

Sparse Error Correction From Nonlinear Measurements With Applications in Bad Data Detection for Power Networks

Weiyu Xu, *Member, IEEE*, Meng Wang, *Member, IEEE*, Jian-Feng Cai, and Ao Tang, *Senior Member, IEEE*

Abstract—In this paper, we consider the problem of sparse error correction from general nonlinear measurements, which has applications in state estimation of electrical power networks, when bad data (outliers) are present. An iterative mixed ℓ_1 and ℓ_2 convex program is used to estimate the true state by locally linearizing the nonlinear measurements. In the special case when the measurements are linear, through using the almost Euclidean property for a linear subspace, we derive a new performance bound for the state estimation error under sparse bad data and additive observation noise. As a byproduct, in this paper we provide sharp bounds on the almost Euclidean property of a linear subspace, using the “escape-through-the-mesh” theorem from geometric functional analysis. When the measurements are nonlinear, we give conditions under which the solution of the iterative algorithm converges to the true state even though the locally linearized measurements may not be the actual nonlinear measurements. We are able to use a semidefinite program to verify the conditions for convergence of the proposed iterative sparse recovery algorithms from nonlinear measurements. We then numerically evaluate our iterative convex programming approach of performing bad data detections in nonlinear electrical power networks problems.

Index Terms—System estimation, electrical power networks, bad data detection, nonlinear sparse recovery, compressed sensing.

I. INTRODUCTION

IN this paper, inspired by state estimation for nonlinear electrical power networks under bad data and additive noise, we study the problem of sparse recovery from nonlinear measurements. The static state of an electrical power network can be described by the vector of bus voltage magnitudes and angles. In smart grid power networks, due to physical constraints, indirect nonlinear measurement results of these quantities are sent to the central control center, where the state estimation of electric power network is performed. On the one hand, these measure-

ment results contain common small additive noises due to the accuracy of meters and equipments. On the other hand, more severely, these results can contain bad data (sometimes referred to as “gross errors”, “outliers”) due to faulty sensors, meters and system malfunctions. In addition, erroneous communications and adversarial compromises of the meters can also introduce gross errors to the measurement quantities received by the central control center. In these scenarios, the observed measurements contain significant or even arbitrary measurement errors, in addition to the more common additive observation noises. So state estimation in power networks needs to detect, identify, and eliminate these large measurement errors [5], [6], [25]. While there are a series of works for dealing with outliers in *linear* measurements [7], [8], the measurements for state estimation in power networks are *nonlinear* functions of the states. This motivates us to study the general problem of state estimation from nonlinear measurements in the presence of bad data.

Suppose that we make n measurements to estimate the state \mathbf{x} described by an m -dimensional ($m < n$) real-numbered vector, then these measurements can be written as an n -dimensional vector \mathbf{y} , which is related to the state vector through the measurement equation

$$\mathbf{y} = h(\mathbf{x}) + \mathbf{v} + \mathbf{e}, \quad (1.1)$$

where $h(\mathbf{x})$ is a set of n general functions, which may be linear or nonlinear, and \mathbf{v} is an n -dimensional vector of additive measurement noise, and \mathbf{e} is the vector of bad data imposed on the measurements. For simplicity, we assume that \mathbf{v} is an n -dimensional vector with i.i.d. zero mean Gaussian elements of variance σ^2 (even though the performance bounds carry over to other types of additive noises). We also assume that \mathbf{e} is an n -dimensional *sparse* vector with at most k nonzero entries, and the nonzero entries can take arbitrary real-numbered values. The sparsity k of gross errors reflects the nature of bad data because generally only a few faulty sensing results are present or an adversary party may control only a few malicious meters. It is natural that a small number of sensors and meters are faulty at a certain moment; an adversary party may be only able to alter the results of a limited number of meters under his control [22]; and communication errors of meter results are often rare.

When there are no bad data present, it is well known that the Least Square (LS) method can be used to suppress the effect of observation noise on state estimations. In nonlinear setting, we need the nonlinear LS method, where we try to find a vector \mathbf{x} minimizing

$$\|\mathbf{y} - h(\mathbf{x})\|_2. \quad (1.2)$$

Manuscript received June 14, 2013; accepted September 10, 2013. Date of publication September 18, 2013; date of current version November 13, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Trac D. Tran. This work was supported by the NSF under CCF-0835706, ONR under N00014-11-1-0131, and DTRA under HDTRA1-11-1-0030. Part of this paper was presented at the Fiftieth IEEE Conference on Decision and Control 2011.

W. Xu is with the Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242 USA (e-mail: weiyu-xu@uiowa.edu).

M. Wang is with the Rensselaer Polytechnic Institute, Troy, NY 12181 USA (e-mail: wangm7@rpi.edu).

J.-F. Cai is with the Department of Mathematics, University of Iowa, Iowa City, IA 52242 USA (e-mail: jianfeng-cai@uiowa.edu).

A. Tang is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: atang@ece.cornell.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2013.2282463

However, the LS method generally only works well when there are no bad data \mathbf{e} corrupting the observation \mathbf{y} . If the magnitudes of bad data are large, the estimation result can be very far from the true state [7], [8]. So further techniques to eliminate abnormal measurements are needed when there are bad data present in the measurement results.

Bad data detection in power networks can be viewed as a sparse error detection problem, which shares similar mathematical structures as sparse recovery problems in compressed sensing [7], [8], [12]. However, state estimation problem in power networks is different from linear sparse error detection problem [8]. In fact, $h(\mathbf{x})$ in (I.1) is a nonlinear mapping instead of a linear mapping as in [7]. It is the goal of this paper to provide sparse error correction algorithms, and performance analysis for general sparse error correction based on nonlinear measurements, with applications in bad data detection for electrical power networks.

Toward this end, we first consider the simplified problem when $h(\mathbf{x})$ is linear, which serves as a basis for solving and analyzing sparse recovery problems with nonlinear measurements. For this sparse recovery problem with linear measurements, a mixed least ℓ_1 norm and ℓ_2 norm convex program is used to simultaneously detect bad data and subtract additive noise from the observations. In our theoretical analysis of the decoding performance, we assume $h(\mathbf{x})$ is a linear transformation $H\mathbf{x}$, where H is an $n \times m$ matrix with i.i.d. standard zero mean Gaussian entries. Through using the almost Euclidean property for the linear subspace generated by H , we derive a new performance bound for the state estimation error under sparse bad data and additive observation noise. In our analysis, using the “escape-through-the-mesh” theorem from geometric functional analysis [15], we are able to significantly improve on the bounds for the almost Euclidean property of a linear subspace, which may be interesting in a more general mathematical setting. Compared with earlier analysis on the same optimization problem in [7], we are able to give explicit bounds on the error performance, which is generally sharper than the result in [7] in terms of recoverable sparsity.

We then consider the nonlinear measurement setting. Generalizing the algorithm and results for linear measurements, we propose an iterative convex programming approach to perform joint noise reduction and bad data detection from nonlinear measurements. We establish conditions under which the iterative algorithm converges to the true state in the presence of bad data even when the measurements are nonlinear. We are also able to explicitly verify when the conditions hold through a semidefinite programming formulation. Our iterative convex programming based algorithm is shown to work well in this nonlinear setting by numerical examples.

Compared with [21], which originally proposed to apply ℓ_1 minimization in bad data detection in power networks, our approach offers a better decoding error performance when both bad data and additive observation noises are present. [19], [20] considered state estimations under malicious data attacks, and formulated state estimation under malicious attacks as a hypothesis testing problem by assuming a prior probability distribution on the state \mathbf{x} . In contrast, our approach does not rely on any prior information on the signal \mathbf{x} itself, and the performance bounds hold for an arbitrary state \mathbf{x} . Compressive sensing with

nonlinear measurements were studied in [4] by extending the restricted isometry condition. Our sparse recovery problem is different from the compressive sensing problem considered in [4] since our measurements are overcomplete and are designed to perform sparse error corrections instead of compressive sensing. Our analysis does not rely on extensions of the restricted isometry condition.

The rest of this paper is organized as follows. In Section II, we study joint bad data detection and denoising for linear measurements, and derive the performance bound on the decoding error based on the almost Euclidean property of linear subspaces. In Section III, a sharp bound on the almost Euclidean property is given through the “escape-through-the-mesh” theorem. In Section IV, we present explicitly computed bounds on the estimation error for linear measurements. In Section V, we propose our iterative convex programming algorithm to perform sparse recovery from nonlinear measurements and give theoretical analysis on the performance guarantee of the iterative algorithm. In Section VI, we present simulation results of our iterative algorithm to show its performance in power networks.

II. BAD DATA DETECTION FOR LINEAR SYSTEMS

In this section, we introduce a convex program formulation to do bad data detection in linear systems, and characterize its decoding error performance. In a linear system, the corresponding $n \times 1$ observation vector in (I.1) is $\mathbf{y} = H\mathbf{x} + \mathbf{e} + \mathbf{v}$, where \mathbf{x} is an $m \times 1$ signal vector ($m < n$), H is an $n \times m$ matrix, \mathbf{e} is a sparse error vector with at most k nonzero elements, and \mathbf{v} is a noise vector with $\|\mathbf{v}\|_2 \leq \epsilon$. In what follows, we denote the part of any vector \mathbf{w} over any index set K as \mathbf{w}_K .

We solve the following optimization problem involving optimization variables \mathbf{x} and \mathbf{z} , and we then estimate the state \mathbf{x} to be $\hat{\mathbf{x}}$, which is the optimizer value for \mathbf{x} .

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{z}} \quad & \|\mathbf{y} - H\mathbf{x} - \mathbf{z}\|_1, \\ \text{subject to} \quad & \|\mathbf{z}\|_2 \leq \epsilon. \end{aligned} \quad (\text{II.1})$$

This optimization problem was proposed in a slightly different form in [7] by restricting \mathbf{z} in the null space of H^T . We are now ready to give a theorem which bounds the decoding error performance of (II.1), using the almost Euclidean property [9], [18].

Definition 2.1 (Almost Euclidean Property): A subspace in \mathbb{R}^n satisfies the *almost Euclidean* property for a constant $\alpha \leq 1$, if

$$\alpha\sqrt{n}\|\mathbf{w}\|_2 \leq \|\mathbf{w}\|_1$$

holds true for every \mathbf{w} in the subspace.

Theorem 2.2: Let \mathbf{y} , H , \mathbf{x} , \mathbf{e} and \mathbf{v} be specified as above. Suppose that the minimum nonzero singular value of H is σ_{\min} . Let C be a real number larger than 1, and suppose that every vector \mathbf{w} in range of the matrix H satisfies $C\|\mathbf{w}_K\|_1 \leq \|\mathbf{w}_{\bar{K}}\|_1$ for any subset $K \subseteq \{1, 2, \dots, n\}$ with cardinality $|K| \leq k$, where k is an integer, and $\bar{K} = \{1, 2, \dots, n\} \setminus K$. We also assume the subspace generated by H satisfies the *almost Euclidean* property for a constant $\alpha \leq 1$.

Then the solution $\hat{\mathbf{x}}$ to (II.1) satisfies

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq \frac{2(C+1)}{\sigma_{\min}\alpha(C-1)}\epsilon. \quad (\text{II.2})$$

Proof: Suppose that one optimal solution pair to (II.1) is $(\hat{\mathbf{x}}, \hat{\mathbf{z}})$. Since $\|\hat{\mathbf{z}}\|_2 \leq \epsilon$, we have $\|\hat{\mathbf{z}}\|_1 \leq \sqrt{n}\|\hat{\mathbf{z}}\|_2 \leq \sqrt{n}\epsilon$.

