges that connect adhesion nodes in $\mathcal{W}_t^{\rm lk}$ with infected nodes by $\mathcal{L}_{\rm ad}(\mathcal{W}_t^{\rm lk}) \subseteq \mathcal{L}_{\rm ad}$.*

Definition 22 (Attacked, boundary and safe variables and measurements for tree decomposition)**.** *The set of "infected variables" $\mathcal{X}_{\rm if}$ includes all variables on lines induced by $\mathcal{W}^{\rm if}$ and on nodes in $\mathcal{W}^{\rm if}$ except for adhesion nodes $\mathcal{N}_{\rm ad}$. The set of "link variables" $\mathcal{X}_{\rm lk}$ includes variables on nodes in $\mathcal{W}^{\rm lk}$ and the induced lines. The set of "safe variables" $\mathcal{X}_{\rm sf}$ includes all other variables. The set of "infected measurements" $\mathcal{M}_{\rm if}$ includes measurements on lines induced by nodes in $\mathcal{W}^{\rm if}$ and on nodes in $\mathcal{W}^{\rm if}$ except for voltage magnitude measurements on $\mathcal{N}_{\rm ad}$. The set of "adhesion measurements" $\mathcal{M}_{\rm ad}$ includes nodal power injections on nodes in $\mathcal{N}_{\rm ad}$ and line measurements on $\mathcal{L}_{\rm ad}$, and the set of "outer link measurements" $\mathcal{M}_{\rm ol}$ includes voltage magnitude on nodes in $\mathcal{W}^{\rm lk}$ and line measurements induced by nodes in $\mathcal{W}^{\rm lk}$. Together, they form the "boundary measurements" $\mathcal{M}_{\rm bd} \coloneqq \mathcal{M}_{\rm ad} \cup \mathcal{M}_{\rm ol}$. The rest of the measurements $\mathcal{M}_{\rm sf}$ are "safe measurements."*

Next, we introduce some useful properties associated with the above definitions. If $\mathcal{T}'$ is a subtree of $\mathcal{T}$, we use $\mathcal{G}_{\mathcal{T}'}$ to denote the subgraph of $\mathcal{G}$ induced by the nodes in all the bags associated with $\mathcal{T}'$, namely $\cup_{t\in\mathcal{T}'}\mathcal{W}_t$.

Lemma 23. *The following properties are satisfied:*

(i) There are no shared nodes between the safe bags and the infected bags.

(ii) There are no shared nodes between the set of outer link nodes and the infected bags.

(iii) Suppose that the infected bags form a subtree of $\mathcal{T}$. Then, there are no shared outer link nodes between any link bags.

(iv) Suppose that the infected bags form a subtree of $\mathcal{T}$. Consider any link bag $\mathcal{W}_t^{\text{lk}}$ that is adjacent to only one infected bag $\mathcal{W}_{t'}^{\text{if}}$ connected by an edge $\mathcal{L}(\mathcal{W}_t^{\text{lk}}, \mathcal{W}_{t'}^{\text{if}})$. If we delete the edge, the tree falls apart into two connected components, $\mathcal{T}_1$ and $\mathcal{T}_2$. Deleting the adhesion nodes $\mathcal{W}_t^{\text{lk}} \cap \mathcal{W}_{t'}^{\text{if}}$ from $\mathcal{N}$ disconnects $\mathcal{G}$ into the two subgraphs $\mathcal{G}_{\mathcal{T}_1} - (\mathcal{W}_t^{\text{lk}} \cap \mathcal{W}_{t'}^{\text{if}})$ and $\mathcal{G}_{\mathcal{T}_2} - (\mathcal{W}_t^{\text{lk}} \cap \mathcal{W}_{t'}^{\text{if}})$. Furthermore, all the infected nodes are contained in only one of the subgraph, and there is no edge across the two subgraphs.

Proof. (i): For any safe bag $\mathcal{W}_t^{\text{sf}}$ and affected bag $\mathcal{W}_t^{\text{if}}$, if there exists a node i that is shared between them, it contradicts the definition of a safe bag.

(ii): For any link bag $\mathcal{W}_t^{\text{lk}}$ and affected bag $\mathcal{W}_t^{\text{if}}$, if there exists a node i that is shared between $\mathcal{W}_t^{\text{if}}$ and the outer link nodes in $\mathcal{W}_t^{\text{lk}}$, then by the running intersection property, it must also appear in the infected bag connected to $\mathcal{W}_t^{\text{lk}}$. This is contradictory, because it makes i an adhesion node.

(iii): For any two link bags $\mathcal{W}_t^{\text{lk}}$ and $\mathcal{W}_{t'}^{\text{lk}}$, suppose that they share an outer link node i. By the running intersection property, there must exist a path of bags between $\mathcal{W}_t^{\text{lk}}$ and $\mathcal{W}_{t'}^{\text{lk}}$. Since this path cannot go through the infected bags, it must be outside the infected region. Since the infected bags form a subtree, this would create a loop within $\mathcal{T}$, which is impossible. Therefore, there cannot be any shared outer link nodes between any two link bags.

(iv) Assume that there is a node i that belongs to both $\mathcal{G}_{\mathcal{T}_1} - (\mathcal{W}_t^{\text{lk}} \cap \mathcal{W}_{t'}^{\text{if}})$ and $\mathcal{G}_{\mathcal{T}_2} - (\mathcal{W}_t^{\text{lk}} \cap \mathcal{W}_{t'}^{\text{if}})$. Therefore, by the node coverage property, there must exist $\mathcal{W}_x$ with $x \in \mathcal{T}_1$ and $\mathcal{W}_y$ with $y \in \mathcal{T}_2$ such that $i \in \mathcal{W}_x$ and $i \in \mathcal{W}_y$. Since $\mathcal{W}_t^{\text{lk}}$ and $\mathcal{W}_{t'}^{\text{if}}$ lie on a $x - y$ path in $\mathcal{T}$, by the running intersection property, $i \in \mathcal{W}_t^{\text{lk}} \cap \mathcal{W}_{t'}^{\text{if}}$. Hence, i belongs to neither $\mathcal{G}_{\mathcal{T}_1} - (\mathcal{W}_t^{\text{lk}} \cap \mathcal{W}_{t'}^{\text{if}})$ nor $\mathcal{G}_{\mathcal{T}_2} - (\mathcal{W}_t^{\text{lk}} \cap \mathcal{W}_{t'}^{\text{if}})$.

Now, assume that there is an edge (i, j) in $\mathcal{G}$ such that $i \in \mathcal{G}_{\mathcal{T}_1} - (\mathcal{W}_t^{\text{lk}} \cap \mathcal{W}_{t'}^{\text{if}})$ and $j \in \mathcal{G}_{\mathcal{T}_2} - (\mathcal{W}_t^{\text{lk}} \cap \mathcal{W}_{t'}^{\text{if}})$. Then, by the edge coverage property, there must be a bag $\mathcal{W}_x$ containing both i and j. However, x cannot be in both $\mathcal{T}_1$ and $\mathcal{T}_2$, otherwise, i and j will belong to $\mathcal{W}_t^{\text{lk}} \cap \mathcal{W}_{t'}^{\text{if}}$. Assume that $x \notin \mathcal{T}_2$. Since j is in $\mathcal{G}_{\mathcal{T}_2} - (\mathcal{W}_t^{\text{lk}} \cap \mathcal{W}_{t'}^{\text{if}})$, it must be in a bag $y \in \mathcal{T}_2$ different than x. Since j belongs to both $\mathcal{W}_x$ and $\mathcal{W}_y$, it lies on a $x - y$ path in $\mathcal{T}$. By the running intersection property, we have $j \in \mathcal{W}_t^{\text{lk}} \cap \mathcal{W}_{t'}^{\text{if}}$, which is a contradiction. □

If the infected bags form a subtree and we can find a link bag that is adjacent to only one infected bag, then by property (iv) in Lemma 23, if we remove the adhesion nodes, we can separate the infected region with the rest of the safe region.

Now, we can define a generalized version of vulnerability index using tree decomposition.

Definition 24 (Bag vulnerability index)**.** *For each adhesion link $\mathcal{L}(\mathcal{W}_t^{\mathrm{lk}}, \mathcal{W}_t^{\mathrm{if}})$, define the measurement and variable partitions according to Def. 22. The bag vulnerability index $\alpha_{\mathcal{W}_t^{\mathrm{if}} \to \mathcal{W}_t^{\mathrm{lk}}}$ is given by the optimal value of the following minimax program:*

$$\alpha_{\mathcal{W}_t^{\mathrm{if}} \to \mathcal{W}_t^{\mathrm{lk}}} = \max_{\boldsymbol{\xi} \in \{-1,+1\}^{|\mathcal{M}_{\mathrm{ad}}|}} \quad \min_{\alpha \in \mathbb{R}, \boldsymbol{h} \in \mathbb{R}^{|\mathcal{M}_{\mathrm{ol}}|}} \alpha \tag{68a}$$

$$\text{subject to} \qquad \boldsymbol{A}_{\mathcal{M}_{\mathrm{ol}}, \mathcal{X}_{\mathrm{lk}}}^{\top} \boldsymbol{h} + \boldsymbol{A}_{\mathcal{M}_{\mathrm{ad}}, \mathcal{X}_{\mathrm{lk}}}^{\top} \boldsymbol{\xi} = \boldsymbol{0} \tag{68b}$$

$$\|\boldsymbol{h}\|_\infty \le \alpha, \tag{68c}$$

where $\mathcal{M}_{\mathrm{ol}}, \mathcal{M}_{\mathrm{ad}}$ and $\mathcal{X}_{\mathrm{lk}}$ are the boundary measurement and variable indices introduced in Def. 22.

Similarly, we can extend the definition to incorporate SOCs.

Definition 25 (Bag vulnerability index for SOCP)**.** *For each adhesion link $\mathcal{L}(\mathcal{W}_t^{\mathrm{lk}}, \mathcal{W}_t^{\mathrm{if}})$, define the measurement and variable partitions according to Def. 22. The bag vulnerability index $\alpha^{\mathrm{SOCP}}_{\mathcal{W}_t^{\mathrm{if}} \to \mathcal{W}_t^{\mathrm{lk}}}$ for a given $\boldsymbol{x} \in \mathcal{K}$ that satisfies primal feasibility is given by the optimal value of the following minimax program:*

$$\alpha^{\mathrm{SOCP}}_{\mathcal{W}_t^{\mathrm{if}} \to \mathcal{W}_t^{\mathrm{lk}}} = \max_{\boldsymbol{\xi} \in \{-1,+1\}^{|\mathcal{M}_{\mathrm{ad}}|}} \quad \min_{\alpha \in \mathbb{R}, \boldsymbol{h} \in \mathbb{R}^{|\mathcal{M}_{\mathrm{ol}}|}} \tag{69a}$$

$$\text{subject to } \boldsymbol{A}_{\mathcal{M}_{\mathrm{ol}}, \mathcal{X}_{\mathrm{lk}}}^{\top} \boldsymbol{h} + \boldsymbol{A}_{\mathcal{M}_{\mathrm{ad}}, \mathcal{X}_{\mathrm{lk}}}^{\top} \boldsymbol{\xi} + \sum_{\ell \in \mathcal{L}(\mathcal{W}_t^{\mathrm{lk}})} \omega_\ell \boldsymbol{T}_\ell \boldsymbol{x} = \boldsymbol{0} \tag{69b}$$

$$\omega_\ell \ge 0, \qquad \forall \ell \in \mathcal{L}(\mathcal{W}_t^{\mathrm{lk}}) \tag{69c}$$

$$\|\boldsymbol{h}\|_\infty \le \alpha, \tag{69d}$$

where $\mathcal{M}_{\mathrm{ol}}, \mathcal{M}_{\mathrm{ad}}$ and $\mathcal{X}_{\mathrm{lk}}$ are the boundary measurement and variable indices introduced in Def. 22, $\mathcal{L}(\mathcal{W}_t^{\mathrm{lk}})$ is the set of lines induced by nodes in $\mathcal{W}_t^{\mathrm{lk}}$. Also, we define $\boldsymbol{T}_\ell = \boldsymbol{c}_\ell \boldsymbol{c}_\ell^{\top} - \boldsymbol{D}_\ell^{\top} \boldsymbol{D}_\ell$, where $\boldsymbol{c}_\ell$ and $\boldsymbol{D}_\ell$ are defined in (6)*.*

With the above definition of bag vulnerability index, we can show the following key results for SE robustness.

