Sparse Recovery from Nonlinear Measurements with Applications in Bad Data Detection for Power Networks

Weiyu Xu, Meng Wang, and Ao Tang
School of ECE, Cornell University, Ithaca, NY 14853

Abstract

In this paper, we consider the problem of sparse recovery from nonlinear measurements, which has applications in state estimation and bad data detection for power networks. An iterative mixed ℓ_1 and ℓ_2 convex programming is used to estimate the true state by locally linearizing the nonlinear measurements. 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. 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 also numerically evaluate an iterative convex programming approach to perform bad data detections in nonlinear electrical power networks problems. As a byproduct, in this paper we provide sharp bounds on the almost Euclidean property of a linear subspace, using the "escape-through-a-mesh" theorem from geometric functional analysis.

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 electric power network can be described by the vector of bus voltage magnitudes and angles. However, in smart grid power networks, the measurement of these quantities can be corrupted due to errors in the sensors, communication errors in transmitting the measurement results, and adversarial compromises of the meters. In these settings, the observed measurements contain abnormally large measurement errors,

Part of this paper was presented in the 50th IEEE Conference on Decision and Control 2011 [18].

called bad data, in addition to the usual additive observation noise. So the state estimation of power networks needs to detect, identify, and eliminate these large measurement errors [2], [14], [3]. To make the problem more challenging, the measurements in power networks are generally nonlinear functions of the states. This motivates us to study the general problem of state estimation from nonlinear measurements and bad data.

In general, 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

$$\mathbf{y} = h(\mathbf{x}) + \mathbf{v} + \mathbf{e}, \tag{I.1}$$

where $h(\mathbf{x})$ is a set of n general functions, which may be linear or a nonlinear, and $\mathbf{v}$ is the vector of additive measurement noise, and $\mathbf{e}$ is the vector of bad data imposed on the measurements. In this paper, we assume that $\mathbf{v}$ is an m-dimensional vector with i.i.d. zero mean Gaussian elements of variance σ^2. We also assume that $\mathbf{e}$ is a vector with only k nonzero entries, and the nonzero entries can take arbitrary real-numbered values, reflecting the nature of bad data.

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 the LS method, we try to find a vector $\mathbf{x}$ minimizing

$$\|\mathbf{y} - h(\mathbf{x})\|_2. \tag{I.2}$$

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. So bad data detection to eliminate abnormal measurements is needed when there are bad data present in the measurement results.

Since the probability of large measurement errors occurring is very small, it is reasonable to assume that bad data are only present in a small fraction of the n available meter measurements results. So bad data detection in power networks can be viewed as a sparse error detection problem, which shares similar mathematical structures as sparse recovery problem in compressive sensing [5], [4]. However, this problem in power networks is very different from ordinary sparse error detection problem [5]. In fact, $h(\mathbf{x})$ in (I.1) is a nonlinear mapping instead of a linear mapping as in [4]. It is the goal of this paper to provide a sparse recovery algorithm and performance analysis for sparse recovery from nonlinear measurements with applications in bad data detection for electrical power networks.

We first consider the simplified problem when $h(\mathbf{x})$ is linear, which serves as a basis for solving and analyzing the sparse recovery problem with nonlinear measurements. For this sparse recovery problem with linear measurements, a mixed least ℓ_1 norm and least square convex programming 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-a-mesh" theorem from geometric functional analysis [8], 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 [4], the analysis in this paper is new using the almost Euclidean property rather than the restricted isometry conditions used in [4], and we are able to give explicit bounds on the error performance, which is generally sharper than the result in [4] in terms of recoverable sparsity.

Generalizing the algorithm and results from 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. Our iterative convex programming based algorithm is shown to work well in this nonlinear setting by numerical examples. Compared with [12], which 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 noise are present. [10][11] 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 [1] by extending the restricted isometry condition. Our sparse recovery problem is different from the compressive sensing problem considered in [1] since our measurements are overcomplete and are designed to perform sparse error corrections instead of compressive sensing. Our analysis also 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-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, give theoretical analysis on the performance guarantee of the iterative algorithm, and give an example to illustrate the algorithm and performance bounds. In Section VI, we present simulation results of our iterative algorithm to show its performance in power networks. Section VII concludes this paper.