Since \mathbf{x} and $\mathbf{z} = \mathbf{v}$ are feasible for (II.1) and $\mathbf{y} = H\mathbf{x} + \mathbf{e} + \mathbf{v}$, then

$$\begin{aligned} & \|\mathbf{y} - H\hat{\mathbf{x}} - \hat{\mathbf{z}}\|_1 \\ &= \|H(\mathbf{x} - \hat{\mathbf{x}}) + \mathbf{e} + \mathbf{v} - \hat{\mathbf{z}}\|_1 \\ &\leq \|H(\mathbf{x} - \mathbf{x}) + \mathbf{e} + \mathbf{v} - \mathbf{v}\|_1 \\ &= \|\mathbf{e}\|_1. \end{aligned}$$

Applying the triangle inequality to $\|H(\mathbf{x} - \hat{\mathbf{x}}) + \mathbf{e} + \mathbf{v} - \hat{\mathbf{z}}\|_1$, we further obtain

$$\|H(\mathbf{x} - \hat{\mathbf{x}}) + \mathbf{e}\|_1 - \|\mathbf{v}\|_1 - \|\hat{\mathbf{z}}\|_1 \leq \|\mathbf{e}\|_1.$$

Denoting $H(\mathbf{x} - \hat{\mathbf{x}})$ as \mathbf{w} , because \mathbf{e} is supported on a set K with cardinality $|K| \leq k$, by the triangle inequality for ℓ_1 norm again,

$$\|\mathbf{e}\|_1 - \|\mathbf{w}_K\|_1 + \|\mathbf{w}_{\bar{K}}\|_1 - \|\mathbf{v}\|_1 - \|\hat{\mathbf{z}}\|_1 \leq \|\mathbf{e}\|_1.$$

So we have

$$-\|\mathbf{w}_K\|_1 + \|\mathbf{w}_{\bar{K}}\|_1 \leq \|\hat{\mathbf{z}}\|_1 + \|\mathbf{v}\|_1 \leq 2\sqrt{n}\epsilon \quad (\text{II.3})$$

With $C\|\mathbf{w}_K\|_1 \leq \|\mathbf{w}_{\bar{K}}\|_1$, we know

$$\frac{C-1}{C+1}\|\mathbf{w}\|_1 \leq -\|\mathbf{w}_K\|_1 + \|\mathbf{w}_{\bar{K}}\|_1.$$

Combining this with (II.3), we obtain

$$\frac{C-1}{C+1}\|\mathbf{w}\|_1 \leq 2\sqrt{n}\epsilon.$$

By the almost Euclidean property $\alpha\sqrt{n}\|\mathbf{w}\|_2 \leq \|\mathbf{w}\|_1$, it follows:

$$\|\mathbf{w}\|_2 \leq \frac{2(C+1)}{\alpha(C-1)}\epsilon. \quad (\text{II.4})$$

By the definition of singular values,

$$\sigma_{\min}\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq \|H(\mathbf{x} - \hat{\mathbf{x}})\|_2 = \|\mathbf{w}\|_2, \quad (\text{II.5})$$

so combining (II.4), we get

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq \frac{2(C+1)}{\sigma_{\min}\alpha(C-1)}\epsilon.$$

Note that when there are no sparse errors present, the decoding error bound using the standard LS method satisfies $\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq \frac{1}{\sigma_{\min}}\epsilon$ [7]. Theorem 2.2 shows that the decoding error bound of (II.1) is oblivious to the amplitudes of these bad data. This phenomenon was also observed in [7] by using the restricted isometry condition for compressive sensing.

We remark that, for given \mathbf{y} and ϵ , by strong Lagrange duality theory, the solution $\hat{\mathbf{x}}$ to (II.1) corresponds to the solution to \mathbf{x} in the following problem (II.6) for some Lagrange dual variable $\lambda \geq 0$.

$$\min_{\mathbf{x}, \mathbf{z}} \quad \|\mathbf{y} - H\mathbf{x} - \mathbf{z}\|_1 + \lambda\|\mathbf{z}\|_2. \quad (\text{II.6})$$

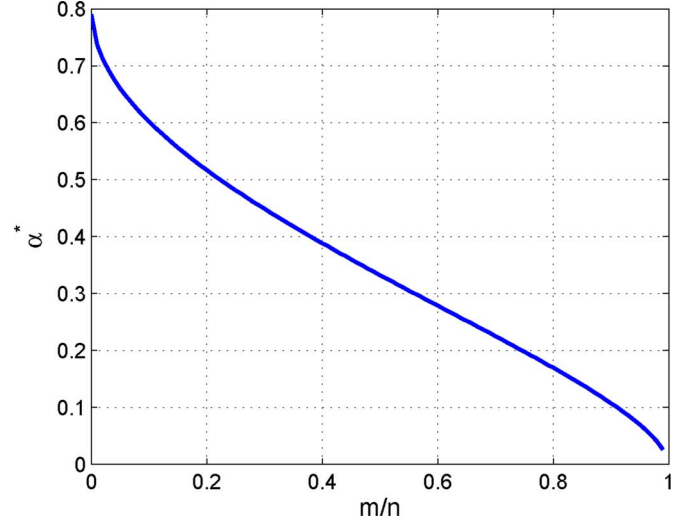


Fig. 1. α^* over m/n .

In fact, when $\lambda \rightarrow \infty$, the optimizer $\|\mathbf{z}\|_2 \rightarrow 0$, and (II.6) approaches

$$\min_{\mathbf{x}} \quad \|\mathbf{y} - H\mathbf{x}\|_1,$$

and when $\lambda \rightarrow 0$, the optimizer $\mathbf{z} \rightarrow \mathbf{y} - H\mathbf{x}$, and (II.6) approaches

$$\min_{\mathbf{x}} \quad \|\mathbf{y} - H\mathbf{x}\|_2.$$

In the next two sections, we aim at explicitly computing $\frac{2(C+1)}{\sigma_{\min}\alpha(C-1)} \times \sqrt{n}$ appearing in the error bound (II.2), which is subsequently denoted as ϖ in this paper. The appearance of the \sqrt{n} factor is to compensate for the energy scaling of large random matrices and its meaning will be clear in later context. We first compute explicitly the almost Euclidean property constant α , and then use the almost Euclidean property to get a direct estimate of the constant C in the error bound (II.2).

III. BOUNDING THE ALMOST EUCLIDEAN PROPERTY

In this section, we would like to give a quantitative bound on the almost Euclidean property constant α such that with high probability (with respect to the measure for the subspace generated by random H), $\alpha\sqrt{n}\|\mathbf{w}\|_2 \leq \|\mathbf{w}\|_1$ holds for every vector \mathbf{w} from the subspace generated by H . Here we assume that each element of H is independently generated from the standard Gaussian distribution $N(0, 1)$. Hence the subspace generated by H is a uniformly distributed m -dimensional subspace.

To ensure that the subspace generated from H satisfies the almost Euclidean property with $\alpha > 0$, we must have the event that the subspace generated by H does not intersect the set $\{\mathbf{w} \in \mathbb{S}^{n-1} \mid \|\mathbf{w}\|_1 < \alpha\sqrt{n}\|\mathbf{w}\|_2 = \alpha\sqrt{n}\}$, where \mathbb{S}^{n-1} is the unit Euclidean sphere in R^n . To evaluate the probability that this event happens, we will need the following “escape-through-mesh” theorem.

Theorem 3.1 (Escape Through the Mesh [15]): Let S be a subset of the unit Euclidean sphere \mathbb{S}^{n-1} in R^n . Let Y be a random m -dimensional subspace of R^n , distributed uniformly in the Grassmanian with respect to the Haar measure. Let us further take $w(S) = E(\sup_{\mathbf{w} \in S} (\mathbf{h}^T \mathbf{w}))$, where \mathbf{h} is a random

column vector in R^n with i.i.d. $N(0, 1)$ components. Assume that $w(S) < (\sqrt{n-m} - \frac{1}{2\sqrt{n-m}})$. Then

$$P(Y \cap S = \emptyset) > 1 - 3.5e^{-\frac{(\sqrt{n-m} - \frac{1}{2\sqrt{n-m}}) - w(S)}{18}}.$$

We derive the following upper bound of $w(\mathbf{h}, S) = \sup_{\mathbf{w} \in S} (\mathbf{h}^T \mathbf{w})$ for an arbitrary but fixed \mathbf{h} . Because the set $\{\mathbf{w} \in S^{n-1} \mid \|\mathbf{w}\|_1 < \alpha\sqrt{n}\|\mathbf{w}\|_2\}$ is symmetric, without loss of generality, we assume that the elements of \mathbf{h} follow i.i.d. half-normal distributions, namely the distribution for the absolute value of a standard zero mean Gaussian random variables. With h_i denoting the i -th element of \mathbf{h} , $\sup_{\mathbf{w} \in S} (\mathbf{h}^T \mathbf{w})$ is equivalent to

$$\begin{aligned} \max \quad & \sum_{i=1}^n h_i w_i \\ \text{subject to} \quad & w_i \geq 0, 1 \leq i \leq n \\ & \sum_{i=1}^n w_i \leq \alpha\sqrt{n} \\ & \sum_{i=1}^n w_i^2 = 1. \end{aligned}$$

Following the method from [29], we use the Lagrange duality to find an upper bound for the objective function of (III.1):

$$\begin{aligned} \min_{u_1 \geq 0, u_2 \geq 0, \lambda \geq 0} \quad & \max_{\mathbf{w}} \mathbf{h}^T \mathbf{w} - u_1 \left(\sum_{i=1}^n w_i^2 - 1 \right) \\ & - u_2 \left(\sum_{i=1}^n w_i - \alpha\sqrt{n} \right) + \sum_{i=1}^n \lambda_i w_i, \end{aligned}$$

where λ is a vector $(\lambda_1, \lambda_2, \dots, \lambda_n)$. Note that restricting u_1 to be nonnegative still gives an upper bound even though it corresponds to an equality.

First, we maximize (III.1) over $w_i, i = 1, 2, \dots, n$ for fixed u_1, u_2 and λ . By setting the derivatives to be zero, the maximizing w_i is given by

$$w_i = \frac{h_i + \lambda_i - u_2}{2u_1}, 1 \leq i \leq n$$

Plugging this back into the objective function in (III.1), we get

$$\begin{aligned} \mathbf{h}^T \mathbf{w} - u_1 \left(\sum_{i=1}^n w_i^2 - 1 \right) \\ - u_2 \left(\sum_{i=1}^n w_i - \alpha\sqrt{n} \right) + \sum_{i=1}^n \lambda_i w_i \\ = \frac{\sum_{i=1}^n (-u_2 + \lambda_i + h_i)^2}{4u_1} + u_1 + \alpha\sqrt{n}u_2. \end{aligned} \quad (\text{III.2})$$

Next, we minimize (III.2) over $u_1 \geq 0$. It is not hard to see the minimizing u_1^* is

$$u_1^* = \frac{\sqrt{\sum_{i=1}^n (-u_2 + \lambda_i + h_i)^2}}{2},$$

and the corresponding minimized value is

$$\sqrt{\sum_{i=1}^n (-u_2 + \lambda_i + h_i)^2} + \alpha\sqrt{n}u_2. \quad (\text{III.3})$$

Then, we minimize (III.3) over $\lambda \geq 0$. Given \mathbf{h} and $u_2 \geq 0$, it is easy to see that the minimizing λ is

$$\lambda_i = \begin{cases} u_2 - h_i & \text{if } h_i \leq u_2; \\ 0 & \text{otherwise,} \end{cases}$$

and the corresponding minimized value is

$$\sqrt{\sum_{1 \leq i \leq n: h_i > u_2} (u_2 - h_i)^2} + \alpha\sqrt{n}u_2. \quad (\text{III.4})$$

Now if we take any $u_2 \geq 0$, (III.4) serves as an upper bound for (III.1), and thus also an upper bound for $\sup_{\mathbf{w} \in S} (\mathbf{h}^T \mathbf{w})$. Since $\sqrt{\cdot}$ is a concave function, by Jensen's inequality, we have for any given $u_2 \geq 0$,

$$E(\sup_{\mathbf{w} \in S} (\mathbf{h}^T \mathbf{w})) \leq \sqrt{E\left\{ \sum_{1 \leq i \leq n: h_i > u_2} (u_2 - h_i)^2 \right\}} + \alpha\sqrt{n}u_2. \quad (\text{III.5})$$

Since \mathbf{h} has i.i.d. half-normal components, the righthand side of (III.5) equals to

$$(\sqrt{(u_2^2 + 1)\text{erfc}(u_2/\sqrt{2})} - \sqrt{2/\pi}u_2e^{-u_2^2/2} + \alpha u_2)\sqrt{n}, \quad (\text{III.6})$$

where erfc is the complementary error function.