Lemma 26 (Local property implies global property in tree decomposition)**.** *Consider a tree decomposition $\mathcal{T}$ and the associated set partitioning (c.f., Def. 22). Suppose that the infected bags form a subtree of $\mathcal{T}$, and that there exists a link bag $\mathcal{W}_t^{\mathrm{lk}}$ that is adjacent to only one infected bag $\mathcal{W}_t^{\mathrm{if}}$. For simplicity of presentation, we also treat the rest of the bags in the subtree as infected. Let $\boldsymbol{A}^{\circ} = \begin{bmatrix} \boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}}, \mathcal{X}_{\mathrm{sf}}} & \boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}}, \mathcal{X}_{\mathrm{lk}}} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\mathrm{ol}}, \mathcal{X}_{\mathrm{lk}}} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\mathrm{ad}}, \mathcal{X}_{\mathrm{lk}}} \end{bmatrix}$ be a submatrix of the sensing matrix. If $\alpha_{\mathcal{W}_t^{\mathrm{if}} \to \mathcal{W}_t^{\mathrm{lk}}} \le 1 - \gamma$ for some $\gamma > 0$, then for any $\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{ad}}} \in [-1, 1]^{|\mathcal{M}_{\mathrm{ad}}|}$, there exists an $\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{sf}} \cup \mathcal{M}_{\mathrm{ol}}}$ such that $\|\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{sf}} \cup \mathcal{M}_{\mathrm{ol}}}\|_\infty \le 1 - \gamma$ and*

$$\boldsymbol{A}^{\circ\top}_{\mathcal{M}_{\mathrm{sf}} \cup \mathcal{M}_{\mathrm{ol}}} \hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{sf}} \cup \mathcal{M}_{\mathrm{ol}}} + \boldsymbol{A}^{\circ\top}_{\mathcal{M}_{\mathrm{ad}}} \hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{ad}}} = \boldsymbol{0}. \tag{70}$$

Proof. The proof is similar to Lemma 9. First, we show that a sufficient condition for the existence of $\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{sf}} \cup \mathcal{M}_{\mathrm{ol}}} = \begin{bmatrix} \hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{sf}}}^{\top} & \hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{ol}}}^{\top} \end{bmatrix}^{\top}$ such that $\|\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{sf}} \cup \mathcal{M}_{\mathrm{ol}}}\|_\infty \le 1 - \gamma$ and (70) is satisfied is that for any $\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{ad}}}$, there exists a vector $\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{ol}}}$ such that $\|\hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{ol}}}\|_\infty \le 1 - \gamma$ and

$$\boldsymbol{A}^{\top}_{\mathcal{M}_{\mathrm{ol}}, \mathcal{X}_{\mathrm{lk}}} \hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{ol}}} + \boldsymbol{A}^{\top}_{\mathcal{M}_{\mathrm{ad}}, \mathcal{X}_{\mathrm{lk}}} \hat{\boldsymbol{h}}_{\mathcal{M}_{\mathrm{ad}}} = \boldsymbol{0}. \tag{71}$$

This is immediate by simply choosing $\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm ol}} = \begin{bmatrix}\mathbf{0}^\top & \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm ol}}^\top\end{bmatrix}^\top$. Since it is guaranteed that there exists a vector $\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm ol}}$ to satisfy (71) under the condition that $\alpha_{\mathcal{W}_t^{\rm if}\to\mathcal{W}_t^{\rm lk}} \le 1-\gamma$, the claim is proved. □

Lemma 27 (Local property implies global property for SOCP with tree decomposition)**.** *Consider a tree decomposition $\mathcal{T}$ and the associated set partitioning (c.f., Def. 22). Suppose that the infected bags form a subtree of $\mathcal{T}$, and that there exists a link bag $\mathcal{W}_t^{\rm lk}$ that is adjacent to only one infected bag $\mathcal{W}_t^{\rm if}$. For simplicity of presentation, we also treat the rest of the bags in the subtree as infected. Let $\boldsymbol{A}^\circ = \begin{bmatrix}\boldsymbol{A}_{\mathcal{M}_{\rm sf},\mathcal{X}_{\rm sf}} & \boldsymbol{A}_{\mathcal{M}_{\rm sf},\mathcal{X}_{\rm lk}} \\ \mathbf{0} & \boldsymbol{A}_{\mathcal{M}_{\rm ol},\mathcal{X}_{\rm lk}} \\ \mathbf{0} & \boldsymbol{A}_{\mathcal{M}_{\rm ad},\mathcal{X}_{\rm lk}}\end{bmatrix}$ be a submatrix of the sensing matrix, and $\boldsymbol{c}_\ell^\circ$ and $\boldsymbol{D}_\ell^\circ$ be the subvector and submatrix of $\boldsymbol{c}_\ell$ and $\boldsymbol{D}_\ell$ indexed by $\mathcal{X}_{\rm sf}\cup\mathcal{X}_{\rm lk}$. If $\alpha^{\rm SOCP}_{\mathcal{W}_t^{\rm if}\to\mathcal{W}_t^{\rm lk}} \le 1-\gamma$ for some $\gamma>0$, then for any $\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm ad}} \in [-1,1]^{|\mathcal{M}_{\rm ad}|}$, there exist $\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm ol}}$ and $\{\hat\nu_\ell, \hat{\boldsymbol{u}}_\ell\}_{\ell\in\mathcal{L}_{\rm ad}(\mathcal{W}_t^{\rm lk})\cup\mathcal{L}(\mathcal{W}_t^{\rm lk})\cup\mathcal{L}_{\rm sf}}$ such that $\|\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm ol}}\|_\infty \le 1-\gamma$ and*

$$\boldsymbol{A}^{\circ\top}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm ol}}\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm ol}} + \boldsymbol{A}^{\circ\top}_{\mathcal{M}_{\rm ad}}\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm ad}} + \sum_{\ell\in\mathcal{L}_{\rm ad}(\mathcal{W}_t^{\rm lk})\cup\mathcal{L}(\mathcal{W}_t^{\rm lk})\cup\mathcal{L}_{\rm sf}} \hat\nu_\ell \boldsymbol{c}_\ell^\circ + \boldsymbol{D}_\ell^{\circ\top}\hat{\boldsymbol{u}}_\ell = \mathbf{0}. \tag{72}$$

Proof. The proof is similar to the one for Lemma 15. First, we show that a sufficient condition for Lemma 27 is that for any $\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm ad}}$, there exists a vector $\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm ol}}$ and $\{\hat\nu_\ell, \hat{\boldsymbol{u}}_\ell\}_{\ell\in\mathcal{L}(\mathcal{W}_t^{\rm lk})}$ such that $\|\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm ol}}\|_\infty \le 1-\gamma$ and

$$\boldsymbol{A}^\top_{\mathcal{M}_{\rm ol},\mathcal{X}_{\rm lk}}\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm ol}} + \boldsymbol{A}^\top_{\mathcal{M}_{\rm ad},\mathcal{X}_{\rm lk}}\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm ad}} + \sum_{\ell\in\mathcal{L}_{\rm ad}(\mathcal{W}_t^{\rm lk})\cup\mathcal{L}(\mathcal{W}_t^{\rm lk})} \left[\hat\nu_\ell \boldsymbol{c}_\ell + \boldsymbol{D}_\ell^\top\hat{\boldsymbol{u}}_\ell\right]_{\mathcal{X}_{\rm lk}} = \mathbf{0}. \tag{73}$$

This is immediate by simply choosing $\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm ol}} = \begin{bmatrix}\mathbf{0}^\top & \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm ol}}^\top\end{bmatrix}^\top$ and $\hat\nu_\ell = 0$ and $\hat{\boldsymbol{u}}_\ell = \mathbf{0}$ for $\ell\in\mathcal{L}_{\rm sf}$. Since it is guaranteed that there exist $\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm ol}}$ and $\{\hat\nu_\ell, \hat{\boldsymbol{u}}_\ell\}_{\ell\in\mathcal{L}(\mathcal{W}_t^{\rm lk})}$ to satisfy (73) under the condition that $\alpha^{\rm SOCP}_{\mathcal{W}_t^{\rm if}\to\mathcal{W}_t^{\rm lk}} \le 1-\gamma$, the claim is proved. □

Theorem 28 (Robust SE with ($\mathrm{S}^{(1)}$: $\ell_2\ell_1$) for tree decomposition)**.** *Consider a tree decomposition $\mathcal{T}$ and the associated set partitioning (c.f., Def. 22). Suppose that the infected bags form a subtree of $\mathcal{T}$, and that there exists a link bag $\mathcal{W}_t^{\rm lk}$ that is adjacent to only one infected bag $\mathcal{W}_t^{\rm if}$. Given the measurements $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_\natural + \boldsymbol{w}_\natural + \boldsymbol{b}_\natural$, where $\boldsymbol{w}_\natural$ has independent entries with zero mean and subgaussian parameter σ and $\mathrm{supp}(\boldsymbol{b}_\natural)\subseteq\mathcal{M}_{\rm if}$, suppose that the rows of $\boldsymbol{A}$ are normalized (c.f., Def. 1) and the regularization parameter λ is chosen such that*

$$\lambda > \frac{2}{n_m\gamma}\sqrt{2\sigma^2\log n_m}. \tag{74}$$

In addition, assume that the following conditions hold:

- *(Full column rank for the safe and boundary region) $\boldsymbol{A}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm lk},\mathcal{X}_{\rm sf}\cup\mathcal{X}_{\rm lk}}$ and*

  $$\boldsymbol{Q}_{\mathcal{M}_{\rm lk},\mathcal{X}_{\rm lk}} = \begin{bmatrix}\boldsymbol{A}_{\mathcal{M}_{\rm lk},\mathcal{X}_{\rm lk}} & \boldsymbol{I}^{(|\mathcal{M}_{\rm lk}|)\top}_{\mathcal{M}_{\rm ad}}\end{bmatrix}$$

  have full column rank.