II. Bad Data Detection for Linear Systems

In this section, we introduce a convex programming 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 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 $\mathbf{x}$.

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

This optimization problem appeared in a slightly different form in [4] by restricting $\mathbf{z}$ in the null space of H. We are now ready to give a theorem which bounds the decoding error performance of (II.1), using the almost Euclidean property [6], [9].

Theorem 2.1: Let $\mathbf{y}$, H, $\mathbf{x}$, $\mathbf{e}$ and $\mathbf{v}$ are 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 $C\|\mathbf{w}_K\|_1 \leq \|\mathbf{w}_{\overline{K}}\|_1$ for any subset $K \subseteq \{1, 2, ..., n\}$ with cardinality $|K| \leq k$, where k is an integer, and $\overline{K} = \{1, 2, ..., n\} \setminus K$. We also assume the subspace generated by H satisfies the *almost Euclidean* property for a constant $\alpha \leq 1$, namely

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

holds true for every $\mathbf{w}$ in the subspace generated by H.

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

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

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,

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

So we have

$$-\|\mathbf{w}_K\|_1 + \|\mathbf{w}_{\overline{K}}\|_1 \leq \|\hat{\mathbf{z}}\|_1 + \|\mathbf{v}\|_1 \leq 2\sqrt{n}\epsilon \tag{II.3}$$

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

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

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:

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

By the definition of singular values,

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

so combining (II.4), we get

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

■

Note that when there are no sparse errors present, the decoding error bound using the standard LS method satisfies $\|\mathbf{x}-\hat{\mathbf{x}}\|_2 \le \frac{1}{\sigma_{\min}}\epsilon$ [4]. Theorem 2.1 shows that the decoding error bound of (II.1) is oblivious to the amplitudes of these bad data. This phenomenon was also observed in [4] 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 $\lambda \ge 0$.

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

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

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

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

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

Thus, (II.6) can be viewed as a weighed version of ℓ_1 minimization and ℓ_2 minimization (or equivalently the LS method). We will later use numerical experiments to show that in order to recover a sparse vector from measurements with both noise and errors, this weighted version outperforms both ℓ_1 minimization and the LS method.

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), $\alpha\sqrt{n}\|\mathbf{w}\|_2 \le \|\mathbf{w}\|_1$ holds for every vector $\mathbf{w}$ from the subspace generated by H. Here we assume that each element of H is generated from the standard Gaussian distribution $N(0,1)$. Hence the subspace generated by H is a uniformly distributed $(n-m)$-dimensional subspaces.

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 $\{\mathbf{w} \in S^{n-1} | \|\mathbf{w}\|_1 < \alpha\sqrt{n}\|\mathbf{w}\|_2\}$, where 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: [8] Let S be a subset of the unit Euclidean sphere 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

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

From Theorem 3.1, we can use the following programming to get an estimate of the upper bound of $w(\mathbf{h}, S) = \sup_{\mathbf{w}\in S}(\mathbf{h}^T\mathbf{w})$. Because the set $\{\mathbf{w} \in S^{n-1} | \|\mathbf{w}\|_1 < \alpha\sqrt{n}\|\mathbf{w}\|_2\}$ 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

$$\max \quad \sum_{i=1}^{n} h_i y_i \tag{III.1}$$

$$\text{subject to} \quad y_0 \geq 0, 1 \leq i \leq n \tag{III.2}$$

$$\sum_{i=1}^{n} y_i \leq \alpha\sqrt{n} \tag{III.3}$$

$$\sum_{i=1}^{n} y_i^2 = 1. \tag{III.4}$$

Following the method from [17], we use the Lagrange duality to find an upper bound for the objective function of (III.1).

$$\min_{u_1\geq 0, u_2\geq 0, \lambda\geq 0} \max_{w} \mathbf{h}^T\mathbf{w} - u_1(\sum_{i=1}^{n} w_i^2 - 1)$$
$$-u_2(\sum_{i=1}^{n} w_i - \alpha\sqrt{n}) + \sum_{i=1}^{n} \lambda_i w_i, \tag{III.5}$$

where λ is a vector $(\lambda_1, \lambda_2, ..., \lambda_n)$.