One can check that (III.6) is convex in u_2 . Given α , we minimize (III.6) over $u_2 \geq 0$ and let $g(\alpha)\sqrt{n}$ denote the minimum value. Then from (III.5) and (III.6) we know

$$w(S) = E(\sup_{\mathbf{w} \in S} (\mathbf{h}^T \mathbf{w})) \leq g(\alpha)\sqrt{n}. \quad (\text{III.7})$$

Given $\delta = \frac{m}{n}$, we pick the largest α^* such that $g(\alpha^*) < \sqrt{1-\delta}$. Then as n goes to infinity, it holds that

$$w(S) \leq g(\alpha^*)\sqrt{n} < (\sqrt{n-m} - \frac{1}{2\sqrt{n-m}}). \quad (\text{III.8})$$

Then from Theorem 3.1, with high probability $\|\mathbf{w}\|_1 \geq \alpha^*\sqrt{n}\|\mathbf{w}\|_2$ holds for every vector \mathbf{w} in the subspace generated by H . We numerically calculate how α^* changes over δ and plot the curve in Fig. 1. For example, when $\delta = 0.5$, $\alpha^* = 0.332$, thus $\|\mathbf{w}\|_1 \geq 0.332\sqrt{n}\|\mathbf{w}\|_2$ for all \mathbf{w} in the subspace generated by H .

Note that when $\frac{m}{n} = \frac{1}{2}$, we get $\alpha = 0.332$. That is much larger than the known α used in [33], which is approximately 0.07 (see (12) in [33]). When applied to the sparse recovery problem considered in [33], we are able to recover any vector with no more than $0.0289n = 0.0578m$ nonzero elements, which are 20 times more than the $\frac{1}{384}m$ bound in [33].

IV. EVALUATING THE ROBUST ERROR CORRECTION BOUND AND COMPARISONS WITH OTHER BOUNDS

If the elements in the measurement matrix H are i.i.d. drawn from the Gaussian distribution $N(0, 1)$, following upon the

work of Marchenko and Pastur [23], Geman [14] and Silverstein [28] proved that for $m/n = \delta$, as $n \rightarrow \infty$, the smallest nonzero singular value

$$\frac{1}{\sqrt{n}}\sigma_{\min} \rightarrow 1 - \sqrt{\delta}$$

almost surely as $n \rightarrow \infty$.

Now that we have already explicitly bounded α and σ_{\min} , we now proceed to characterize C . It turns out that our earlier result on the almost Euclidean property can be used to compute C .

Lemma 4.1: Suppose an n -dimensional vector \mathbf{w} satisfies $\|\mathbf{w}\|_1 \geq \alpha\sqrt{n}\|\mathbf{w}\|_2$, and for some set $K \subseteq \{1, 2, \dots, n\}$ with cardinality $|K| = k \leq n$, $\frac{\|\mathbf{w}_K\|_1}{\|\mathbf{w}\|_1} = \beta$. Then β satisfies

$$\frac{\beta^2}{k} + \frac{(1-\beta)^2}{n-k} \leq \frac{1}{\alpha^2 n}.$$

Proof: Without loss of generality, we let $\|\mathbf{w}\|_1 = 1$. Then by the Cauchy-Schwarz inequality,

$$\begin{aligned} \|\mathbf{w}\|_2^2 &= \|\mathbf{w}_K\|_2^2 + \|\mathbf{w}_{\bar{K}}\|_2^2 \\ &\geq \left(\frac{\|\mathbf{w}_K\|_1}{\sqrt{k}}\right)^2 + \left(\frac{\|\mathbf{w}_{\bar{K}}\|_1}{\sqrt{n-k}}\right)^2 \\ &= \left(\frac{\beta^2}{k} + \frac{(1-\beta)^2}{n-k}\right) \|\mathbf{w}\|_1^2. \end{aligned}$$

At the same time, by the almost Euclidean property,

$$\alpha^2 n \|\mathbf{w}\|_2^2 \leq \|\mathbf{w}\|_1^2,$$

so we must have

$$\frac{\beta^2}{k} + \frac{(1-\beta)^2}{n-k} \leq \frac{1}{\alpha^2 n}.$$

Corollary 4.2: If a nonzero n -dimensional vector \mathbf{w} satisfies $\|\mathbf{w}\|_1 \geq \alpha\sqrt{n}\|\mathbf{w}\|_2$, and if for any set $K \subseteq \{1, 2, \dots, n\}$ with cardinality $|K| = k \leq n$, $C\|\mathbf{w}_K\|_1 = \|\mathbf{w}_{\bar{K}}\|_1$ for some number $C \geq 1$, then

$$\frac{1}{\frac{k}{n}} + \frac{C^2}{1 - \frac{k}{n}} \leq \frac{(C+1)^2}{\alpha^2}.$$

Proof: If $C\|\mathbf{w}_K\|_1 = \|\mathbf{w}_{\bar{K}}\|_1$, we have

$$\frac{\|\mathbf{w}_K\|_1}{\|\mathbf{w}\|_1} = \frac{1}{C+1}.$$

So by Lemma 4.1, $\beta = \frac{1}{C+1}$ satisfies

$$\frac{\beta^2}{k} + \frac{(1-\beta)^2}{n-k} \leq \frac{1}{\alpha^2 n}.$$

This is equivalent to

$$\frac{1}{\frac{k}{n}} + \frac{C^2}{1 - \frac{k}{n}} \leq \frac{(C+1)^2}{\alpha^2}.$$

Corollary 4.3: Let $\mathbf{y}, \mathbf{x}, H, \mathbf{e}$ and \mathbf{v} be specified as above. Assume that H is drawn i.i.d. from the standard Gaussian distribution $N(0, 1)$ and the subspace generated by H satisfies the *almost Euclidean* property for a constant $\alpha \leq 1$ with overwhelming probability as $n \rightarrow \infty$. Then for any small number

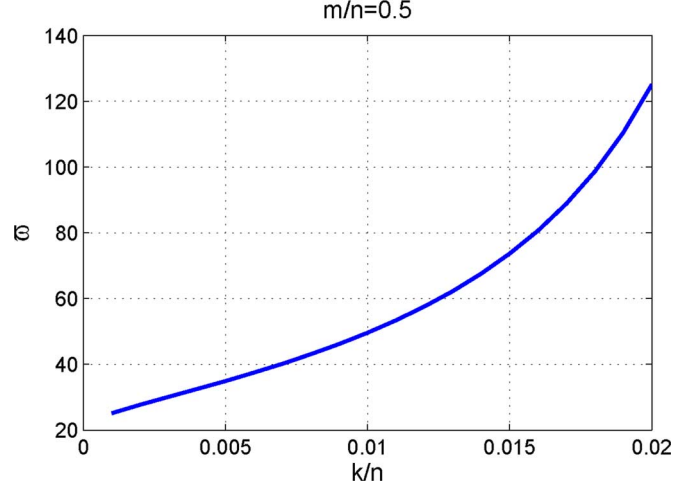


Fig. 2. ϖ versus $\frac{k}{n}$.

$\gamma > 0$, almost surely as $n \rightarrow \infty$, *simultaneously* for any state \mathbf{x} and for any sparse error \mathbf{e} with at most k nonzero elements, the solution $\hat{\mathbf{x}}$ to (II.1) satisfies

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq \frac{2(1+\gamma)(C+1)}{(\sqrt{n} - \sqrt{m})\alpha(C-1)}\epsilon, \quad (\text{IV.1})$$

where C is the smallest nonnegative number such that $\frac{1}{\frac{k}{n}} + \frac{C^2}{1 - \frac{k}{n}} \leq \frac{(C+1)^2}{\alpha^2}$ holds; and $\frac{k}{n}$ needs to be small enough such that $C > 1$.

Proof: This follows from 4.2 and $\frac{1}{\sqrt{n}}\sigma_{\min} \rightarrow 1 - \sqrt{\delta}$ as $n \rightarrow \infty$. ■

So for a sparsity ratio $\frac{k}{n}$, we can use the procedure in Section III to calculate the constant α for the almost Euclidean property. Then we can use α to find the value for C in Corollary 4.3. In Fig. 2, we plot $\frac{2(C+1)}{\sigma_{\min}\alpha(C-1)}\sqrt{n} = \varpi$ appearing in Corollary 4.3 for $\delta = \frac{m}{n} = \frac{1}{2}$ as a function $\frac{k}{n}$. Apparently, when the sparsity $\frac{k}{n}$ increases, the recovery error bound also increases.

A. Comparisons With Existing Performance Bounds

In this subsection, we would like to explore the connection between robust sparse error correction and compressive sensing, and compare our performance bound with the bounds in the compressive sensing literature.

As already noted in [7] and [8], sparse error correction can be seen as a dual to the compressive sensing problem. If we multiply $\mathbf{y} = H\mathbf{x} + \mathbf{e} + \mathbf{v}$ with an $(n-m) \times n$ matrix A satisfying $A \times H = 0$,

$$A\mathbf{y} = AH\mathbf{x} + A\mathbf{e} + A\mathbf{v} = A\mathbf{e} + A\mathbf{v}.$$

Since \mathbf{e} is a sparse vector, this corresponds to a compressive sensing problem with sensing matrix A , observation result $A\mathbf{y}$ and observation noise $A\mathbf{v}$.

[10] and [12] used the theory of high dimensional convex geometry to establish the phase transition on the recoverable sparsity for perfectly sparse signals under noiseless observations. Compared with [10] and [12], our method using the almost Euclidean property applies to the setting of noisy observations. [31] also used the convex geometry tool to get the precise stability phase transition for compressive sensing. However, [31]

mainly focused on obtaining ℓ_1 - ℓ_1 error bounds, namely the recovery error measured in ℓ_1 norm is bounded in terms of the ℓ_1 norm of perturbations. No explicitly computed error bound measured in ℓ_2 norm was given in [31].

In a remarkable paper [11], precise noise-sensitivity phase transition was provided for compressive sensing problems with Gaussian observation noises. In [11], average recovery error in ℓ_2 norm was considered, and the average was taken over the signal input distribution and the Gaussian observation noise. Compared with [11], this paper derives a *worst-case* bound on the recovery error which holds true for *all* k -sparse vectors *simultaneously*; the performance bound also applies to arbitrary observation noises, including but not limited to Gaussian observation noise. Compared with another worst-case recovery error bound obtained in [7] through restricted isometry condition, the bound in this paper greatly improves on the sparsity up to which robust error correction happens. To our best knowledge, currently the best bound on the sparsity for a restricted isometry to hold is still very small. For example, when $\frac{n-m}{n} = \frac{1}{2}$, the proven sparsity $\frac{k}{n}$ for strong $\ell_0 - \ell_1$ equivalence is about 1.5×10^{-3} according to [3], which is smaller than the bound $\frac{k}{n} = 0.0289$ obtained in this paper. In fact, as illustrated in Fig. 1, the almost Euclidean bound exists for any $\frac{m}{n} < 1$. So combined with the recovery error bound in ℓ_1 norm, we can show the recovery error is bounded in ℓ_2 norm up to the stability sparsity threshold in [31]. In this paper, we also obtain much sharper bounds on the almost Euclidean property than known in the literature [33].

V. SPARSE ERROR CORRECTION FROM NONLINEAR MEASUREMENTS

In power network applications, measurement outcomes are nonlinear functions of system states. Let us denote the i -th measurement by $h_i(\mathbf{x})$, where $1 \leq i \leq n$ and $h_i(\mathbf{x})$ can be a nonlinear function of \mathbf{x} . In this section, we study the theoretical performance guarantee of sparse recovery from nonlinear measurements and give an iterative algorithm to do sparse recovery from nonlinear measurements, for which we provide conditions under which the iterative algorithm converges to the true state.

In Subsection V.A, we explore the conditions under which sparse recovery from nonlinear measurements are theoretically possible. In Subsection V.B, we describe our iterative algorithm to perform sparse recovery from nonlinear measurements. In Subsection V.C, we study the algorithm performance guarantees when the measurements are with or without additive noise.