- *(Localized mutual incoherence for bags) for the link bag $\mathcal{W}_t^{\rm lk}$ that is adjacent to only one infected bag $\mathcal{W}_t^{\rm if}$, we have $\alpha_{\mathcal{W}_t^{\rm if}\to\mathcal{W}_t^{\rm lk}} \le 1-\gamma$ for some $\gamma>0$.*

Then, the following properties hold for the solution to ($\mathrm{S}^{(1)}$: $\ell_2\ell_1$)*, denoted as* $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{b}})$*:*

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

2. *(Large bad data detection) Let* $\boldsymbol{A}^\circ := \begin{bmatrix} \boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}},\mathcal{X}_{\mathrm{sf}}} & \boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}},\mathcal{X}_{\mathrm{lk}}} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\mathrm{ol}},\mathcal{X}_{\mathrm{lk}}} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\mathrm{ad}},\mathcal{X}_{\mathrm{lk}}} \end{bmatrix}$ *and* $\boldsymbol{Q}^\circ_{\mathcal{M}_{\mathrm{ad}}} = \begin{bmatrix} \boldsymbol{A}^\circ & \boldsymbol{I}^{\circ\top}_{\mathcal{M}_{\mathrm{ad}}} \end{bmatrix}$*, and*

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

be a threshold value, and let $\tilde{\boldsymbol{b}}_{\mathcal{M}_{\mathrm{ad}}} = \boldsymbol{A}_{\mathcal{M}_{\mathrm{ad}},\mathcal{X}_{\mathrm{if}}} (\boldsymbol{x}_{\natural \mathrm{if}} - \hat{\boldsymbol{x}}_{\mathrm{if}})$ *be the error at the boundary. 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*

$$\|\boldsymbol{x}_{\natural \mathcal{X}_{\mathrm{sf}} \cup \mathcal{X}_{\mathrm{lk}}} - \hat{\boldsymbol{x}}_{\mathcal{X}_{\mathrm{sf}} \cup \mathcal{X}_{\mathrm{lk}}}\|_2 \leq t \frac{\sqrt{|\mathcal{X}_{\mathrm{sf}}| + |\mathcal{X}_{\mathrm{lk}}| + |\mathcal{M}_{\mathrm{ad}}|}}{C_{min}} + n_m \lambda \|\boldsymbol{I}_x (\boldsymbol{Q}^{\circ\top}_{\mathcal{M}_{\mathrm{ad}}} \boldsymbol{Q}^\circ_{\mathcal{M}_{\mathrm{ad}}})^{-1} \boldsymbol{I}_b^\top \|_{\infty,2}$$

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

Theorem 29 (SE robustness with ($\mathrm{S}^{(1)}$: $\ell_2\ell_1$-$\mathcal{K}$) for tree decomposition)**.** *Given a tree decomposition* $\mathcal{T}$ *and the associated set partitioning (c.f., Def. 22), suppose that the infected bags form a subtree of* $\mathcal{T}$ *and that there exists a link bag* $\mathcal{W}_t^{\mathrm{lk}}$ *that is adjacent to only one infected bag* $\mathcal{W}_t^{\mathrm{if}}$*. Consider the measurements* $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_\natural + \boldsymbol{w}_\natural + \boldsymbol{b}_\natural$*, where* $\boldsymbol{w}_\natural$ *has independent entries with zero mean and subgaussian parameter* σ *and* $\mathrm{supp}(\boldsymbol{b}_\natural) \subseteq \mathcal{M}_{\mathrm{if}}$*. Let* $\mathcal{K}_{\mathrm{lk}}$ *and* $\mathcal{K}_{\mathrm{if}}$ *be the subsets of SOCP constraints* $\mathcal{K}$ *restricted to the variables* $\boldsymbol{x}_{\mathrm{lk}}$ *and* $\boldsymbol{x}_{\mathrm{if}}$*, respectively, and let*

$$\tilde{\mathcal{K}}_{\mathrm{if}}(\hat{\boldsymbol{x}}_{\mathrm{lk}}) = \Bigg\{ \boldsymbol{x}_{\mathrm{if}} \Bigg| \begin{bmatrix} x_i^{mg} & x_\ell^{re} + j x_\ell^{im} \\ x_\ell^{re} - j x_\ell^{im} & x_j^{mg} \end{bmatrix} \succeq 0, \quad \forall \ell = (i,j) \in \mathcal{L}_{\mathrm{if}} \cup \mathcal{L}_{\mathrm{ad}}(\mathcal{W}_t^{\mathrm{lk}}), \text{ where } x_i^{mg} = \hat{x}_i^{mg} \quad \forall i \in \mathcal{N}_{\mathrm{ad}}(\mathcal{W}_t^{\mathrm{lk}}, \mathcal{W}_t^{\mathrm{if}}) \Bigg\},$$

be the confined feasible set for $\boldsymbol{x}_{\mathrm{if}}$*, which fixes the boundary variables* $\hat{\boldsymbol{x}}_{\mathrm{lk}}$ *in the SOCP constraints. Suppose that rows of* $\boldsymbol{A}$ *are normalized (c.f., Def. 1), and the regularization parameter* λ *is chosen such that*

$$\lambda > \frac{2}{n_m \gamma} \sqrt{2\sigma^2 \log n_m}. \tag{75}$$

In addition, suppose that the following conditions hold:

- *(Full column rank for the safe and boundary region)* $\boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}} \cup \mathcal{M}_{\mathrm{lk}},\mathcal{X}_{\mathrm{sf}} \cup \mathcal{X}_{\mathrm{lk}}}$ *and*

$$\boldsymbol{Q}_{\mathcal{M}_{\mathrm{lk}},\mathcal{X}_{\mathrm{lk}}} = \begin{bmatrix} \boldsymbol{A}_{\mathcal{M}_{\mathrm{lk}},\mathcal{X}_{\mathrm{lk}}} & \boldsymbol{I}^{(|\mathcal{M}_{\mathrm{lk}}|)\top}_{\mathcal{M}_{\mathrm{ad}}} \end{bmatrix}$$

have full column rank.

- *(Localized mutual incoherence for bags) for the link bag $\mathcal{W}_t^{\mathrm{lk}}$ that is adjacent to only one infected bag $\mathcal{W}_t^{\mathrm{if}}$, we have $\alpha^{\mathrm{SOCP}}_{\mathcal{W}_t^{\mathrm{if}} \to \mathcal{W}_t^{\mathrm{lk}}} \leq 1 - \gamma$ for some $\gamma > 0$.*
- *(Nonbinding SOCP constraints in the boundary) the solution for the attacked states satisfies $\hat{\boldsymbol{x}}_{\mathrm{if}} \in \tilde{\mathcal{K}}_{\mathrm{if}}(\boldsymbol{x}_{\natural \mathrm{lk}})$.*

Then, the following properties hold for the solution to ($\mathrm{S}^{(1)}$: $\ell_2\ell_1$)*, denoted as* $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{b}})$*:*

1. *(No false inclusion) The solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{b}})$ has no false bad data inclusion (i.e., $\mathrm{supp}(\hat{\boldsymbol{b}}) \subset \mathrm{supp}(\boldsymbol{b}_\natural)$) with probability greater than $1 - \frac{c_0}{n_m}$, for some constant $c_0 > 0$.*
2. *(Large bad data detection) Let $\boldsymbol{A}^\circ := \begin{bmatrix} \boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}}, \mathcal{X}_{\mathrm{sf}}} & \boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}}, \mathcal{X}_{\mathrm{lk}}} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\mathrm{ol}}, \mathcal{X}_{\mathrm{lk}}} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\mathrm{ad}}, \mathcal{X}_{\mathrm{lk}}} \end{bmatrix}$ and $\boldsymbol{Q}^\circ_{\mathcal{M}_{\mathrm{ad}}} = \begin{bmatrix} \boldsymbol{A}^\circ & \boldsymbol{I}^\top_{\mathcal{M}_{\mathrm{ad}}} \end{bmatrix}$, and*

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

   be a threshold value, and let $\tilde{\boldsymbol{b}}_{\mathcal{M}_{\mathrm{ad}}} = \boldsymbol{A}_{\mathcal{M}_{\mathrm{ad}}, \mathcal{X}_{\mathrm{if}}} (\boldsymbol{x}_{\natural \mathrm{if}} - \hat{\boldsymbol{x}}_{\mathrm{if}})$ be the error at the boundary. 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*

$$\| \boldsymbol{x}_{\natural \mathcal{X}_{\mathrm{sf}} \cup \mathcal{X}_{\mathrm{lk}}} - \hat{\boldsymbol{x}}_{\mathcal{X}_{\mathrm{sf}} \cup \mathcal{X}_{\mathrm{lk}}} \|_2 \leq t \frac{\sqrt{|\mathcal{X}_{\mathrm{sf}}| + |\mathcal{X}_{\mathrm{lk}}| + |\mathcal{M}_{\mathrm{ad}}|}}{C_{min}} + n_m \lambda \| \boldsymbol{I}_x (\boldsymbol{Q}^{\circ\top}_{\mathcal{M}_{\mathrm{ad}}} \boldsymbol{Q}^\circ_{\mathcal{M}_{\mathrm{ad}}})^{-1} \boldsymbol{Q}^{\circ\top}_{\mathcal{M}_{\mathrm{ad}}} \|_{\infty,2}$$

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

The proofs of Theorems 28 and 29 are similar to those of Theorems 11 and 18 in Section 5 and are omitted for brevity. As shown in our analysis, tree decomposition provides an efficient way to define the boundary between infected and safe nodes. Tree decomposition has been employed in semidefinite programming (SDP) to efficiently deal with network with chordal sparsity [15]. The smaller the treewidth, the faster it is to solve SDP [21]. Our analysis shows that with smaller treewidth, it is generally easier to certify robustness for SE. This is mainly due to the fact that the adhesion set is bounded by the treewidth, which limits the number of nodes that an infected bag can influence.

4 Experimental details

Noisy measurements: For each simulation, we randomly generate dense noise $\boldsymbol{w}$ and sparse bad data $\boldsymbol{b}$, and add them to the clean data according to (4). The dense noise for each measurement is zero-mean Gaussian variable, with standard deviation of 1e-5 (per unit) for voltage magnitude measurements and 0.005 (per unit) for all the other measurements. The difference in standard deviation is due to the fact that voltage magnitude sensors have higher standards of accuracy compared to power meters. For the sparse bad data, its support is randomly selected among the line measurements. We randomly select a set of lines, whose branch flow measurements are all compromised accordingly. The values for the sparse noise can be arbitrarily large, and we assume these parameters are uniformly chosen from the set $[-4.25, -3.75] \cup [3.75, 4.25]$ (per unit).

Performance metrics: We use the root-mean-square error (RMSE) as the metric for estimation accuracy, which is defined as $\sqrt{\frac{1}{n_b}\sum_{i\in\mathcal{N}}|v_i-\hat{v}_i|^2}$, where v_i and $\hat{v}_i$ are the true and estimated complex voltage at bus $i\in\mathcal{N}$. To evaluate the bad data detection accuracy, we use the F1 score, which is defined as $\frac{2*\text{precision}\times\text{recall}}{\text{precision}+\text{recall}}$, where *precision* is given by $\frac{\#\text{True positives }|\mathcal{J}\cap\hat{\mathcal{J}}|}{\#\text{Conditional positives }|\hat{\mathcal{J}}|}$, and *recall* is given by $\frac{\#\text{True positives }|\mathcal{J}\cap\hat{\mathcal{J}}|}{\#\text{Conditional positives }|\mathcal{J}|}$, and $\mathcal{J}$ and $\hat{\mathcal{J}}$ denote the true and estimated supports of bad data (# indicates the number of elements). The F1 score is the harmonic average of the precision and recall, which reaches its best value at 1 (perfect precision and recall) and worst at 0.

Experimental setup: We evaluate the proposed method (step-1 estimators include ($\text{S}^{(1)}$: ℓ_1), ($\text{S}^{(1)}$: $\ell_2\ell_1$), ($\text{S}^{(1)}$: ℓ_1-$\mathcal{K}$) or ($\text{S}^{(1)}$: $\ell_2\ell_1$-$\mathcal{K}$)) combined with step-2 recovery method ($\text{S}^{(2)}$: ℓ_2) or ($\text{S}^{(2)}$: $\ell_2\ell_1$)), and compare it with the current practice of nonlinear least square (NLS) method based on Newton's algorithm. We use SeDuMi [13] as the optimization solver and the MATPOWER implementation of NLS. Throughout the experiment, we choose λ in ($\text{S}^{(1)}$: $\ell_2\ell_1$) to be $3\times10^{-4}/n_m$, λ_2 in ($\text{S}^{(1)}$: $\ell_2\ell_1$-$\mathcal{K}$) to be 0.1, and a bad data detection threshold of 0.01 for stage-1 estimators. After the removal of bad data (i.e., cleaning step), we perform the estimation with the remaining data. All the experiments are performed on a standard laptop with 3.3GHz Intel Core i7 and 16GB memory.

Convergence issue of Newton's method: We performed a simple experiment, where there is no noise in the measurements, and we use both Newton's method and our proposed method to estimate the state for the IEEE 300-bus system [20]. Since Newton's method depends on the initial point, we randomly generate an initial point, where we add a complex vector on top of the ground truth. The magnitude of each entry is uniformly chosen from $[1-\tau, 1+\tau]$, and angle (in degrees) uniformly chosen from $[-100\times\tau, 100\times\tau]$. We increase τ to enlarge the initialization distance. As shown in Figure 2, as we increase τ, Newton's method becomes less and less reliable. This can be due to several factors, for example, if the initial point is far from the ground truth, the algorithm can become stuck at a local optimal. On the contrary, our proposed method based on ($\text{S}^{(1)}$: $\ell_2\ell_1$) does not depend on the initial point and can recover the ground truth for all the experiments.

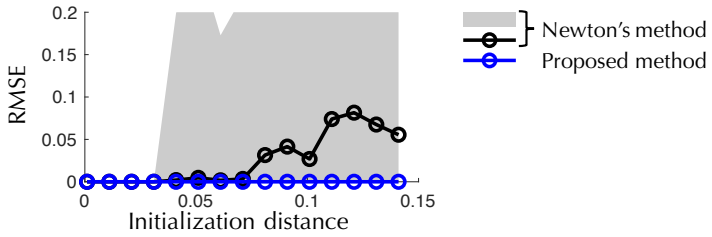

Figure 2: Plots of RMSE against initilization distance τ for Newton's method. The RMSE is averaged over 20 simulations. For the Newton's method, we show both the mean performance (circled line) and the min/max range (black shades).

Simulations on measurement redundancy: In the main paper, we demonstrated the performance of ($\text{S}^{(1)}$: $\ell_2\ell_1$-$\mathcal{K}$) for different sensor measurement profiles. We have tested three different methods to add additional sensors: the first method (Method 1) starts from a spanning tree of the network and incrementally adds a set of lines to the tree. In this method, each bus is equipped with only voltage magnitude measurements, and each line has 3 out of 4 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 measurements, 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. In Figure 3, we compared the performance of ($S^{(1)}$: $\ell_2\ell_1$) with ($S^{(1)}$: $\ell_2\ell_1$-$\mathcal{K}$) in terms of both estimation accuracy and bad data detection rates. It can be seen that ($S^{(1)}$: $\ell_2\ell_1$-$\mathcal{K}$) consistently outperforms ($S^{(1)}$: $\ell_2\ell_1$) at different redundancy rates. We can also observe that Method 1 is more efficient in terms of improvement of performance with additional sensors.

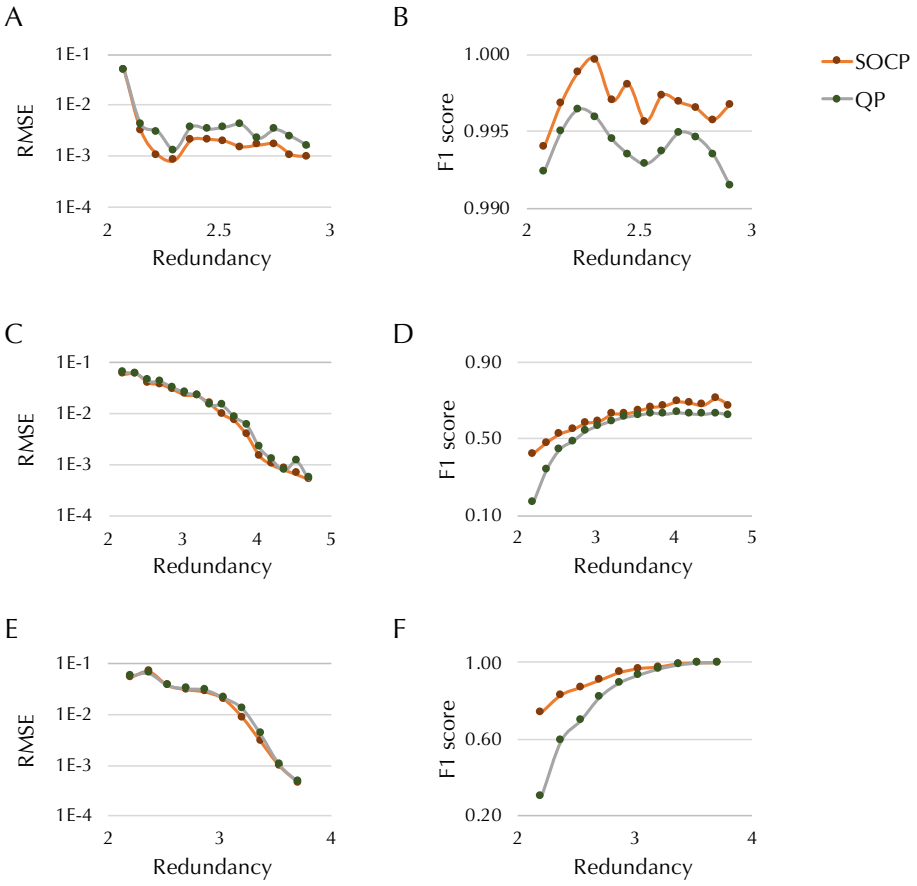

Figure 3: Performance of proposed algorithms with different rates of measurement redundancy. Plots for different methods to add measurements: **(A, B)** Method 1, **(C, D)** Method 2, **(E, F)** Method 3. Results for ($S^{(1)}$: $\ell_2\ell_1$-$\mathcal{K}$) (red) and ($S^{(1)}$: $\ell_2\ell_1$) (green) are shown, which are averaged over 100 independent simulations.

Visualization of vulnerability maps for different measurement profiles: In Figure 9 from the main text, we show statistics regarding vulnerability index and critical index for different measurement profiles. Figures 4 and 5 show the geographical distributions of VI and CI, respectively. It can be seen that ($S^{(1)}$: $\ell_2\ell_1$-$\mathcal{K}$) is consistently more robust in terms of VI and CI than ($S^{(1)}$: $\ell_2\ell_1$). This is also theoretically proven in Proposition 16. We also see that the more vulnerable lines exist, the higher the bus critical index tends to be. By comparing Figure **(B, G)** with Figure **(C, H)**, we see that the inclusion of nodal power injections is likely to cause vulnerable lines. By including more branch flow measurements, as shown in

Figure **(A, C, D)** and Figure **(F, H, I)**, or more voltage magnitude measurements, as shown in Figure **(B, C, E)** or Figure **(G, H, J)**, it is more likely to robustify the network.

5 Proofs

5.1 Proof of Theorem 11

For an arbitrary set of attacked measurements $\mathcal{M}_{\rm at}$, their boundary $\mathcal{M}_{\rm bd} \coloneqq \mathcal{M}_{\rm bi} \cup \mathcal{M}_{\rm bo}$ and unaffected measurements $\mathcal{M}_{\rm sf}$, as well as the associated variables $\boldsymbol{x}_{\rm at}$, $\boldsymbol{x}_{\rm bd}$ and $\boldsymbol{x}_{\rm sf}$, respectively, we design the 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 $\hat{\boldsymbol{x}} = \begin{bmatrix}\hat{\boldsymbol{x}}_{\rm sf}^\top & \hat{\boldsymbol{x}}_{\rm bd}^\top & \hat{\boldsymbol{x}}_{\rm at}^\top\end{bmatrix}^\top$ and $\hat{\boldsymbol{b}} = \begin{bmatrix}\boldsymbol{0}^\top & \boldsymbol{0}^\top & \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}}^\top & \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm at}}^\top\end{bmatrix}^\top$ by solving the following program:

$$\min_{\boldsymbol{b}\in\mathbb{R}^{n_m},\boldsymbol{x}\in\mathbb{R}^{n_x}} \frac{1}{2n_m}\left\|\begin{bmatrix}\boldsymbol{y}_{\mathcal{M}_{\rm sf}}\\ \boldsymbol{y}_{\mathcal{M}_{\rm bo}}\\ \boldsymbol{y}_{\mathcal{M}_{\rm bi}}\\ \boldsymbol{y}_{\mathcal{M}_{\rm at}}\end{bmatrix}-\begin{bmatrix}\boldsymbol{A}_{\mathcal{M}_{\rm sf},\mathcal{X}_{\rm sf}} & \boldsymbol{A}_{\mathcal{M}_{\rm sf},\mathcal{X}_{\rm bd}} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\rm bo},\mathcal{X}_{\rm bd}} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\rm bi},\mathcal{X}_{\rm bd}} & \boldsymbol{A}_{\mathcal{M}_{\rm bi},\mathcal{X}_{\rm at}}\\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\rm at},\mathcal{X}_{\rm at}}\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_{\rm sf}\\ \boldsymbol{x}_{\rm bd}\\ \boldsymbol{x}_{\rm at}\end{bmatrix}-\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\\ \boldsymbol{b}_{\mathcal{M}_{\rm bi}}\\ \boldsymbol{b}_{\mathcal{M}_{\rm at}}\end{bmatrix}\right\|_2^2+\lambda\left\|\begin{bmatrix}\boldsymbol{b}_{\mathcal{M}_{\rm bi}}\\ \boldsymbol{b}_{\mathcal{M}_{\rm at}}\end{bmatrix}\right\|_1, \tag{76}$$

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

$$-\frac{1}{n_m}(\boldsymbol{y}_{\mathcal{M}_{\rm at}} - \boldsymbol{A}_{\mathcal{M}_{\rm at},\mathcal{X}_{\rm at}}\hat{\boldsymbol{x}}_{\rm at} - \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm at}}) + \lambda\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm at}} = \boldsymbol{0}, \tag{77a}$$

$$-\frac{1}{n_m}\left(\boldsymbol{y}_{\mathcal{M}_{\rm bi}} - \boldsymbol{A}_{\mathcal{M}_{\rm bi},\mathcal{X}_{\rm bd}}\hat{\boldsymbol{x}}_{\rm bd} - \boldsymbol{A}_{\mathcal{M}_{\rm bi},\mathcal{X}_{\rm at}}\hat{\boldsymbol{x}}_{\rm at} - \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}}\right) + \lambda\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}} = \boldsymbol{0}. \tag{77b}$$

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

$$-\frac{1}{n_m}\left(\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}} - \hat{\boldsymbol{b}}\right) + \lambda\hat{\boldsymbol{h}} = \boldsymbol{0}, \tag{78}$$

where $\hat{\boldsymbol{x}} = \begin{bmatrix}\hat{\boldsymbol{x}}_{\mathcal{B}_{\rm sf}}^\top & \hat{\boldsymbol{x}}_{\mathcal{B}_{\rm bd}}^\top & \hat{\boldsymbol{x}}_{\mathcal{B}_{\rm at}}^\top\end{bmatrix}^\top$ and $\hat{\boldsymbol{b}} = \begin{bmatrix}\boldsymbol{0}^\top & \boldsymbol{0}^\top & \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}}^\top & \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm at}}^\top\end{bmatrix}^\top$ are solutions obtained in (76), and $\hat{\boldsymbol{h}} = \begin{bmatrix}\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}}^\top & \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bo}}^\top & \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}}^\top & \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm at}}^\top\end{bmatrix}^\top$ where $(\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}}, \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm at}})$ are given in (77). 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.