A. Theoretical Guarantee for Direct ℓ_0 and ℓ_1 -Minimization

We first give a general condition which guarantees recovering correctly the state \mathbf{x} from the corrupted observation \mathbf{y} without considering the computational cost.

Theorem 5.1: Let $h(\cdot)$, \mathbf{x} , and \mathbf{e} be specified as above; and $\mathbf{y} = h(\mathbf{x}) + \mathbf{e}$. A state \mathbf{x} can be recovered correctly from any error \mathbf{e} with $\|\mathbf{e}\|_0 \leq k$ from solving the optimization

$$\min_{\mathbf{x}} \|\mathbf{y} - h(\mathbf{x})\|_0, \quad (\text{V.1})$$

if and only if for any $\mathbf{x}^* \neq \mathbf{x}$, $\|h(\mathbf{x}) - h(\mathbf{x}^*)\|_0 \geq 2k + 1$.

Proof: We first prove the sufficiency part, namely if for any $\mathbf{x}^* \neq \mathbf{x}$, $\|h(\mathbf{x}) - h(\mathbf{x}^*)\|_0 \geq 2k + 1$, we can always correctly recover \mathbf{x} from \mathbf{y} corrupted with any error \mathbf{e} with $\|\mathbf{e}\|_0 \leq k$. Sup-

pose that instead a solution to the optimization problem (V.1) is an $\mathbf{x}^* \neq \mathbf{x}$. Then

$$\begin{aligned} \|\mathbf{y} - h(\mathbf{x}^*)\|_0 &= \|(h(\mathbf{x}) + \mathbf{e}) - h(\mathbf{x}^*)\|_0 \\ &\geq \|h(\mathbf{x}) - h(\mathbf{x}^*)\|_0 - \|\mathbf{e}\|_0 \\ &\geq (2k + 1) - k \\ &> \|\mathbf{e}\|_0 = \|\mathbf{y} - h(\mathbf{x})\|_0. \end{aligned}$$

So $\mathbf{x}^* \neq \mathbf{x}$ can not be a solution to (V.1), which is a contradiction.

For the necessary part, suppose that there exists an $\mathbf{x}^* \neq \mathbf{x}$ such that $\|h(\mathbf{x}) - h(\mathbf{x}^*)\|_0 \leq 2k$. Let I be the index set where $h(\mathbf{x})$ and $h(\mathbf{x}^*)$ differ and its size $|I| \leq 2k$. Let $\gamma = h(\mathbf{x}^*) - h(\mathbf{x})$. We pick \mathbf{e} such that $\mathbf{e}_i = \gamma_i, \forall i \in I'$, where $I' \subseteq I$ is an index set with cardinality $|I'| = k$; and \mathbf{e}_i to be 0 otherwise. Then

$$\begin{aligned} \|\mathbf{y} - h(\mathbf{x}^*)\|_0 &= \|h(\mathbf{x}) - h(\mathbf{x}^*) + \mathbf{e}\|_0 \\ &= |I| - k \\ &\leq k = \|\mathbf{e}\|_0 = \|\mathbf{y} - h(\mathbf{x})\|_0, \end{aligned}$$

which means that \mathbf{x} is certainly not a unique solution to (V.1). ■

Theorem 5.2: Let $h(\cdot)$, \mathbf{x} , and \mathbf{e} be specified as above; and $\mathbf{y} = h(\mathbf{x}) + \mathbf{e}$. A state \mathbf{x} can be recovered correctly from any error \mathbf{e} with $\|\mathbf{e}\|_0 \leq k$ from solving the optimization

$$\min_{\mathbf{x}} \|\mathbf{y} - h(\mathbf{x})\|_1, \quad (\text{V.2})$$

if and only if for any $\mathbf{x}^* \neq \mathbf{x}$, $\|(h(\mathbf{x}) - h(\mathbf{x}^*))_K\|_1 < \|(h(\mathbf{x}) - h(\mathbf{x}^*))_{\bar{K}}\|_1$, where K is the support of the error vector \mathbf{e} .

Proof: We first prove if any $\mathbf{x}^* \neq \mathbf{x}$, $\|(h(\mathbf{x}) - h(\mathbf{x}^*))_K\|_1 < \|(h(\mathbf{x}) - h(\mathbf{x}^*))_{\bar{K}}\|_1$, where K is the support of the error vector \mathbf{e} , we can correctly recover state \mathbf{x} from (V.2). Suppose that instead a solution to the optimization problem (V.1) is an $\mathbf{x}^* \neq \mathbf{x}$. Then

$$\begin{aligned} \|\mathbf{y} - h(\mathbf{x}^*)\|_1 &= \|(h(\mathbf{x}) + \mathbf{e}) - h(\mathbf{x}^*)\|_1 \\ &= \|\mathbf{e}_K - (h(\mathbf{x}^*) - h(\mathbf{x}))_K\|_1 + \|(h(\mathbf{x}^*) - h(\mathbf{x}))_{\bar{K}}\|_1 \\ &\geq \|\mathbf{e}_K\|_1 - \|(h(\mathbf{x}^*) - h(\mathbf{x}))_K\|_1 + \|(h(\mathbf{x}^*) - h(\mathbf{x}))_{\bar{K}}\|_1 \\ &> \|\mathbf{e}_K\|_1 = \|\mathbf{y} - h(\mathbf{x})\|_1. \end{aligned}$$

So $\mathbf{x}^* \neq \mathbf{x}$ can not be a solution to (V.2), and this leads to a contradiction.

Now suppose that there exists an $\mathbf{x}^* \neq \mathbf{x}$ such that $\|(h(\mathbf{x}) - h(\mathbf{x}^*))_K\|_1 \geq \|(h(\mathbf{x}) - h(\mathbf{x}^*))_{\bar{K}}\|_1$, where K is the support of the error vector \mathbf{e} . Then we can pick \mathbf{e} to be $(h(\mathbf{x}^*) - h(\mathbf{x}))_K$ over its support K and to be 0 over \bar{K} . Then

$$\begin{aligned} \|\mathbf{y} - h(\mathbf{x}^*)\|_1 &= \|h(\mathbf{x}) - h(\mathbf{x}^*) + \mathbf{e}\|_1 \\ &= \|(h(\mathbf{x}^*) - h(\mathbf{x}))_{\bar{K}}\|_1 \\ &\leq \|(h(\mathbf{x}) - h(\mathbf{x}^*))_K\|_1 = \|\mathbf{e}\|_1 = \|\mathbf{y} - h(\mathbf{x})\|_1, \end{aligned}$$

which means that \mathbf{x} is certainly not a unique solution to (V.2). ■

However, direct ℓ_0 and ℓ_1 minimization may be computationally costly because ℓ_0 norm and nonlinear $h(\cdot)$ may lead to non-convex optimization problems. In the next subsection,

we introduce our computationally efficient iterative sparse recovery algorithm in the general setting when the additive noise \mathbf{v} is present.

B. Iterative ℓ_1 -Minimization Algorithm

Let $h(\cdot)$, \mathbf{x} , \mathbf{e} and \mathbf{v} be specified as above; and $\mathbf{y} = h(\mathbf{x}) + \mathbf{e} + \mathbf{v}$ with $\|\mathbf{v}\|_2 \leq \epsilon$. Now let us consider the algorithm which recovers the state variables iteratively. Ideally, an estimate of the state variables, $\hat{\mathbf{x}}$, can be obtained by solving the following minimization problem,

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{z}} \quad & \|\mathbf{y} - h(\mathbf{x}) - \mathbf{z}\|_1, \\ \text{subject to} \quad & \|\mathbf{z}\|_2 \leq \epsilon. \end{aligned} \quad (\text{V.3})$$

where $\hat{\mathbf{x}}$ is the optimal solution \mathbf{x} . Even though the ℓ_1 norm is a convex function, the function $h(\cdot)$ may make the objective function non-convex.

Since h is nonlinear, we linearize the equations and apply an iterative procedure to obtain a solution. We start with an initial state \mathbf{x}^0 . In the k -th ($k \geq 1$) iteration, let $\Delta\mathbf{y}^k = \mathbf{y} - h(\mathbf{x}^{k-1})$, then we solve the following convex optimization problem,

$$\begin{aligned} \min_{\Delta\mathbf{x}, \mathbf{z}} \quad & \|\Delta\mathbf{y}^k - H^{local} \Delta\mathbf{x} - \mathbf{z}\|_1, \\ \text{subject to} \quad & \|\mathbf{z}\|_2 \leq \epsilon, \end{aligned} \quad (\text{V.4})$$

where H^{local} is the $n \times m$ Jacobian matrix of h evaluated at the point \mathbf{x}^{k-1} . Let $\Delta\mathbf{x}^k$ denote the optimal solution $\Delta\mathbf{x}$ to (V.4), then the state estimation is updated by

$$\mathbf{x}^k = \mathbf{x}^{k-1} + \Delta\mathbf{x}^k. \quad (\text{V.5})$$

We repeat the process until $\Delta\mathbf{x}^k$ approaches 0 close enough or k reaches a specified maximum value.

Note that when there is no additive noise, we can take $\epsilon = 0$ in this iterative algorithm; and the algorithm is exactly the same as the state estimation algorithm from [21].

C. Convergence Conditions for the Iterative Sparse Recovery Algorithm

In this subsection, we discuss the convergence of the proposed algorithm in Subsection V.B. First, we give a necessary condition (Theorem 5.3) for recovering the true state when there is no additive noise, and then give a sufficient condition (Corollary 5.5) for the iterative algorithm to converge to the true state in the absence of additive noise. Secondly, we give the performance bounds (Theorem 5.4) for the iterative sparse error correction algorithm when there is additive noise.

Theorem 5.3 (Necessary Recovery Condition): Let $h(\cdot)$, \mathbf{x} , and \mathbf{e} be specified as above; and $\mathbf{y} = h(\mathbf{x}) + \mathbf{e}$. The iterative algorithm converges to the true state \mathbf{x} only if for the Jacobian matrix H^{local} at the point of \mathbf{x} and for any $\mathbf{x}^* \neq \mathbf{x}$, $\|(H^{local}\mathbf{x}^*)_K\|_1 > \|(H^{local}\mathbf{x}^*)_{\bar{K}}\|_1$, where K is the support of the error vector \mathbf{e} .

Proof: The proof follows from the proof for Theorem 5.2, with the linear function $g(\Delta\mathbf{x}) = h(\mathbf{x}) + H^{local}\Delta\mathbf{x}$, where H^{local} is the Jacobian matrix at the true state \mathbf{x} . ■

Theorem 5.3 shows that for nonlinear measurements, the local Jacobian matrix needs to satisfy the same condition as the matrix for linear measurements. This assumes that the iterative algorithm starts with the correct initial state. However, the

iterative algorithm generally does not start at the true state \mathbf{x} . In the following theorem, we give a sufficient condition for the algorithm to have an upper bounded estimation error when additive noises are present, even though the starting point is not precisely the same one. As a corollary of this theorem, the proposed algorithm converges to the true state when there is no additive noise.

Theorem 5.4 (Guarantee With Additive Noise): Let $h(\cdot)$, \mathbf{x} , \mathbf{e} , and n be specified as above; and $\mathbf{y} = h(\mathbf{x}) + \mathbf{e} + \mathbf{v}$ with $\|\mathbf{v}\|_2 \leq \epsilon$. We assume that each function $h_i(\cdot)$, $1 \leq i \leq n$ has continuous gradients. Suppose that at every point \mathbf{x} , the local $n \times m$ Jacobian matrix H^{local} is full rank and satisfies that for every \mathbf{z} in the range of H^{local} , $C\|\mathbf{z}_K\|_1 \leq \|\mathbf{z}_{\bar{K}}\|_1$, where K is the support of the error vector \mathbf{e} and C a constant larger than 1. Moreover, for a fixed constant $\beta < 1$, we assume that

$$\frac{2(C+1)}{C-1} \frac{\sigma_{\max}^1(H^{true} - H^{local})}{\sigma_{\min}^1(H^{local})} \leq \beta \quad (\text{V.6})$$

holds for any two states \mathbf{x}_1 and \mathbf{x}_2 , where H^{local} is the local Jacobian matrix at the point \mathbf{x}_1 , H^{true} is a matrix such that $h(\mathbf{x}_2) - h(\mathbf{x}_1) = H^{true}(\mathbf{x}_2 - \mathbf{x}_1)$, $\sigma_{\max}^1(A)$ is the induced ℓ_1 matrix norm for A , and $\sigma_{\min}^1(A)$ for a matrix A is defined as $\sigma_{\min}^1(A) = \min\{\|A\mathbf{z}\|_1 : \|\mathbf{z}\|_1 = 1\}$.