Lemma 30. *If the PDW procedure succeeds, then $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{b}})$ that is optimal for* (76) *is also optimal for* ($\mathrm{S}^{(1)}$: $\ell_2\ell_1$). *Furthermore, for any optimal solution $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{b}})$, if $\hat{\boldsymbol{x}}_{\rm at} = \tilde{\boldsymbol{x}}_{\rm at}$, we must have $\hat{\boldsymbol{x}}_{\rm sf} = \tilde{\boldsymbol{x}}_{\rm sf}$ and $\hat{\boldsymbol{x}}_{\rm bd} = \tilde{\boldsymbol{x}}_{\rm bd}$ (i.e., uniqueness property in the weak sense).*

Proof. The KKT conditions of ($\mathrm{S}^{(1)}$: $\ell_2\ell_1$) for a given primal-dual pair $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{b}})$ and $\hat{\boldsymbol{h}}$ are given by:

$$\boldsymbol{A}^\top(\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}} - \hat{\boldsymbol{b}}) = \boldsymbol{0}, \tag{79a}$$

$$-\frac{1}{n_m}\left(\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}} - \hat{\boldsymbol{b}}\right) + \lambda\hat{\boldsymbol{h}} = \boldsymbol{0}, \tag{79b}$$

$$\|\hat{\boldsymbol{h}}\|_\infty \le 1 \tag{79c}$$

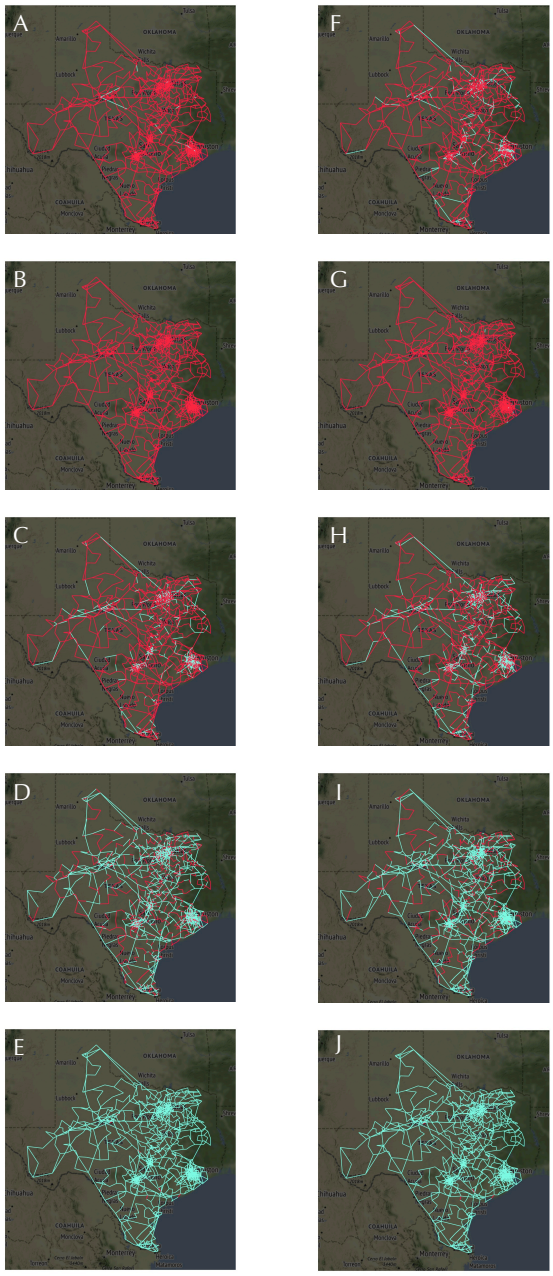

Figure 4: Vulnerability maps for different measurement profiles and optimization techniques. **(A–E)** and **(F–J)** are series of maps without and with the SOCs, respectively. **(A, F)**, **(C, H)** and **(D, I)** correspond to PV or PQ nodal measurements together with 2, 3, and 4 branch power flows, respectively. **(B, G)** and **(E, J)** correspond to only PQ or only voltage magnitude nodal measurements with 3 branch power flows, respectively.

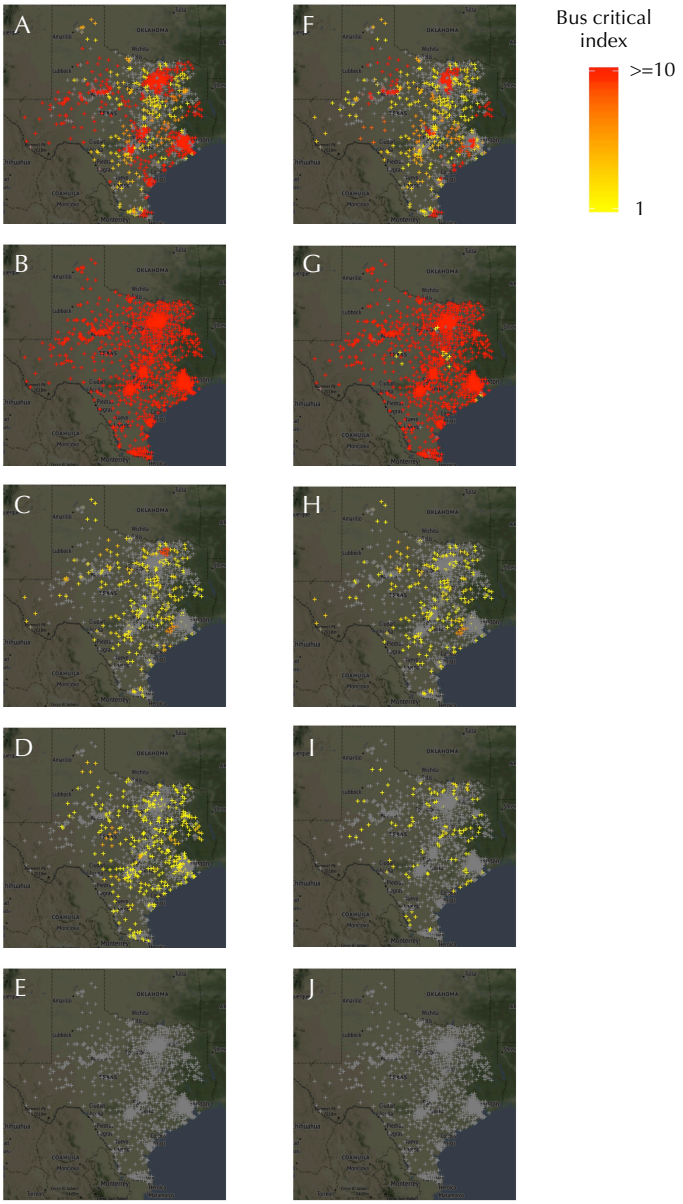

Figure 5: Bus critical index maps for different measurement profiles and optimization techniques. **(A–E)** and **(F–J)** are series of maps without and with the SOCs, respectively. **(A, F)**, **(C, H)** and **(D, I)** correspond to PV or PQ nodal measurements together with 2, 3, and 4 branch power flows, respectively. **(B, G)** and **(E, J)** correspond to only PQ or only voltage magnitude nodal measurements with 3 branch power flows, respectively. Color indicates low (yellowish) to high (reddish) critical index. If the critical index is 0, which occurs when attacking the bus does not affect any of its neighbors, the grey color is shown.

If PDW succeeds, then the optimality conditions (79) are satisfied, which certify the optimality of $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{b}})$. 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 we have

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

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 optimality conditions in (79), 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

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

due to convexity. Therefore, $\|\tilde{\boldsymbol{b}}\|_1 \leq \left\langle \hat{\boldsymbol{h}}, \tilde{\boldsymbol{b}} \right\rangle$. Since by Holder's inequality, we also have $\left\langle \hat{\boldsymbol{h}}, \tilde{\boldsymbol{b}} \right\rangle \leq \|\hat{\boldsymbol{h}}\|_\infty \|\tilde{\boldsymbol{b}}\|_1$, and $\|\hat{\boldsymbol{h}}\|_\infty \leq 1$, it holds that $\|\tilde{\boldsymbol{b}}\|_1 = \left\langle \hat{\boldsymbol{h}}, \tilde{\boldsymbol{b}} \right\rangle$. Since by the success of PDW, $\|\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}}\|_\infty < 1$ and $\|\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bo}}\|_\infty < 1$, we have $\tilde{\boldsymbol{b}}_j = 0$ for all $j \in \mathcal{M}_{\rm sf} \cup \mathcal{M}_{\rm bo}$. To show the weak uniqueness, let $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{b}})$ be another optimal solution, and assume that $\hat{\boldsymbol{x}}_{\rm at} = \tilde{\boldsymbol{x}}_{\rm at}$. Then, by fixing $\boldsymbol{x}_{\rm at}$ in the optimization (76) as $\hat{\boldsymbol{x}}_{\rm at}$ and by the lower eigenvalue condition, the the function is strictly convex in $\boldsymbol{x}_{\rm sf}$, $\boldsymbol{x}_{\rm bd}$ and $\boldsymbol{b}_{\mathcal{M}_{\rm bi}}$. □

Proof of Theorem 11

Proof. **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 $\hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bo}} = \boldsymbol{b}_{\natural \mathcal{M}_{\rm bo}} = \boldsymbol{0}$. In the following, we allow the optimal solution $\hat{\boldsymbol{x}}_{\rm at}$ and $\hat{\boldsymbol{b}}_{\mathcal{M}_{\rm at}}$ of (76) to take any value. 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 (76) and solve the following smaller program:

$$\min_{\boldsymbol{b}_{\mathcal{M}_{\rm bi}}, \boldsymbol{x}_{\rm sf}, \boldsymbol{x}_{\rm bd}} \frac{1}{2n_m} \left\| \underbrace{\begin{bmatrix} \boldsymbol{y}_{\mathcal{M}_{\rm sf}} \\ \boldsymbol{y}_{\mathcal{M}_{\rm bo}} \\ \boldsymbol{z}_{\mathcal{M}_{\rm bi}} \end{bmatrix}}_{\boldsymbol{z}^\circ} - \underbrace{\begin{bmatrix} \boldsymbol{A}_{\mathcal{M}_{\rm sf}, \mathcal{X}_{\rm sf}} & \boldsymbol{A}_{\mathcal{M}_{\rm sf}, \mathcal{X}_{\rm bd}} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\rm bo}, \mathcal{X}_{\rm bd}} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\rm bi}, \mathcal{X}_{\rm bd}} \end{bmatrix}}_{\boldsymbol{A}^\circ} \underbrace{\begin{bmatrix} \boldsymbol{x}_{\rm sf} \\ \boldsymbol{x}_{\rm bd} \end{bmatrix}}_{\boldsymbol{x}^\circ} - \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{b}_{\mathcal{M}_{\rm bi}} \end{bmatrix} \right\|_2^2 + \lambda \left\| \boldsymbol{b}_{\mathcal{M}_{\rm bi}} \right\|_1, \tag{80}$$

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 $\tilde{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}} = \boldsymbol{A}_{\mathcal{M}_{\rm bi}, \mathcal{X}_{\rm at}} (\boldsymbol{x}_{\natural {\rm at}} - \hat{\boldsymbol{x}}_{\rm at})$. Let $\boldsymbol{I}^\circ$ be an identity matrix of size $n_m - |\mathcal{M}_{\rm at}|$, and $\boldsymbol{x}^\circ$ and $\boldsymbol{w}^\circ$ be the subvectors of $\boldsymbol{x}$ and $\boldsymbol{w}$ indexed by $\mathcal{M}_{\rm sf} \cup \mathcal{M}_{\rm bo} \cup \mathcal{M}_{\rm bi}$, respectively. Thus, we have $\boldsymbol{z}^\circ = \boldsymbol{A}^\circ \boldsymbol{x}^\circ_\natural + \boldsymbol{w}^\circ_\natural + \boldsymbol{I}^{\circ\top}_{\mathcal{M}_{\rm bi}} \tilde{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}}$. The solution $(\boldsymbol{x}_{\rm sf}, \boldsymbol{x}_{\rm bd}, \boldsymbol{b}_{\mathcal{M}_{\rm bi}})$ of (80) is unique and coincides with that of (76) due to the lower eigenvalue condition. Thus, the zero-subgradient condition (77) is satisfied, which together with (78) can be written as:

$$-\frac{1}{n_m} \left(\begin{bmatrix} \boldsymbol{A}_{\mathcal{M}_{\rm sf}, \mathcal{X}_{\rm sf}} & \boldsymbol{A}_{\mathcal{M}_{\rm sf}, \mathcal{X}_{\rm bd}} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\rm bo}, \mathcal{X}_{\rm bd}} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\rm bi}, \mathcal{X}_{\rm bd}} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_{\natural {\rm sf}} - \hat{\boldsymbol{x}}_{\rm sf} \\ \boldsymbol{x}_{\natural {\rm bd}} - \hat{\boldsymbol{x}}_{\rm bd} \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \tilde{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}} - \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}} \end{bmatrix} \right) - \frac{1}{n_m} \begin{bmatrix} \boldsymbol{w}_{\natural \mathcal{M}_{\rm sf}} \\ \boldsymbol{w}_{\natural \mathcal{M}_{\rm bo}} \\ \boldsymbol{w}_{\natural \mathcal{M}_{\rm bi}} \end{bmatrix} + \lambda \begin{bmatrix} \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}} \\ \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bo}} \\ \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}} \end{bmatrix} = \boldsymbol{0}. \tag{81}$$

We can partition the above relation into equations indexed by $\mathcal{M}_{\rm bi}$, which can be rearranged as:

$$\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}} = \frac{1}{n_m \lambda} \begin{bmatrix} \boldsymbol{I}^\circ_{\mathcal{M}_{\rm bi}} \boldsymbol{A}^\circ & \boldsymbol{I}^\circ_{\mathcal{M}_{\rm bi}} \boldsymbol{I}^{\circ\top}_{\mathcal{M}_{\rm bi}} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}^\circ_\natural - \hat{\boldsymbol{x}}^\circ \\ \tilde{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}} - \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}} \end{bmatrix} + \frac{1}{n_m \lambda} \boldsymbol{I}^\circ_{\mathcal{M}_{\rm bi}} \boldsymbol{w}^\circ_\natural, \tag{82}$$

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

$$\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bo}} = \frac{1}{n_m\lambda}\boldsymbol{I}^\circ_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bo}}\left(\boldsymbol{A}^\circ(\boldsymbol{x}^\circ_\natural - \hat{\boldsymbol{x}}^\circ) + \boldsymbol{w}^\circ_\natural\right). \tag{83}$$

Since $\hat{\boldsymbol{x}}^\circ$ is the optimal solution of (80), it satisfies the optimality condition:

$$\boldsymbol{A}^{\circ\top}\left(\boldsymbol{A}^\circ(\boldsymbol{x}^\circ_\natural - \hat{\boldsymbol{x}}^\circ) + \boldsymbol{w}^\circ_\natural + \boldsymbol{I}^{\circ\top}_{\mathcal{M}_{\rm bi}}(\tilde{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}} - \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}})\right) = \boldsymbol{0} \tag{84}$$

Combining (82), (83) and (84) and after some elementary operations, we have

$$\boldsymbol{A}^{\circ\top}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bo}}\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bo}} + \boldsymbol{A}^{\circ\top}_{\mathcal{M}_{\rm bi}}\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}} = \boldsymbol{0}. \tag{85}$$

By Lemma 9, for any $\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}} \in \partial\|\hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}}\|_1$, there always exists $\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bo}}$ such that $\|\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bo}}\|_\infty < 1$. Thus, the strict feasibility condition is satisfied deterministically.

Part 2): By the lower eigenvalue condition the and definition of $\boldsymbol{Q}^\circ_{\mathcal{M}_{\rm bi}} = \begin{bmatrix}\boldsymbol{A}^\circ & \boldsymbol{I}^{\circ\top}_{\mathcal{M}_{\rm bi}}\end{bmatrix}$, we can solve (82) and (84):

$$\boldsymbol{\Delta} \coloneqq \begin{bmatrix}\boldsymbol{x}^\circ_\natural - \hat{\boldsymbol{x}}^\circ \\ \tilde{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}} - \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}}\end{bmatrix} = -(\boldsymbol{Q}^{\circ\top}_{\mathcal{M}_{\rm bi}}\boldsymbol{Q}^\circ_{\mathcal{M}_{\rm bi}})^{-1}\boldsymbol{Q}^{\circ\top}_{\mathcal{M}_{\rm bi}}\boldsymbol{w}^\circ_\natural + n_m\lambda(\boldsymbol{Q}^{\circ\top}_{\mathcal{M}_{\rm bi}}\boldsymbol{Q}^\circ_{\mathcal{M}_{\rm bi}})^{-1}\begin{bmatrix}\boldsymbol{0} \\ \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}}\end{bmatrix} \tag{86}$$

Let $\boldsymbol{I}_x$ and $\boldsymbol{I}_b$ denote the matrices that consist of the first $|\mathcal{X}_{\rm sf}| + |\mathcal{X}_{\rm bd}|$ rows and the last $|\mathcal{M}_{\rm bi}|$ rows of the identity matrix of size $|\mathcal{X}_{\rm sf}| + |\mathcal{X}_{\rm bd}| + |\mathcal{M}_{\rm bi}|$, respectively. Then, we can bound the estimation error $\boldsymbol{\Delta}$ in (86). 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,

$$\|\boldsymbol{I}_b\boldsymbol{\Delta}\|_\infty \le \|\boldsymbol{I}_b(\boldsymbol{Q}^{\circ\top}_{\mathcal{M}_{\rm bi}}\boldsymbol{Q}^\circ_{\mathcal{M}_{\rm bi}})^{-1}\boldsymbol{Q}^{\circ\top}_{\mathcal{M}_{\rm bi}}\boldsymbol{w}^\circ_\natural\|_\infty + n_m\lambda\|\boldsymbol{I}_b(\boldsymbol{Q}^{\circ\top}_{\mathcal{M}_{\rm bi}}\boldsymbol{Q}^\circ_{\mathcal{M}_{\rm bi}})^{-1}\boldsymbol{I}_b^\top\|_\infty. \tag{87}$$

Since the second term is deterministic, we will now bound the first term. By the normalized measurement condition (1) (we assume all measurement vectors are normalized by 1 without loss of generality) and the lower eigenvalue condition, each entry of $(\boldsymbol{Q}^{\circ\top}_{\mathcal{M}_{\rm bi}}\boldsymbol{Q}^\circ_{\mathcal{M}_{\rm bi}})^{-1}\boldsymbol{Q}^{\circ\top}_{\mathcal{M}_{\rm bi}}\boldsymbol{w}^\circ_\natural$ is zero-mean sub-Gaussian with parameter at most

$$\sigma^2\|(\boldsymbol{Q}^{\circ\top}_{\mathcal{M}_{\rm bi}}\boldsymbol{Q}^\circ_{\mathcal{M}_{\rm bi}})^{-1}\|_2 \le \frac{\sigma^2}{C_{\rm min}}. \tag{88}$$

Thus, by the union bound, we have

$$\mathbb{P}\left(\|\boldsymbol{I}_b(\boldsymbol{Q}^{\circ\top}_{\mathcal{M}_{\rm bi}}\boldsymbol{Q}^\circ_{\mathcal{M}_{\rm bi}})^{-1}\boldsymbol{Q}^{\circ\top}_{\mathcal{M}_{\rm bi}}\boldsymbol{w}^\circ_\natural\|_\infty > t\right) \le 2\exp\left(-\frac{C_{\rm min}t^2}{2\sigma^2} + \log|\mathcal{M}_{\rm bi}|\right). \tag{89}$$

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

$$\|\tilde{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}} - \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}}\|_\infty \le n_m\lambda\left(\frac{1}{2\sqrt{C_{\rm min}}} + \|\boldsymbol{I}_b(\boldsymbol{Q}^{\circ\top}_{\mathcal{M}_{\rm bi}}\boldsymbol{Q}^\circ_{\mathcal{M}_{\rm bi}})^{-1}\boldsymbol{I}_b^\top\|_\infty\right) \tag{90}$$

with probability greater than $1 - 2\exp(-c_2n_m^2\lambda^2)$. This indicates that all bad data entries greater than

$$g(\lambda) = n_m\lambda\left(\frac{1}{2\sqrt{C_{\rm min}}} + \|\boldsymbol{I}_b(\boldsymbol{Q}^{\circ\top}_{\mathcal{M}_{\rm bi}}\boldsymbol{Q}^\circ_{\mathcal{M}_{\rm bi}})^{-1}\boldsymbol{I}_b^\top\|_\infty\right) \tag{91}$$

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

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

$$\|\boldsymbol{I}_x\boldsymbol{\Delta}\|_2 \leq \|\boldsymbol{I}_x(\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ})^{-1}\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\boldsymbol{w}_\natural^\circ\|_2 + n_m\lambda\|\boldsymbol{I}_x(\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ})^{-1}\boldsymbol{I}_b^\top\|_{\infty,2}. \tag{92}$$

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

$$\mathbb{P}\left(\|\boldsymbol{I}_x(\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ})^{-1}\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\boldsymbol{w}_\natural^\circ\|_2 > \|\boldsymbol{I}_x(\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ})^{-1}\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\|_F + t\|\boldsymbol{I}_x(\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ})^{-1}\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\|_2\right)$$

is upper bounded by $\exp\left(-\frac{c_1t^2}{\sigma^4}\right)$. Since

$$\|\boldsymbol{I}_x(\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ})^{-1}\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\|_F \leq \|\boldsymbol{I}_x\|_2\|(\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ})^{-1}\|_2\|\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\|_F \leq \frac{\sqrt{|\mathcal{X}_{\mathrm{sf}}| + |\mathcal{X}_{\mathrm{bd}}| + |\mathcal{M}_{\mathrm{bi}}|}}{C_{\mathrm{min}}}$$

due to the lower eigenvalue condition and the normalized measurement condition, and similarly it holds that

$$\|\boldsymbol{I}_x(\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ})^{-1}\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\|_2 \leq \|\boldsymbol{I}_x\|_2\|(\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ})^{-1}\|_2\|\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\|_F \leq \frac{\sqrt{|\mathcal{X}_{\mathrm{sf}}| + |\mathcal{X}_{\mathrm{bd}}| + |\mathcal{M}_{\mathrm{bi}}|}}{C_{\mathrm{min}}}.$$

Moreover,

$$\mathbb{P}\left(\|\boldsymbol{I}_x(\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ})^{-1}\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\boldsymbol{w}_\natural^\circ\|_2 > t\frac{\sqrt{|\mathcal{X}_{\mathrm{sf}}| + |\mathcal{X}_{\mathrm{bd}}| + |\mathcal{M}_{\mathrm{bi}}|}}{C_{\mathrm{min}}}\right) \leq \exp\left(-\frac{c_1t^2}{\sigma^4}\right).$$

Together, we conclude that

$$\|\boldsymbol{x}_\natural - \hat{\boldsymbol{x}}\|_2 \leq t\frac{\sqrt{|\mathcal{X}_{\mathrm{sf}}| + |\mathcal{X}_{\mathrm{bd}}| + |\mathcal{M}_{\mathrm{bi}}|}}{C_{\mathrm{min}}} + n_m\lambda\|\boldsymbol{I}_x(\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ})^{-1}\boldsymbol{I}_b^\top\|_{\infty,2} \tag{93}$$

with probability greater than $1 - \exp\left(-\frac{c_1t^2}{\sigma^4}\right)$.