Then for any true state \mathbf{x} , the estimation $\mathbf{x}^{k+1} = \mathbf{x}^k + \Delta\mathbf{x}^{k+1}$, where $\Delta\mathbf{x}^{k+1}$ is the solution to the $(k+1)$ -th iteration optimization

$$\begin{aligned} \min_{\Delta\mathbf{x}^{k+1}, \mathbf{z}} \quad & \|\Delta\mathbf{y}^{k+1} - H^{local} \Delta\mathbf{x}^{k+1} - \mathbf{z}\|_1, \\ \text{subject to} \quad & \|\mathbf{z}\|_2 \leq \epsilon \end{aligned} \quad (\text{V.7})$$

satisfies

$$\begin{aligned} \|\mathbf{x} - \mathbf{x}^{k+1}\|_1 &\leq \frac{2(C+1)}{(C-1)\sigma_{\min}^1(H^{local})} \times 2\sqrt{n}\epsilon \\ &+ \frac{2(C+1)}{C-1} \frac{\sigma_{\max}^1(H^{true} - H^{local})}{\sigma_{\min}^1(H^{local})} \|\mathbf{x} - \mathbf{x}^k\|_1, \end{aligned}$$

where H^{local} is the local Jacobian matrix at \mathbf{x}^k , and H^{true} is a matrix satisfying $h(\mathbf{x}) - h(\mathbf{x}^k) = H^{true}(\mathbf{x} - \mathbf{x}^k)$.

As the number of iteration $k \rightarrow \infty$, the estimation error

$$\|\mathbf{x} - \mathbf{x}^k\|_1 \leq \frac{2(C+1)}{(1-\beta)(C-1)\sigma_{\min}^1} \times 2\sqrt{n}\epsilon + \gamma,$$

where $\sigma_{\min}^1 = \min_{\mathbf{x}_1} \sigma_{\min}^1(H^{local}(\mathbf{x}_1))$, $H^{local}(\mathbf{x}_1)$ is the local Jacobian matrix at an arbitrary state \mathbf{x}_1 , and $\gamma > 0$ is an arbitrarily small positive number.

Remarks: When the function is linear and therefore $H^{true} = H^{local}$, the condition (V.6) will always be satisfied with $\beta = 0$. So nonlinearity is captured by the term $\sigma_{\max}^1(H^{true} - H^{local})$. We also remark that H^{true} is not necessarily a Jacobian matrix at a certain point, but however, from Mean Value Theorem for multiple variables, such a matrix exists.

Proof: At the k -th iteration of the iterative state estimation algorithm

$$\mathbf{y} = H^{true} \Delta\mathbf{x}^* + h(\mathbf{x}^k) + \mathbf{e} + \mathbf{v}, \quad (\text{V.8})$$

where H^{true} is an $n \times m$ matrix, \mathbf{x}^k is the state estimate at the k -th step, and $\Delta\mathbf{x}^* = \mathbf{x} - \mathbf{x}^k$, namely the estimation error at the k -th step.

Since at the $(k+1)$ -th step, we are solving the following optimization problem

$$\begin{aligned} \min_{\Delta \mathbf{x}, \mathbf{z}} \quad & \|\Delta \mathbf{y}^{k+1} - H^{local} \Delta \mathbf{x} - \mathbf{z}\|_1, \\ \text{subject to} \quad & \|\mathbf{z}\|_2 \leq \epsilon. \end{aligned} \quad (\text{V.9})$$

Plugging (V.8) into (V.9), we are really solving

$$\begin{aligned} \min_{\Delta \mathbf{x}, \mathbf{z}} \quad & \|H^{true} \Delta \mathbf{x}^* + \mathbf{e} + \mathbf{v} - H^{local} \Delta \mathbf{x} - \mathbf{z}\|_1, \\ \text{subject to} \quad & \|\mathbf{z}\|_2 \leq \epsilon. \end{aligned} \quad (\text{V.10})$$

Denoting $(H^{true} - H^{local})\Delta \mathbf{x}^*$ as \mathbf{w} , which is the measurement gap generated by using the local Jacobian matrix H^{local} instead of H^{true} , then (V.10) is equivalent to

$$\begin{aligned} \min_{\Delta \mathbf{x}, \mathbf{z}} \quad & \|H^{local}(\Delta \mathbf{x}^* - \Delta \mathbf{x}) + \mathbf{w} + \mathbf{e} + \mathbf{v} - \mathbf{z}\|_1, \\ \text{subject to} \quad & \|\mathbf{z}\|_2 \leq \epsilon. \end{aligned} \quad (\text{V.11})$$

Suppose that the solution to (V.7) is $\Delta \mathbf{x} = \Delta \mathbf{x}^* - \text{error}$. We are minimizing the objective ℓ_1 norm, and $(\Delta \mathbf{x}^*, \mathbf{v})$ is a feasible solution with an objective function value $\|\mathbf{w} + \mathbf{e}\|_1$, so we have

$$\|H^{local} \times \text{error} + \mathbf{w} + \mathbf{e} + \mathbf{v} - \mathbf{z}\|_1 \leq \|\mathbf{w} + \mathbf{e}\|_1. \quad (\text{V.12})$$

By the triangular inequality and the property of H^{local} , using the same line of reasoning as in the proof of Theorem 2.2, we have

$$\begin{aligned} \|\mathbf{e}\|_1 + \frac{C-1}{C+1} \|H^{local} \times \text{error}\|_1 - \|\mathbf{w}\|_1 - \|\mathbf{v}\|_1 - \|\mathbf{z}\|_1 \\ \leq \|\mathbf{e}\|_1 + \|\mathbf{w}\|_1. \end{aligned} \quad (\text{V.13})$$

So

$$\|H^{local} \times \text{error}\|_1 \leq \frac{2(C+1)}{C-1} (\|\mathbf{w}\|_1 + \|\mathbf{v}\|_1 + \|\mathbf{z}\|_1). \quad (\text{V.14})$$

Since $\|\mathbf{v}\|_1$ and $\|\mathbf{z}\|_1$ are both no larger than $2\sqrt{n}\epsilon$, $\text{error} = \Delta \mathbf{x}^* - \Delta \mathbf{x}$, $(\mathbf{x} - \mathbf{x}^k) = \Delta \mathbf{x}^*$, and $\mathbf{x} - \mathbf{x}^{k+1} = (\mathbf{x} - \mathbf{x}^k) - (\mathbf{x}^{k+1} - \mathbf{x}^k) = \Delta \mathbf{x}^* - \Delta \mathbf{x}$, we have

$$\begin{aligned} \|\mathbf{x} - \mathbf{x}^{k+1}\|_1 & \leq \frac{2(C+1)}{(C-1)\sigma_{\min}^1(H^{local})} \times 2\sqrt{n}\epsilon \\ & + \frac{2(C+1)}{C-1} \frac{\sigma_{\max}^1(H^{true} - H^{local})}{\sigma_{\min}^1(H^{local})} \|\mathbf{x} - \mathbf{x}^k\|_1, \end{aligned}$$

where $\sigma_{\max}^1(H^{true} - H^{local})$ and $\sigma_{\min}^1(H^{local})$ are respectively the matrix quantities defined in the statement of the theorem.

So as long as $\frac{2(C+1)}{C-1} \frac{\sigma_{\max}^1(H^{true} - H^{local})}{\sigma_{\min}^1(H^{local})} \leq \beta$, for some fixed constant $\beta < 1$, the error upper bound converges to $\frac{2(C+1)}{(1-\beta)(C-1)\sigma_{\min}^1(H^{local})} \times 2\sqrt{n}\epsilon$. ■

When there is no additive noise, as a corollary of Theorem 5.4, we know the algorithm converges to the true state.

Corollary 5.5 (Correct Recovery Without Additive Noise): Let $\mathbf{y}, h(\cdot), \mathbf{x}, H$, and \mathbf{e} be specified as above; and $\mathbf{y} = h(\mathbf{x}) + \mathbf{e}$. Suppose that at every point \mathbf{x} , the local Jacobian matrix H^{local} is full rank and satisfies that for every

\mathbf{z} in the range of H^{local} , $C\|\mathbf{z}_K\|_1 \leq \|\mathbf{z}_K\|_1$, where K is the support of the error vector \mathbf{e} . Moreover, for a fixed constant $\beta < 1$, we assume that

$$\frac{2(C+1)}{C-1} \frac{\sigma_{\max}^1(H^{true} - H^{local})}{\sigma_{\min}^1(H^{local})} \leq \beta, \quad (\text{V.15})$$

holds true for any two states \mathbf{x}_1 and \mathbf{x}_2 , where H^{local} is the local Jacobian matrix at the point \mathbf{x}_1 , H^{true} is a matrix such that $h(\mathbf{x}_2) - h(\mathbf{x}_1) = H^{true}(\mathbf{x}_2 - \mathbf{x}_1)$, $\sigma_{\max}^1(A)$ is the induced ℓ_1 matrix norm for A , and $\sigma_{\min}^1(A)$ for a matrix A is defined as $\sigma_{\min}^1(A) = \min\{\|A\mathbf{z}\|_1 : \|\mathbf{z}\|_1 = 1\}$.

Then any state \mathbf{x} can be recovered correctly from the observation \mathbf{y} from the iterative algorithm in Subsection V.B, regardless of the initial starting state of the algorithm.

D. Verifying the Conditions for Nonlinear Sparse Error Correction

To verify our conditions for sparse error correction from nonlinear measurements, we need to verify:

- at every iteration of the algorithm, for every \mathbf{z} in the range of the local Jacobian matrix H^{local} , $C\|\mathbf{z}_K\|_1 \leq \|\mathbf{z}_K\|_1$, where K is the support of the error vector \mathbf{e} and $C > 1$ is a constant.
- for some constant $\beta < 1$,

$$\frac{2(C+1)}{C-1} \frac{\sigma_{\max}^1(H^{true} - H^{local})}{\sigma_{\min}^1(H^{local})} \leq \beta \quad (\text{V.16})$$

To simplify our analysis, we only focus on verifying these conditions under which the algorithm converges to the true state when there is no observation noise (namely Corollary 5.5), even though similar verifications also apply to the noisy setting (Theorem 5.4).

1) *Verifying the Balancedness Condition:* In nonlinear setting, the first condition, namely the balancedness condition [31], needs to hold *simultaneously* for many (or even an infinite number of) local Jacobian matrices H^{local} , while in the linear setting, the condition only needs to hold for a single linear subspace. The biggest challenge is then to verify that the balancedness condition holds *simultaneously* for a set of linear subspaces. In fact, verifying the balancedness condition for a single linear space is already very difficult [2], [17], [30]. In this subsection, we show how to convert checking this balancedness condition into a semidefinite optimization problem, through which we can explicitly verify the conditions for nonlinear sparse error corrections.

One can verify the balancedness condition for each individual local Jacobian matrix H^{local} in an instance of algorithm execution, however, the results are not general enough: our state estimation algorithm may pass through different trajectories of H^{local} 's, with different algorithm inputs and initializations. Instead, we propose to check the balancedness condition simultaneously for all H^{local} 's in a neighborhood of H_0 , where H_0 is the local Jacobian matrix at the true system state. Once the algorithm gets into this neighborhood in which the balancedness condition holds for every local Jacobian matrix, the algorithm is guaranteed to converge to the true system state.