□

Lemma 31. *Suppose that $\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ}$ is invertible, where $\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ} = \begin{bmatrix}\boldsymbol{A}^\circ & \boldsymbol{I}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\end{bmatrix}$. Then, it holds that*

$$\boldsymbol{I}_{\mathcal{M}_{\mathrm{sf}}\cup\mathcal{M}_{\mathrm{bo}}}\boldsymbol{A}^\circ\boldsymbol{I}_x(\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ})^{-1}\boldsymbol{I}_b^\top = -\boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}}\cup\mathcal{M}_{\mathrm{bo}}}^{\top+}\boldsymbol{A}_{\mathcal{M}_{\mathrm{bi}}}^\top. \tag{94}$$

Proof. By the definition of $\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ}$ and block matrix inversion formula, we have

$$\begin{aligned}
&\boldsymbol{I}_x(\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ})^{-1}\boldsymbol{I}_b^\top \\
&= -(\boldsymbol{A}^{\circ\top}\boldsymbol{A}^\circ)^{-1}\boldsymbol{A}_{\mathcal{M}_{\mathrm{bi}}}^\top(\boldsymbol{I} - \boldsymbol{A}_{\mathcal{M}_{\mathrm{bi}}}(\boldsymbol{A}^{\circ\top}\boldsymbol{A}^\circ)^{-1}\boldsymbol{A}_{\mathcal{M}_{\mathrm{bi}}}^\top)^{-1} \\
&= -(\boldsymbol{A}^{\circ\top}\boldsymbol{A}^\circ)^{-1}\boldsymbol{A}_{\mathcal{M}_{\mathrm{bi}}}^\top(\boldsymbol{I} + \boldsymbol{A}_{\mathcal{M}_{\mathrm{bi}}}(\boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}}\cup\mathcal{M}_{\mathrm{bo}}}^\top\boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}}\cup\mathcal{M}_{\mathrm{bo}}})^{-1}\boldsymbol{A}_{\mathcal{M}_{\mathrm{bi}}}^\top) \\
&= -(\boldsymbol{A}^{\circ\top}\boldsymbol{A}^\circ)^{-1}(\boldsymbol{I} + \boldsymbol{A}_{\mathcal{M}_{\mathrm{bi}}}^\top\boldsymbol{A}_{\mathcal{M}_{\mathrm{bi}}}(\boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}}\cup\mathcal{M}_{\mathrm{bo}}}^\top\boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}}\cup\mathcal{M}_{\mathrm{bo}}})^{-1})\boldsymbol{A}_{\mathcal{M}_{\mathrm{bi}}}^\top \\
&= -(\boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}}\cup\mathcal{M}_{\mathrm{bo}}}^\top\boldsymbol{A}_{\mathcal{M}_{\mathrm{sf}}\cup\mathcal{M}_{\mathrm{bo}}})^{-1}\boldsymbol{A}_{\mathcal{M}_{\mathrm{bi}}}^\top,
\end{aligned}$$

where the first equation follows from the Sherman-Morrison-Woodbury formula and the rest are elementary operations. □

Lemma 32. *Suppose that $\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ\top}\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ}$ is invertible. Then, it holds that*

$$\boldsymbol{I}_b(\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ\top}\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ})^{-1}\boldsymbol{I}_b^\top = \boldsymbol{I} + \boldsymbol{A}_{\mathcal{M}_{\rm bi}}(\boldsymbol{A}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bo}}^\top\boldsymbol{A}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bo}})^{-1}\boldsymbol{A}_{\mathcal{M}_{\rm bi}}^\top \tag{95}$$

Proof. By the definition of $\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ}$ and block matrix inversion formula, we have

$$\begin{aligned}\boldsymbol{I}_b(\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ\top}\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ})^{-1}\boldsymbol{I}_b^\top &= (\boldsymbol{I} - \boldsymbol{A}_{\mathcal{M}_{\rm bi}}(\boldsymbol{A}^{\circ\top}\boldsymbol{A}^{\circ})^{-1}\boldsymbol{A}_{\mathcal{M}_{\rm bi}}^\top)^{-1}\\ &= \boldsymbol{I} + \boldsymbol{A}_{\mathcal{M}_{\rm bi}}(\boldsymbol{A}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bo}}^\top\boldsymbol{A}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bo}})^{-1}\boldsymbol{A}_{\mathcal{M}_{\rm bi}}^\top,\end{aligned}$$

where the second equation follows from the Sherman-Morrison-Woodbury formula. □

5.2 Proof of Theorem 18

For an arbitrary set of attacked measurements $\mathcal{M}_{\rm at}$, their boundary $\mathcal{M}_{\rm bd} \coloneqq \mathcal{M}_{\rm bi}\cup\mathcal{M}_{\rm bo}$ and unaffected measurements $\mathcal{M}_{\rm sf}$, as well as the associated variables $\boldsymbol{x}_{\rm at}$, $\boldsymbol{x}_{\rm bd}$ and $\boldsymbol{x}_{\rm sf}$, respectively, we design the primal-dual witness 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 $\hat{\boldsymbol{x}} = \begin{bmatrix}\hat{\boldsymbol{x}}_{\rm sf}^\top & \hat{\boldsymbol{x}}_{\rm bd}^\top & \hat{\boldsymbol{x}}_{\rm at}^\top\end{bmatrix}^\top$ and $\hat{\boldsymbol{b}} = \begin{bmatrix}\boldsymbol{0}^\top & \boldsymbol{0}^\top & \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}}^\top & \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm at}}^\top\end{bmatrix}^\top$ by solving the following program:

$$\min_{\boldsymbol{b}\in\mathbb{R}^{n_m},\boldsymbol{x}\in\mathbb{R}^{n_x}} \quad \frac{1}{2n_m}\left\|\begin{bmatrix}\boldsymbol{y}_{\mathcal{M}_{\rm sf}}\\ \boldsymbol{y}_{\mathcal{M}_{\rm bo}}\\ \boldsymbol{y}_{\mathcal{M}_{\rm bi}}\\ \boldsymbol{y}_{\mathcal{M}_{\rm at}}\end{bmatrix} - \begin{bmatrix}\boldsymbol{A}_{\mathcal{M}_{\rm sf},\mathcal{X}_{\rm sf}} & \boldsymbol{A}_{\mathcal{M}_{\rm sf},\mathcal{X}_{\rm bd}} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\rm bo},\mathcal{X}_{\rm bd}} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\rm bi},\mathcal{X}_{\rm bd}} & \boldsymbol{A}_{\mathcal{M}_{\rm bi},\mathcal{X}_{\rm at}}\\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\rm at},\mathcal{X}_{\rm at}}\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_{\rm sf}\\ \boldsymbol{x}_{\rm bd}\\ \boldsymbol{x}_{\rm at}\end{bmatrix} - \begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\\ \boldsymbol{b}_{\mathcal{M}_{\rm bi}}\\ \boldsymbol{b}_{\mathcal{M}_{\rm at}}\end{bmatrix}\right\|_2^2 + \lambda\left\|\begin{bmatrix}\boldsymbol{b}_{\mathcal{M}_{\rm bi}}\\ \boldsymbol{b}_{\mathcal{M}_{\rm at}}\end{bmatrix}\right\|_1, \tag{96a}$$

$$\text{subject to } \boldsymbol{c}_\ell^\top\boldsymbol{x} \geq \|\boldsymbol{D}_\ell\boldsymbol{x}\|_2, \qquad \forall \ell\in\mathcal{L}, \tag{96b}$$

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

$$-\frac{1}{n_m}(\boldsymbol{y}_{\mathcal{M}_{\rm at}} - \boldsymbol{A}_{\mathcal{M}_{\rm at},\mathcal{X}_{\rm at}}\hat{\boldsymbol{x}}_{\rm at} - \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm at}}) + \lambda\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm at}} = \boldsymbol{0}, \tag{97a}$$

$$-\frac{1}{n_m}\left(\boldsymbol{y}_{\mathcal{M}_{\rm bi}} - \boldsymbol{A}_{\mathcal{M}_{\rm bi},\mathcal{X}_{\rm bd}}\hat{\boldsymbol{x}}_{\rm bd} - \boldsymbol{A}_{\mathcal{M}_{\rm bi},\mathcal{X}_{\rm at}}\hat{\boldsymbol{x}}_{\rm at} - \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}}\right) + \lambda\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}} = \boldsymbol{0}. \tag{97b}$$

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

$$-\frac{1}{n_m}\left(\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}} - \hat{\boldsymbol{b}}\right) + \lambda\hat{\boldsymbol{h}} = \boldsymbol{0}, \tag{98}$$

where $\hat{\boldsymbol{x}} = \begin{bmatrix}\hat{\boldsymbol{x}}_{\mathcal{B}_{\rm sf}}^\top & \hat{\boldsymbol{x}}_{\mathcal{B}_{\rm bd}}^\top & \hat{\boldsymbol{x}}_{\mathcal{B}_{\rm at}}^\top\end{bmatrix}^\top$ and $\hat{\boldsymbol{b}} = \begin{bmatrix}\boldsymbol{0}^\top & \boldsymbol{0}^\top & \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}}^\top & \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm at}}^\top\end{bmatrix}^\top$ are solutions obtained in (76), and $\hat{\boldsymbol{h}} = \begin{bmatrix}\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}}^\top & \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bo}}^\top & \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}}^\top & \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm at}}^\top\end{bmatrix}^\top$ where $(\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}}, \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm at}})$ are given in (77). 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.

Lemma 33. *If the PDW procedure succeeds, then $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{b}})$ that is optimal for* (96) *is also optimal for* (S$^{(1)}$: $\ell_2\ell_1$-$\mathcal{K}$). *Furthermore, for any optimal solution $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{b}})$, if $\hat{\boldsymbol{x}}_{\rm at} = \tilde{\boldsymbol{x}}_{\rm at}$, it holds that $\hat{\boldsymbol{x}}_{\rm sf} = \tilde{\boldsymbol{x}}_{\rm sf}$ and $\hat{\boldsymbol{x}}_{\rm bd} = \tilde{\boldsymbol{x}}_{\rm bd}$ (i.e., uniqueness property in the weak sense).*

Proof. The KKT conditions of (S$^{(1)}$: $\ell_2\ell_1$-$\mathcal{K}$) for a given primal-dual pair $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{b}})$ and $\hat{\boldsymbol{h}}$ are given by:

$$\frac{1}{n_m}\boldsymbol{A}^\top(\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}} - \hat{\boldsymbol{b}}) + \lambda \sum_{\ell\in\mathcal{L}}(\nu_\ell \boldsymbol{c}_\ell + \boldsymbol{D}_\ell^\top \boldsymbol{\mu}_\ell) = \boldsymbol{0}, \tag{99a}$$

$$-\frac{1}{n_m}\left(\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}} - \hat{\boldsymbol{b}}\right) + \lambda\hat{\boldsymbol{h}} = \boldsymbol{0}, \tag{99b}$$

$$\hat{\boldsymbol{h}} \in \partial\|\hat{\boldsymbol{b}}\|_1, \quad \|\hat{\boldsymbol{h}}\|_\infty \le 1 \tag{99c}$$

If PDW succeeds, then the optimality conditions (99) are satisfied, which certify the optimality of $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{b}})$. 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 optimality conditions in (99), 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

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

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 $j \in \mathcal{M}_{\rm sf} \cup \mathcal{M}_{\rm bo}$. To show the weak uniqueness, let $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{b}})$ be another optimal solution, and assume that $\hat{\boldsymbol{x}}_{\rm at} = \tilde{\boldsymbol{x}}_{\rm at}$. Then, by fixing $\boldsymbol{x}_{\rm at}$ in the optimization (96) at $\hat{\boldsymbol{x}}_{\rm at}$ and by the lower eigenvalue condition, the the function is strictly convex in $\boldsymbol{x}_{\rm sf}$, $\boldsymbol{x}_{\rm bd}$ and $\boldsymbol{b}_{\mathcal{M}_{\rm bi}}$. □

Proof of Theorem 18

Proof. **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 $\hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bo}} = \boldsymbol{b}_{\natural\mathcal{M}_{\rm bo}} = \boldsymbol{0}$. In the following, we allow the optimal solution $\hat{\boldsymbol{x}}_{\rm at}$ and $\hat{\boldsymbol{b}}_{\mathcal{M}_{\rm at}}$ of (96) to take any value as long as the nonbinding SOC constraints assumption 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 (96) and solve the following smaller program:

$$\min_{\boldsymbol{b}_{\mathcal{M}_{\rm bi}}, \boldsymbol{x}_{\rm sf}, \boldsymbol{x}_{\rm bd}} \ \frac{1}{2n_m}\left\| \underbrace{\begin{bmatrix} \boldsymbol{y}_{\mathcal{M}_{\rm sf}} \\ \boldsymbol{y}_{\mathcal{M}_{\rm bo}} \\ \boldsymbol{z}_{\mathcal{M}_{\rm bi}} \end{bmatrix}}_{\boldsymbol{z}^\circ} - \underbrace{\begin{bmatrix} \boldsymbol{A}_{\mathcal{M}_{\rm sf},\mathcal{X}_{\rm sf}} & \boldsymbol{A}_{\mathcal{M}_{\rm sf},\mathcal{X}_{\rm bd}} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\rm bo},\mathcal{X}_{\rm bd}} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\rm bi},\mathcal{X}_{\rm bd}} \end{bmatrix}}_{\boldsymbol{A}^\circ} \underbrace{\begin{bmatrix} \boldsymbol{x}_{\rm sf} \\ \boldsymbol{x}_{\rm bd} \end{bmatrix}}_{\boldsymbol{x}^\circ} - \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{b}_{\mathcal{M}_{\rm bi}} \end{bmatrix} \right\|_2^2 + \lambda\left\|\boldsymbol{b}_{\mathcal{M}_{\rm bi}}\right\|_1, \tag{100a}$$

$$\text{subject to} \quad \boldsymbol{c}_\ell^\top \boldsymbol{x} \ge \|\boldsymbol{D}_\ell \boldsymbol{x}\|_2, \qquad \forall \ell \in \mathcal{L} \setminus \mathcal{L}_{\rm at}, \tag{100b}$$

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 $\tilde{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}} = \boldsymbol{A}_{\mathcal{M}_{\rm bi},\mathcal{X}_{\rm at}}(\boldsymbol{x}_{\natural{\rm at}} - \hat{\boldsymbol{x}}_{\rm at})$. Let $\boldsymbol{I}^\circ$ be an identity matrix of size $n_m - |\mathcal{M}_{\rm at}|$, and $\boldsymbol{x}^\circ$, $\boldsymbol{c}_\ell^\circ$ and $\boldsymbol{D}_\ell^\circ$ be the subvector and submatrix of $\boldsymbol{x}$, $\boldsymbol{c}_\ell$ and $\boldsymbol{D}_\ell$ indexed by $\mathcal{X}_{\rm sf}$ and $\mathcal{X}_{\rm bd}$, respectively, and $\boldsymbol{w}^\circ$ be the subvector of $\boldsymbol{w}$ indexed by $\mathcal{M}_{\rm sf} \cup \mathcal{M}_{\rm bo} \cup \mathcal{M}_{\rm bi}$. 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 $(\boldsymbol{x}_{\rm sf}, \boldsymbol{x}_{\rm bd}, \boldsymbol{b}_{\mathcal{M}_{\rm bi}})$ of (100) is unique and coincides with that of (96) due to the lower eigenvalue condition. Thus, the zero-subgradient condition (97) is satisfied, which together with (98) can be written as:

$$-\frac{1}{n_m}\left(\begin{bmatrix} \boldsymbol{A}_{\mathcal{M}_{\rm sf},\mathcal{X}_{\rm sf}} & \boldsymbol{A}_{\mathcal{M}_{\rm sf},\mathcal{X}_{\rm bd}} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\rm bo},\mathcal{X}_{\rm bd}} \\ \boldsymbol{0} & \boldsymbol{A}_{\mathcal{M}_{\rm bi},\mathcal{X}_{\rm bd}} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_{\natural \rm sf} - \hat{\boldsymbol{x}}_{\rm sf} \\ \boldsymbol{x}_{\natural \rm bd} - \hat{\boldsymbol{x}}_{\rm bd} \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \tilde{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}} - \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}} \end{bmatrix}\right) - \frac{1}{n_m}\begin{bmatrix} \boldsymbol{w}_{\natural \mathcal{M}_{\rm sf}} \\ \boldsymbol{w}_{\natural \mathcal{M}_{\rm bo}} \\ \boldsymbol{w}_{\natural \mathcal{M}_{\rm bi}} \end{bmatrix} + \lambda \begin{bmatrix} \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}} \\ \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bo}} \\ \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}} \end{bmatrix} = \boldsymbol{0}. \tag{101}$$

We can partition the above relation into equations indexed by $\mathcal{M}_{\rm bi}$, which can be rearranged as:

$$\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}} = \frac{1}{n_m \lambda}\begin{bmatrix} \boldsymbol{I}_{\mathcal{M}_{\rm bi}}^\circ \boldsymbol{A}^\circ & \boldsymbol{I}_{\mathcal{M}_{\rm bi}}^\circ \boldsymbol{I}_{\mathcal{M}_{\rm bi}}^{\circ\top} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_\natural^\circ - \hat{\boldsymbol{x}}^\circ \\ \tilde{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}} - \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}} \end{bmatrix} + \frac{1}{n_m \lambda} \boldsymbol{I}_{\mathcal{M}_{\rm bi}}^\circ \boldsymbol{w}_\natural^\circ, \tag{102}$$

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

$$\hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bo}} = \frac{1}{n_m \lambda} \boldsymbol{I}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bo}}^\circ \left(\boldsymbol{A}^\circ(\boldsymbol{x}_\natural^\circ - \hat{\boldsymbol{x}}^\circ) + \boldsymbol{w}_\natural^\circ\right). \tag{103}$$

Since $\hat{\boldsymbol{x}}^\circ$ is the optimal solution of (100), it satisfies the optimality condition:

$$\frac{1}{n_m}\boldsymbol{A}^{\circ\top}\left(\boldsymbol{A}^\circ(\boldsymbol{x}_\natural^\circ - \hat{\boldsymbol{x}}^\circ) + \boldsymbol{w}_\natural^\circ + \boldsymbol{I}_{\mathcal{M}_{\rm bi}}^{\circ\top}(\tilde{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}} - \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}})\right) + \sum_{\ell \in \mathcal{L}_{\rm at\cap bi} \cup \mathcal{L}_{\rm bd} \cup \mathcal{L}_{\rm sf}} \hat{\nu}_\ell \boldsymbol{c}_\ell^\circ + \boldsymbol{D}_\ell^{\circ\top} \hat{\boldsymbol{u}}_\ell = \boldsymbol{0} \tag{104}$$

Combining (102), (103) and (104) and after some elementary operations, it yields that

$$\lambda \boldsymbol{A}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bo}}^{\circ\top} \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bo}} + \lambda \boldsymbol{A}_{\mathcal{M}_{\rm bi}}^{\circ\top} \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}} + \sum_{\ell \in \mathcal{L}_{\rm at\cap bi} \cup \mathcal{L}_{\rm bd} \cup \mathcal{L}_{\rm sf}} \hat{\nu}_\ell \boldsymbol{c}_\ell^\circ + \boldsymbol{D}_\ell^{\circ\top} \hat{\boldsymbol{u}}_\ell = \boldsymbol{0}. \tag{105}$$

By Lemma 15, 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 is satisfied deterministically.

Part 2): Thus, by the lower eigenvalue condition and definition of $\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^\circ = \begin{bmatrix} \boldsymbol{A}^\circ & \boldsymbol{I}_{\mathcal{M}_{\rm bi}}^{\circ\top} \end{bmatrix}$ and $\hat{\boldsymbol{h}} = \begin{bmatrix} \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bo}}^\top & \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}}^\top \end{bmatrix}^\top$, we can solve (102), (104) and (105):

$$\begin{aligned} \boldsymbol{\Delta} &\coloneqq \begin{bmatrix} \boldsymbol{x}_\natural^\circ - \hat{\boldsymbol{x}}^\circ \\ \tilde{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}} - \hat{\boldsymbol{b}}_{\mathcal{M}_{\rm bi}} \end{bmatrix} \\ &= -(\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ\top} \boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^\circ)^{-1} \boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ\top} \boldsymbol{w}_\natural^\circ + n_m \lambda (\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ\top} \boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^\circ)^{-1} \begin{bmatrix} \boldsymbol{A}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bo}}^{\circ\top} \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm sf}\cup\mathcal{M}_{\rm bo}} + \boldsymbol{A}_{\mathcal{M}_{\rm bi}}^{\circ\top} \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}} \\ \hat{\boldsymbol{h}}_{\mathcal{M}_{\rm bi}} \end{bmatrix} \\ &= -(\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ\top} \boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^\circ)^{-1} \boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ\top} \boldsymbol{w}_\natural^\circ + n_m \lambda (\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ\top} \boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^\circ)^{-1} \boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ\top} \hat{\boldsymbol{h}}, \end{aligned} \tag{106}$$

Let $\boldsymbol{I}_x$ and $\boldsymbol{I}_b$ denote the matrices that consist of the first $|\mathcal{X}_{\rm sf}| + |\mathcal{X}_{\rm bd}|$ rows and the last $|\mathcal{M}_{\rm bi}|$ rows of the identity matrix of size $|\mathcal{X}_{\rm sf}| + |\mathcal{X}_{\rm bd}| + |\mathcal{M}_{\rm bi}|$, respectively. Then, we can bound the estimation error $\boldsymbol{\Delta}$ in (86). 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,

$$\|\boldsymbol{I}_b \boldsymbol{\Delta}\|_\infty \le \|\boldsymbol{I}_b (\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ\top} \boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^\circ)^{-1} \boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ\top} \boldsymbol{w}_\natural^\circ\|_\infty + n_m \lambda \|\boldsymbol{I}_b (\boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ\top} \boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^\circ)^{-1} \boldsymbol{Q}_{\mathcal{M}_{\rm bi}}^{\circ\top}\|_\infty. \tag{107}$$

Since the second term is deterministic, we will bound the first term similar to Theorem 10. This concludes the proof

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

$$\|\boldsymbol{I}_x \boldsymbol{\Delta}\|_2 \leq \|\boldsymbol{I}_x (\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top} \boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ})^{-1} \boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top} \boldsymbol{w}_\natural^\circ\|_2 + n_m \lambda \|\boldsymbol{I}_x (\boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top} \boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ})^{-1} \boldsymbol{Q}_{\mathcal{M}_{\mathrm{bi}}}^{\circ\top}\|_{\infty,2}. \tag{108}$$

For the first term, we can apply standard sub-gaussian concentration. The second term is deterministic. Combining them together yields the results.

□

References

[1] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Mathematical Programming*, 95(1):3–51, 2003.

[2] A. Bagchi, K. Clements, P. Davis, and F. Maurais. A comparison of algorithms for least absolute value state estimation electric power networks. In *Proceedings of IEEE International Symposium on Circuits and Systems*, volume 6, pages 53–56. IEEE, 1994.

[3] 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*, 12(1):376–382, 1997.

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

[5] M. C. Ferris and T. S. Munson. Complementarity problems in GAMS and the PATH solver. *Journal of Economic Dynamics and Control*, 24(2):165–188, 2000.

[6] J.-J. Fuchs. Recovery of exact sparse representations in the presence of bounded noise. *IEEE Transactions on Information Theory*, 51(10):3601–3608, 2005.

[7] P. J. Huber. *Robust statistics*. Springer, 2011.

[8] M. Jin, J. Lavaei, and K. H. Johansson. Power grid AC-based state estimation: Vulnerability analysis against cyber attacks. *IEEE Transactions on Automatic Control*, 64(5):1784–1799, 2019.

[9] 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.

[10] 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.

[11] J. Lofberg. YALMIP: a toolbox for modeling and optimization in MATLAB. In *IEEE International Conference on Robotics and Automation*, pages 284–289, 2004.

[12] L. Mili, M. Cheniae, N. Vichare, and P. J. Rousseeuw. Robust state estimation based on projection statistics [of power systems]. *IEEE Transactions on Power Systems*, 11(2):1118–1127, 1996.

[13] J. F. Sturm. Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones. *Optimization methods and software*, 11(1-4):625–653, 1999.

[14] J. A. Tropp. Just relax: Convex programming methods for identifying sparse signals in noise. *IEEE Transactions on Information Theory*, 52(3):1030–1051, 2006.

[15] L. Vandenberghe, M. S. Andersen, et al. Chordal graphs and semidefinite optimization. *Foundations and Trends® in Optimization*, 1(4):241–433, 2015.

[16] M. J. Wainwright. Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso). *IEEE Transactions on Information Theory*, 55(5):2183–2202, 2009.

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

[18] J. Zhao, L. Mili, and R. C. Pires. Statistical and numerical robust state estimator for heavily loaded power systems. *IEEE Transactions on Power Systems*, 33(6):6904–6914, 2018.

[19] P. Zhao and B. Yu. On model selection consistency of Lasso. *Journal of Machine Learning Research*, 7(Nov):2541–2563, 2006.

[20] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas. Matpower: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Transactions on Power Systems*, 26(1):12–19, 2010.

[21] F. Zohrizadehb, C. Josza, M. Jina, R. Madanib, J. Lavaeia, and S. Sojoudia. Conic relaxations of power system optimization: Theory and algorithms.