More specifically, we consider a neighborhood of the true system state, where each $H^{local} = H_0 + \Delta H$. We assume

that for each state in this neighborhood, each row of ΔH is bounded by $\epsilon > 0$ in ℓ_2 norm. The question is whether, for every H^{local} , every vector \mathbf{z} in the range of H^{local} satisfies $C\|\mathbf{z}_K\|_1 \leq \|\mathbf{z}_{\bar{K}}\|_1$, for all subsets K with cardinality $|K| \leq k$ (k is the sparsity of bad data). This requirement is satisfied, if the optimal objective value of the following optimization problem is no more than $\alpha = \frac{1}{1+C}$:

$$\begin{aligned} \max_{K, H} \max_{\mathbf{x}} \quad & \|(H\mathbf{x})_K\|_1, \\ \text{subject to} \quad & \|(H\mathbf{x})_{\bar{K}}\|_1 \leq 1. \end{aligned} \quad (\text{V.17})$$

However, this is a difficult problem to solve, because there are at least $\binom{n}{k}$ subsets K , and the objective is not a concave function. We instead solve a series of n optimization problems to find an upper bound on the optimal objective value of (V.17). For each integer $1 \leq i \leq n$, by taking the set $K = \{i\}$, we solve

$$\begin{aligned} \max_H \max_{\mathbf{x}} \quad & \|(H\mathbf{x})_{\{i\}}\|_1, \\ \text{subject to} \quad & \|(H\mathbf{x})_{\bar{i}}\|_1 \leq 1. \end{aligned} \quad (\text{V.18})$$

Suppose the optimal objective value of (V.18) is given by α_i for $1 \leq i \leq n$. Then $\frac{\alpha_i}{1+\alpha_i}$ is the largest fraction $\|(H\mathbf{x})_{\{i\}}\|_1$ can occupy out of $\|H\mathbf{x}\|_1$. Then $\max_{K, |K|=k} \sum_{i \in K} \frac{\alpha_i}{1+\alpha_i}$ is the largest fraction $\|(H\mathbf{x})_K\|_1$ can occupy out of $\|H\mathbf{x}\|_1$ for any subset K with cardinality $|K| \leq k$. Let us then take a constant C satisfying $\frac{1}{C+1} = \max_{K, |K|=k} \sum_{i \in K} \frac{\alpha_i}{1+\alpha_i}$. Therefore $C\|\mathbf{z}_K\|_1 \leq \|\mathbf{z}_{\bar{K}}\|_1$ always holds true, for all support subsets K with $|K| \leq k$, and all \mathbf{z} 's in the ranges of all $H^{local} = H_0 + \Delta H$ with each row of ΔH bounded by ϵ in ℓ_2 norm.

So now we only need to give an upper bound on the optimal objective value α_i of (V.18). By first optimizing over H , α_i is upper bounded by the optimal objective value of the following optimization problem:

$$\begin{aligned} \max_{\mathbf{x}} \quad & \|(H_0\mathbf{x})_{\{i\}}\|_1 + \epsilon\|\mathbf{x}\|_2, \\ \text{subject to} \quad & \|(H_0\mathbf{x})_{\bar{i}}\|_1 - (n-1)\epsilon\|\mathbf{x}\|_2 \leq 1. \end{aligned} \quad (\text{V.19})$$

Since

$$\|(H_0\mathbf{x})_{\bar{i}}\|_1 \geq \|(H_0\mathbf{x})_{\bar{i}}\|_2 \geq \sigma_{\min}((H_0)_{\bar{i}})\|\mathbf{x}\|_2,$$

(V.19) is equivalent to

$$\begin{aligned} \max_{\mathbf{x}} \quad & \|(H_0\mathbf{x})_{\{i\}}\|_1 + \epsilon\|\mathbf{x}\|_2, \\ \text{subject to} \quad & \|(H_0\mathbf{x})_{\bar{i}}\|_1 - (n-1)\epsilon\|\mathbf{x}\|_2 \leq 1, \\ & \|(H_0)_{\bar{i}}\mathbf{x}\|_2 - (n-1)\epsilon\|\mathbf{x}\|_2 \leq 1. \end{aligned} \quad (\text{V.20})$$

From $\|(H_0)_{\bar{i}}\mathbf{x}\|_2 - (n-1)\epsilon\|\mathbf{x}\|_2 \leq 1$, we know $\|\mathbf{x}\|_2$ in the feasible set of (V.20) is upper bounded by $t' = \frac{1}{\sigma_{\min} - (n-1)\epsilon}$, where σ_{\min} is the smallest singular value of the matrix $(H_0)_{\bar{i}}$. This enables us to further relax (V.20) to

$$\begin{aligned} \max_{\mathbf{x}} \quad & \|(H_0\mathbf{x})_{\{i\}}\|_1 + \epsilon t', \\ \text{subject to} \quad & \|(H_0\mathbf{x})_{\bar{i}}\|_1 - (n-1)\epsilon t' \leq 1, \\ & (\sigma_{\min} - (n-1)\epsilon)\|\mathbf{x}\|_2 \leq 1, \end{aligned} \quad (\text{V.21})$$

which can be solved by semidefinite programming algorithms.

2) *Verifying the Second Condition in Nonlinear Sparse Error Correction:* Now we are in a position to verify the second condition for successful nonlinear sparse recovery:

$$\frac{2(C+1)}{C-1} \frac{\sigma_{\max}^1(H^{true} - H^{local})}{\sigma_{\min}^1(H^{local})} \leq \beta \quad (\text{V.22})$$

holds for a constant $\beta < 1$ in this process.

Suppose that the Jacobian matrix at the true state \mathbf{x}_0 is H_0 . Define a set M_ϵ as the set of system states \mathbf{x} at which each row of $H^{local} - H_0$ is no larger than ϵ in ℓ_2 norm, where H^{local} is the local Jacobian matrix.

By picking a small enough constant $\tau > 0$, we consider a diamond neighborhood $N_\tau = \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}_0\|_1 \leq \tau\}$ of the true state \mathbf{x}_0 , such that $N_\tau \subseteq M_\epsilon$. By Mean Value Theorem for multiple variables, for each $\mathbf{x} \in N_\tau$, we can find an $n \times m$ matrix H^{true} such that $h(\mathbf{x}) - h(\mathbf{x}_0) = H^{true}(\mathbf{x} - \mathbf{x}_0)$, and moreover, each row of $H^{true} - H_0$ is also upper bounded by ϵ in ℓ_2 norm.

For this neighborhood N_τ , using the proposed semidefinite programming (V.21), for a fixed k (the sparsity), we can find a C such that for every local Jacobian H^{local} in that neighborhood, every vector \mathbf{z} in the subspace generated by H^{local} satisfies the balancedness property with the parameter C .

Meanwhile, $\sigma_{\max}^1(H^{true} - H^{local}) \leq n \times 2\epsilon$ in that neighborhood since each row of $H^{true} - H_0$ and $H^{local} - H_0$ is no larger than ϵ in ℓ_2 norm.

Similarly, for every H^{local} in that neighborhood N_τ , by the definition of singular values and inequalities between ℓ_1 and ℓ_2 norms, $\sigma_{\min}^1(H^{local}) \geq \sigma_{\min}^1(H_0) - n\epsilon \geq \frac{1}{\sqrt{n}}\sigma_{\min}(H_0) - n\epsilon$, where $\sigma_{\min}(H_0)$ is the smallest singular value of H_0 .

So as long as

$$\frac{2(C+1)}{C-1} \frac{2n\epsilon}{\frac{1}{\sqrt{n}}\sigma_{\min}(H_0) - n\epsilon} \leq \beta < 1, \quad (\text{V.23})$$

once our state estimation algorithm gets into $N_\tau = \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}_0\|_1 \leq \tau\} \subseteq M_\epsilon$, the algorithm will always stay in the region N_τ (because $\|\mathbf{x}^k - \mathbf{x}_0\|_1$ decreases after each iteration, see the proof of Theorem 5.5 with $\epsilon = 0$), and the iterative state estimation algorithm will converge to the true state \mathbf{x}_0 .

In summary, we can explicitly compute a parameter ϵ such that inside a diamond neighborhood $N_\tau = \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}_0\|_1 \leq \tau\} \subseteq M_\epsilon$, the condition $\frac{2(C+1)}{C-1} \frac{\sigma_{\max}^1(H^{true} - H^{local})}{\sigma_{\min}^1(H^{local})} \leq \beta < 1$ is always satisfied and the decoding algorithm always converges to the true state \mathbf{x} once it gets into the region N_τ . The size of the region N_τ depends on specific functions. For example, if $h(\cdot)$ is a linear function and $C > 1$, N_τ can be taken as R^m . This fits with known results in [7], [8] for linear measurements where the local Jacobian is the same everywhere.

VI. NUMERICAL RESULTS

In our simulation, we apply (II.1) to estimate an unknown vector from Gaussian linear measurements with both sparse errors and noise, and also apply the iterative method to recover state information from nonlinear measurements with bad data and noise in a power system.

Linear System: We first consider recovering a signal vector from linear Gaussian measurements. Let $m = 60$ and $n = 150$. We generate the measurement matrix $H^{n \times m}$ with i.i.d. $N(0, 1)$

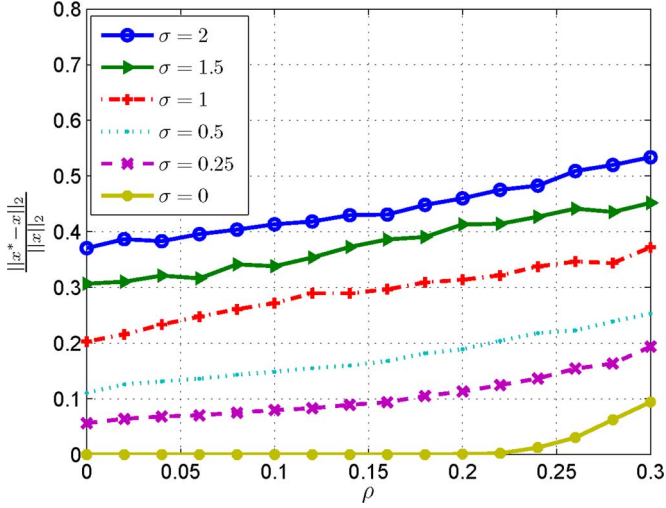


Fig. 3. Estimation error versus ρ for Gaussian measurements with fixed noise level.

entries. We also generate a vector $\mathbf{x} \in R^m$ with i.i.d entries uniformly chosen from interval $[-1, 1]$.

We add to each measurement of $H\mathbf{x}$ with a Gaussian noise independently drawn from $N(0, \sigma^2)$. Let ρ denote the fraction of erroneous measurements. Given ρ , we randomly choose ρn measurements, and each such measurement is added with a Gaussian error independently drawn from $N(0, 4^2)$. We apply (II.1) to estimate \mathbf{x} . Since the noise vector $\mathbf{z} \in R^m$ has i.i.d. $N(0, \sigma^2)$ entries, then $\|\mathbf{z}\|_2/\sigma$ follows the chi distribution with dimension m . Let $F_m(x)$ denote the cumulative distribution function (CDF) of the chi distribution of dimension m , and let $F_m^{-1}(y), y \in [0, 1]$ denote its inverse distribution function. We choose ϵ to be $C\sigma$ where

$$C := F_m^{-1}(0.98). \quad (\text{VI.1})$$

Thus, $\|\mathbf{z}\|_2 \leq \epsilon$ holds with probability 0.98 for randomly generated \mathbf{z} . Let \mathbf{x}^* denote the estimation of \mathbf{x} , and the relative estimation error is represented by $\|\mathbf{x}^* - \mathbf{x}\|_2 / \|\mathbf{x}\|_2$.

We fix the noise level and consider the estimation performance when the number of erroneous measurements changes. Fig. 3 shows how the estimation error changes as ρ increases, and each result is averaged over one hundred runs. We choose different σ between 0 and 2. When $\sigma = 0$, the measurements have errors but no noise, and (II.1) is reduced to conventional ℓ_1 -minimization problem. One can see from Fig. 3 that when there is no noise, \mathbf{x} can be correctly recovered from (II.1) even when twenty percent of measurements contain errors.

We next compare the recovery performance of (II.1) with that of ℓ_1 -minimization. We fix the number of erroneous measurements to be two, i.e., the percentage of erroneous measurements is $\rho = 0.0133$, and increase the noise level σ . One can see from Fig. 4 that (II.1) has a smaller estimation error than that of ℓ_1 -minimization when the noise level is not too small. This is not that surprising since (II.1) takes into account the measurement noise while ℓ_1 -minimization does not.

Power System: We also consider state estimation in power networks. Monitoring the system characteristics is a fundamental prerequisite for the efficient and reliable operation of the power networks. Based on the meter measurements, state estimation provides pertinent information, e.g., the voltage

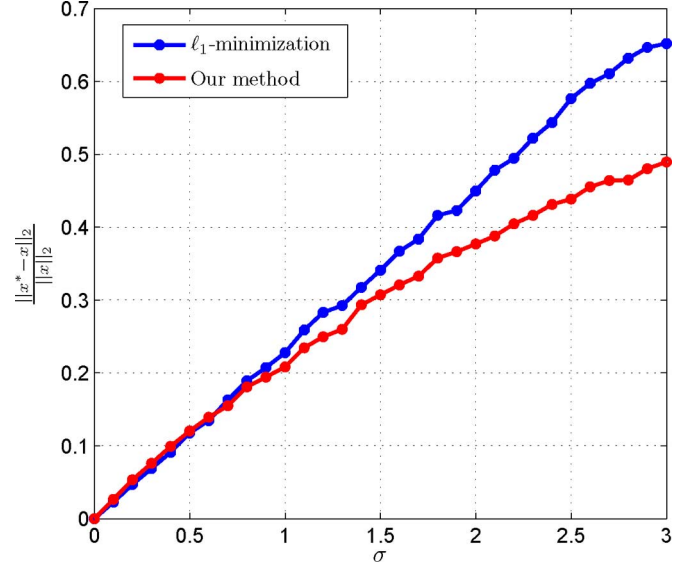


Fig. 4. Comparison of the estimation error with ℓ_1 -minimization.

magnitudes and the voltage angles at each bus, about the operation condition of a power grid [1], [24], [27]. Modern devices like a phasor measurement unit (PMU) [16] can measure the system states directly, but their installation in the current power system is very limited. The current meter measurements are mainly bus voltage magnitudes, real and reactive power injections at each bus, and the real and reactive power flows on the lines. The meter measurements are subject to observation noise, and may contain large errors [5], [13], [25]. Moreover, with the modern information technology introduced in a smart grid, there may exist data attacks from intruders or malicious insiders [19], [20], [22]. Incorrect output of state estimation will mislead the system operator, and can possibly result in catastrophic consequences. Thus, it is important to obtain accurate estimates of the system states from measurements that are noisy and erroneous. If the percentage of the erroneous measurements is small, which is a valid assumption given the massive scale of the power networks, the state estimation is indeed a special case of the sparse error correction problem from nonlinear measurements that we discussed in Section V.

The relationship between the measurements and the state variables for a k' -bus system can be stated as follows [21]:

$$P_i = \sum_{j=1}^{k'} E_i E_j Y_{ij} \cos(\theta_{ij} + \delta_i - \delta_j), \quad (\text{VI.2})$$

$$Q_i = \sum_{j=1}^{k'} E_i E_j Y_{ij} \sin(\theta_{ij} + \delta_i - \delta_j), \quad (\text{VI.3})$$

$$P_{ij} = E_i E_j Y_{ij} \cos(\theta_{ij} + \delta_i - \delta_j) - E_i^2 Y_{ij} \cos \theta_{ij} + E_i^2 Y_{si} \cos \theta_{si} \quad i \neq j, \quad (\text{VI.4})$$

$$Q_{ij} = E_i E_j Y_{ij} \sin(\theta_{ij} + \delta_i - \delta_j) - E_i^2 Y_{ij} \sin \theta_{ij} + E_i^2 Y_{si} \sin \theta_{si} \quad i \neq j, \quad (\text{VI.5})$$

where P_i and Q_i are the real and reactive power injection at bus i respectively, P_{ij} and Q_{ij} are the real and reactive power flow from bus i to bus j , E_i and δ_i are the voltage magnitude and angle at bus i . Y_{ij} and θ_{ij} are the magnitude and phase angle of admittance from bus i to bus j , Y_{si} and θ_{si} are the magnitude

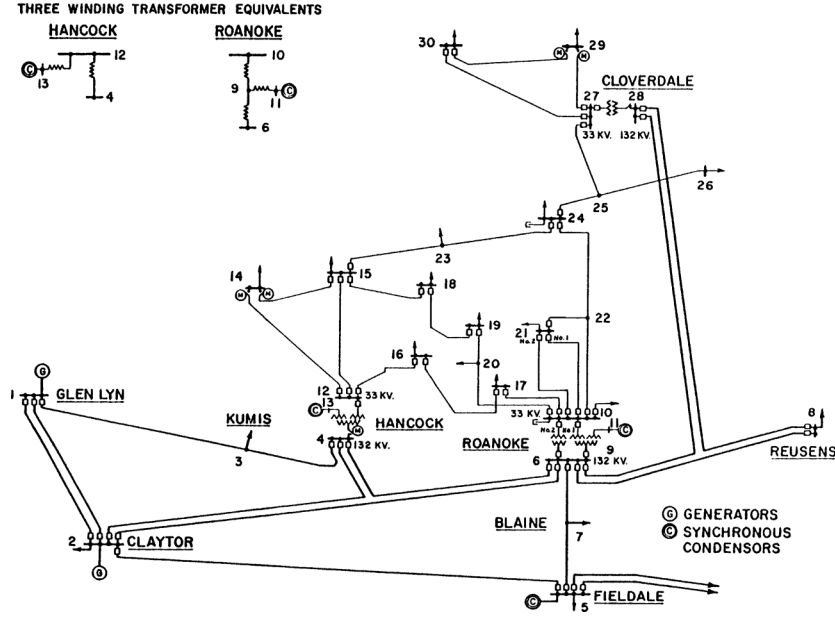


Fig. 5. IEEE 30-bus test system.

and angle of the shunt admittance of line at bus i . Given a power system, all Y_{ij} , θ_{ij} , Y_{si} and θ_{si} are known.

For a k' -bus system, we treat one bus as the reference bus and set the voltage angle at the reference bus to be zero. There are $m = 2k' - 1$ state variables with the first k' variables for the bus voltage magnitudes E_i and the rest $k' - 1$ variables for the bus voltage angles θ_i . Let $\mathbf{x} \in R^m$ denote the state variables and let $\mathbf{y} \in R^n$ denote the n measurements of the real and reactive power injection and power flow. Let $\mathbf{v} \in R^n$ denote the noise and $\mathbf{e} \in R^n$ denote the sparse error vector. Then we can write the equations in a compact form,

$$\mathbf{y} = h(\mathbf{x}) + \mathbf{v} + \mathbf{e}, \quad (\text{VI.6})$$

where $h(\cdot)$ denotes n nonlinear functions defined in (VI.2) to (VI.5). We apply the iterative ℓ_1 -minimization algorithm in Section V.B to recover \mathbf{x} from \mathbf{y} .

We evaluate the performance on the IEEE 30-bus test system. Fig. 5 shows the structure of the test system. Then the state vector \mathbf{x} contains 59 variables. Its first thirty entries correspond to the normalized bus voltage magnitudes, and are all close to 1. Among the thirty buses in this example, the minimum bus voltage magnitude is 0.992, and the maximum bus voltage magnitude is 1.082. The last twenty-nine entries of \mathbf{x} correspond to the relative bus voltage angles to a reference bus. Here, the bus voltage angles are in the range -0.0956 to -0.03131 . We take $n = 100$ measurements including the real and reactive power injection at each bus and some of the real and reactive power flows on the lines. We first characterize the dependence of the estimation performance on the noise level when the number of erroneous measurements is fixed. For each fixed ρ , we randomly choose a set T with cardinality $|T| = \rho n$. Each measurement with its index in T contains a Gaussian error independently drawn from $N(0, 0.5^2)$. Each measurement also contains a Gaussian noise independently drawn from $N(0, \sigma^2)$. For a given noise level σ , we apply the iterative ℓ_1 -minimization algorithm in Section V.B to recover the state vector \mathbf{x} with $\epsilon = C\sigma$ where C is defined in (VI.1).

The initial estimate \mathbf{x}^0 is chosen to have '1's in its first thirty entries and '0's in its last twenty-nine entries. The relative error of the initial estimate is $\|\mathbf{x}^0 - \mathbf{x}\|_2 / \|\mathbf{x}\|_2 = 0.2447$. For example, in one realization when $\rho = 0.02$ and $\sigma = 0$, the iterative method takes seven iterations to converge. Let \mathbf{x}^k be the estimate of \mathbf{x} after the k th iteration, \mathbf{x}^7 is treated as the final estimate \mathbf{x}^* . The relative estimation errors $\|\mathbf{x}^k - \mathbf{x}\|_2 / \|\mathbf{x}\|_2$ are 0.1208, 0.0179, 0.0049, 0.0017, 0.00017, 3×10^{-6} , and 3×10^{-8} respectively. Fig. 6 shows the relative estimation error $\|\mathbf{x}^k - \mathbf{x}\|_2 / \|\mathbf{x}\|_2$ against σ for various ρ . The result is averaged over two hundred runs. Note that when $\sigma = 0$, i.e., the measurements contain sparse errors but no observation noise, the relative estimation error is not zero. For example, $\|\mathbf{x}^* - \mathbf{x}\|_2 / \|\mathbf{x}\|_2$ is 0.018 when ρ is 0.02. That is because the system in Fig. 5 is not proven to satisfy the condition in Theorem 5.5, and the exact recovery with sparse bad measurements is not guaranteed here. However, our next simulation result indicates that our method indeed outperforms some existing method in some cases.

We compare our proposed method with two other recovery methods. In the first method, we assume we know the location of the errors, i.e., the support T of the sparse vector \mathbf{e} . We delete these erroneous measurements and apply the conventional Weighted Least Squares (WLS) method (Section III.A of [5]) to estimate system states based on the remaining measurements. The solution minimizing the weighted least squares function requires the application of an iterative method, and we follow the updating rule in [5] ((6) and (7)). In the second method, we apply the b test in [25] to detect the erroneous measurements. It first applies the WLS method to estimate the system states, then applies a statistical test to each measurement. If some measurement does not pass the test, it is identified as a potential erroneous measurement, and the system states are recomputed by WLS based on the remaining measurements. This procedure is repeated until every remaining measurement passes the statistical test. Fig. 7 shows the recovery performance of three different methods when $\rho = 0.02$. The result is averaged over

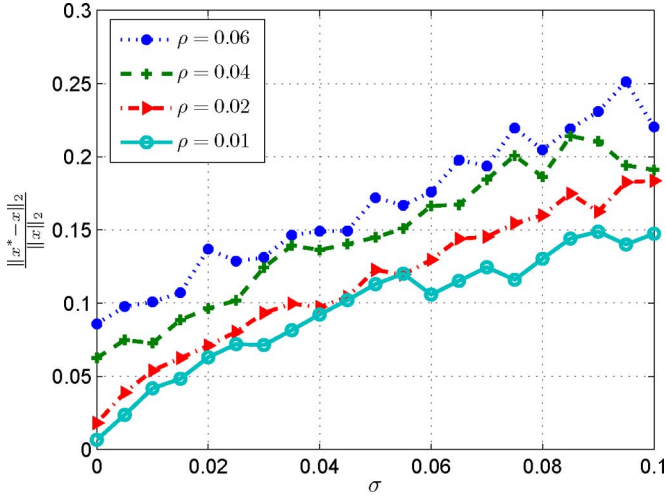


Fig. 6. Estimation error versus σ with fixed percentage of errors in power system.

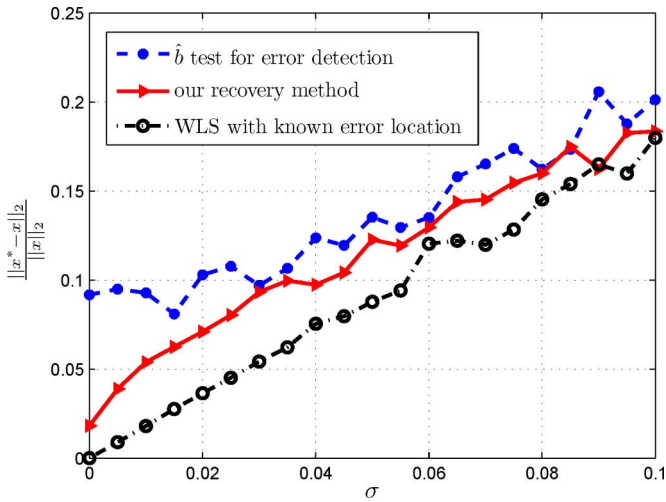


Fig. 7. Estimation error of three recovery methods.

two-hundred runs. The estimation error of our method is less than that of the \hat{b} test. WLS method with known error location has the best recovery performance.

Verification of the Conditions for Nonlinear Sparse Error Correction: We use the verification methods in Subsection V.D to verify the conditions for nonlinear sparse error correction. The Jacobian matrix H_0 for the true state is generated a 200×50 matrix with i.i.d. standard Gaussian entries. We also assume that each row of $\Delta H = H^{local} - H_0$ is upper bounded by $\epsilon = 10^{-3}$ in ℓ_2 norm. We then plot the sparsity k and the corresponding lower bound on $C(k)$ in Fig. 8 such that for every H^{local} , every index set K with cardinality $|K| = k$, and every vector \mathbf{z} in the subspace spanned by H^{local} , we always have $C\|\mathbf{z}_K\|_1 \leq \|\mathbf{z}_{\bar{K}}\|_1$. We can see that for $k = 7$, C is lower bounded by 1.1222. So when there are no more than 7 bad data entries, there exists a convergence region around the true state such that the iterative algorithm converges to the true state once the iterative algorithm gets into that convergence region. Of course, we remark that the computation here is pretty conservative and in practice, we can get much broader convergence scenarios than this theory predicts. However, this is a meaningful step towards establishing sparse recovery conditions for nonlinear measurements.

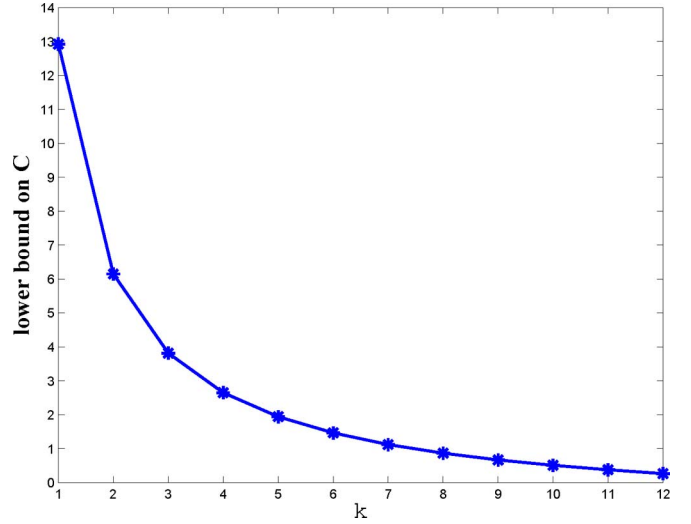


Fig. 8. Lower bound on C when $n = 200$, $m = 50$ and $\epsilon = 10^{-3}$.

REFERENCES

- [1] A. Abur and A. G. Exposito, *Power System State Estimation: Theory and Implementation*, 1st ed. Boca Raton, FL, USA: CRC, 2004.
- [2] A. Aspremont and L. El Ghaoui, "Testing the nullspace property using semidefinite programming," *Math. Program.*, vol. 127, pp. 123–144, 2011.
- [3] J. D. Blanchard, C. Cartis, and J. Tanner, "Compressed sensing: How sharp is the restricted isometry property?," *SIAM Rev.*, vol. 53, no. 1, pp. 105–125, 2011.
- [4] T. Blumensath, "Compressed sensing with nonlinear observations and related nonlinear optimisation problems," 2012 [Online]. Available: <http://arxiv.org/abs/1205.1650>
- [5] A. Bose and K. Clements, "Real-time modeling of power networks," *Proc. IEEE*, vol. 75, no. 12, pp. 1607–1622, 1987.
- [6] Broussolle, "State estimation in power systems: Detecting bad data through the sparse inverse matrix method," *IEEE Trans. Power Appl. Syst.*, vol. PAS-94, pp. 329–337, Mar./Apr. 1975.
- [7] E. Candès and P. Randall, "Highly robust error correction by convex programming," *IEEE Trans. Inf. Theory*, vol. 54, pp. 2829–2840, 2008.
- [8] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–32, Dec. 2005.
- [9] A. Y. Garnaev and E. D. Gluskin, "The widths of a Euclidean ball," *Sov. Math., Dokl.*, vol. 30, pp. 200–204, 1984.
- [10] D. Donoho, "High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension," *Discr. Comput. Geom.*, vol. 35, no. 4, pp. 617–652, 2006.
- [11] D. Donoho, A. Maleki, and A. Montanari, "The noise-sensitivity phase transition in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6920–6941, 2011.
- [12] D. L. Donoho and J. Tanner, "Neighborliness of randomly projected simplices in high dimensions," in *Proc. Nat. Acad. Sci., USA*, 2005, vol. 102, no. 27, pp. 9452–9457.
- [13] E. Handschin, F. C. Schweppe, J. Kohlas, and A. Fiechter, "Bad data analysis for power system state estimation," *IEEE Trans. Power Appl. Syst.*, vol. 94, no. 2, pp. 329–337, Mar. 1975.
- [14] S. Geman, "A limit theorem for the norm of random matrices," *Ann. Probabil.*, vol. 8, no. 2, pp. 252–261, 1980.
- [15] Y. Gordon, "On Milman's inequality and random subspaces with escape through a mesh in R^n ," *Geometric Aspect of Functional Analysis, Isr. Semin. 1986-87, Lect. Notes Math.*, pp. 1317–, 1988.
- [16] J. Jiang, J. Yang, Y. Lin, C. Liu, and J. Ma, "An adaptive PMU based fault detection/location technique for transmission lines. I. Theory and algorithms," *IEEE Trans. Power Del.*, vol. 15, no. 2, pp. 486–493, Apr. 2000.
- [17] A. Juditsky and A. Nemirovski, "On verifiable sufficient conditions for sparse signal recovery via ℓ_1 minimization," *Math. Program.*, vol. 127, pp. 57–88, 2011.
- [18] B. S. Kasin, "The widths of certain finite-dimensional sets and classes of smooth functions," (in Russian) *Izv. Akad. Nauk SSSR Ser. Mat.*, vol. 41, pp. 334–351, 1977.
- [19] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "On malicious data attacks on power system state estimation," in *Proc. UPEC*, Cardiff, Wales, U.K., 2010.

- [20] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on smart grid state estimation: Attack strategies and countermeasures," *Proc. IEEE SmartGridComm*, pp. 220–225, 2010.
- [21] W. Kotiuga and M. Vidyasagar, "Bad data rejection properties of weighted least absolute value techniques applied to static state estimation," *IEEE Trans. Power Appl. Syst.*, vol. PAS-101, pp. 844–853, Apr. 1982.
- [22] Y. Liu, P. Ning, and M. Reiter, "False data injection attacks against state estimation in electric power grids," *Proc. 16th ACM Conf. Comput. Commun. Secur.*, pp. 21–32, 2009.
- [23] V. A. Marčenko and L. A. Pastur, "Distributions of eigenvalues for some sets of random matrices," *Math. USSR-Sbornik*, vol. 1, pp. 457–483, 1967.
- [24] A. Monticelli, "Electric power system state estimation," *Proc. IEEE*, vol. 88, no. 2, pp. 262–282, Feb. 2000.
- [25] A. Monticelli and A. Garcia, "Reliable bad data processing for real time state estimation," *IEEE Trans. Power Appl. Syst.*, vol. PAS-102, pp. 1126–1139, May 1983.
- [26] M. Rudelson and R. Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *Commun. Pure Appl. Math.*, vol. 61, no. 8, 2007.
- [27] F. C. Schweppe and E. J. Handschin, "Static state estimation in electric power systems," *Proc. IEEE*, vol. 62, no. 7, pp. 972–982, Jul. 1974.
- [28] J. Silverstein, "The smallest eigenvalue of a large dimensional Wishart matrix," *Ann. Probabil.*, vol. 13, pp. 1364–1368, 1985.
- [29] M. Stojnic, "Various thresholds for ℓ_1 -optimization in compressed sensing," *ArXiv* [Online]. Available: <http://arxiv.org/abs/0907.3666>
- [30] G. Tang and A. Nehorai, "Verifiable and computable performance analysis of sparsity recovery," *ArXiv* [Online]. Available: <http://arxiv.org/pdf/1102.4868v2.pdf>
- [31] W. Xu and B. Hassibi, "Precise stability phase transitions for ℓ_1 minimization: A unified geometric framework," *IEEE Trans. Inf. Theory*, vol. 57, pp. 6894–6919, 2011.
- [32] W. Xu, M. Wang, and A. Tang, "On state estimation with bad data detection," in *Proc. IEEE Conf. Decision Contr.*, 2011, pp. 5989–5994.
- [33] Y. Zhang, "A simple proof for recoverability of ℓ_1 -minimization: Go over or under? [Online]. Available: <http://www.caam.rice.edu/~yzhang/reports/tr0509.pdf>

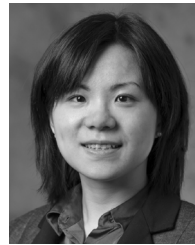


Weiyu Xu (M'11) received the B.E. degree in information engineering from the Beijing University of Posts and Telecommunications in 2002, and the M.S. degree in electronic engineering from Tsinghua University in 2005. He received the M.S. and Ph.D. degrees in electrical engineering in 2006 and 2009, respectively, from the California Institute of Technology (Caltech), with a minor in applied and computational mathematics.

He is currently an Assistant Professor at the Department of Electrical and Computer Engineering,

University of Iowa. His research interests are in signal processing, compressive sensing, communication networks, information and coding theory.

Dr. Xu is a recipient of the Information Science and Technology Fellowship at Caltech, and the recipient of the Charles Wilts doctoral research award in 2010.



Meng Wang (S'06–M'12) received B.E. (Hon.) and M.S. (Hon.) from Tsinghua University, Beijing, China, and Ph.D. from Cornell University, Ithaca, NY, in 2005, 2007, and 2012, all in electrical engineering.

She is currently an Assistant Professor in the Department of Electrical, Computer, and Systems Engineering at Rensselaer Polytechnic Institute. Her research interests include communication networks, signal processing, and nonconvex optimization and its applications.

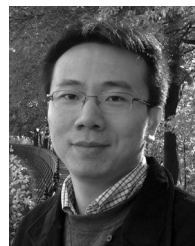
Dr. Wang is a recipient of Jacobs Fellowship of Cornell University in 2008 and 2010.



Jian-Feng Cai received the B.Sc. and M.Sc. degrees in computational mathematics from Fudan University, Shanghai, China, and the Ph.D. degree in mathematics from the Chinese University of Hong Kong, Sha Tin, Hong Kong.

He was a Research Scientist with the National University of Singapore, Singapore, from 2007 to 2009 and a CAM Assistant Adjunct Professor with the University of California at Los Angeles, from 2009 to 2011. He is currently an Assistant Professor with the Department of Mathematics, University of

Iowa, Iowa City. His research interests are in computational harmonic analysis, optimization, numerical linear algebra, and their applications to imaging sciences and related inverse problems.



Ao Tang (S'01–M'07–SM'11) received the B.E. (Hon.) degree in electronics engineering from Tsinghua University, Beijing, China, in 1999, and the Ph.D. degree in electrical engineering with a minor in applied and computational mathematics from the California Institute of Technology (Caltech), Pasadena, in 2006.

He is currently an Associate Professor with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, where he conducts research on the control and optimization of engineering networks including communication networks, power networks and

on-chip networks.

Dr. Tang's recent awards include the Cornell Engineering School Michael Tien'72 Excellence in Teaching Award in 2011, the Young Investigator Award from the Airforce Office of Scientific Research (AFOSR) in 2012, and the Presidential Early Career Award for Scientists and Engineers (PECASE) from the White House in 2012.