First, we maximize (III.5) over w_i, $i = 1, 2, ..., n$ for fixed u_1, u_2 and λ. By making the derivatives to be zero, the minimizing w_i is given by

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

Plugging this back to the objective function in (III.5), we get

$$\begin{aligned} & \mathbf{h}^T\mathbf{w} - u_1(\sum_{i=1}^{n} w_i^2 - 1) \\ & -u_2(\sum_{i=1}^{n} w_i - \alpha\sqrt{n}) + \sum_{i=1}^{n} \lambda_i w_i \\ = & \frac{\sum_{i=1}^{n} (-u_2 + \lambda_i + h_i)^2}{4u_1} + u_1 + \alpha\sqrt{n}u_2. \end{aligned} \tag{III.6}$$

Next, we minimize (III.6) over $u_1 \ge 0$. 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

$$\sqrt{\sum_{i=1}^{n} (-u_2 + \lambda_i + h_i)^2} + \alpha\sqrt{n}u_2. \tag{III.7}$$

Then, we minimize (III.7) over $\lambda \ge 0$. Given $\mathbf{h}$ and $u_2 \ge 0$, it is easy to see that the minimizing λ is

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

and the corresponding minimized value is

$$\sqrt{\sum_{1 \le i \le n : h_i < u_2} (u_2 - h_i)^2} + \alpha\sqrt{n}u_2. \tag{III.8}$$

Now if we take any $u_2 \ge 0$, (III.8) serves as an upper bound for (III.5), 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 $u_2 \ge 0$,

$$E(\sup_{\mathbf{w} \in S}(\mathbf{h}^T w)) \le \sqrt{E\{\sum_{1 \le i \le n : h_i < u_2} (u_2 - h_i)^2\}} + \alpha\sqrt{n}u_2. \tag{III.9}$$

Since $\mathbf{h}$ has i.i.d. half-normal components, the righthand side of (III.9) equals to

$$(\sqrt{(u_2^2+1)\text{erfc}(u_2/\sqrt{2}) - \sqrt{2/\pi}u_2 e^{-u_2^2/2}} + \alpha u_2)\sqrt{n}, \tag{III.10}$$

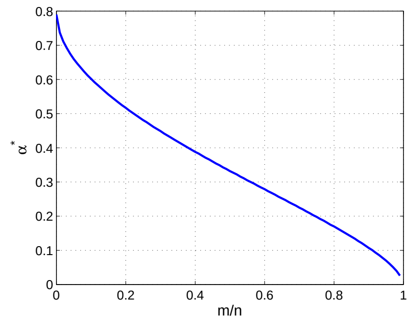

Fig. 1: α^* over m/n

where erfc is the error function.

One can check that (III.10) is convex in u_2. Given α, we minimize (III.10) over $u_2 \geq 0$ and let $g(\alpha)\sqrt{n}$ denote the minimum value. Then from (III.9) and (III.10) we know

$$w(S) = E(\sup_{\mathbf{w}\in S}(\mathbf{h}^T\mathbf{w})) \leq g(\alpha)\sqrt{n}. \tag{III.11}$$

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

$$w(S) \leq g(\alpha^*)\sqrt{n} < (\sqrt{n-m} - \frac{1}{2\sqrt{n-m}}). \tag{III.12}$$

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 [19], which is approximately 0.07 (see Equation (12) in [19]). When applied to the sparse recovery problem considered in [19], we will be 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 [19].

IV. Evaluating the Robust Error Correction Bound

If the elements in the measurement matrix H are i.i.d. as the unit real Gaussian random variables $N(0,1)$, following upon the work of Marchenko and Pastur [13], Geman [7] and Silverstein [16] proved that for $m/n = \delta$, as $n \to \infty$, the smallest nonzero singular value

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

almost surely as $n \to \infty$.

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 $K \subseteq \{1, 2, ..., n\}$ 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,

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

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 $K \subseteq \{1, 2, ..., n\}$ with cardinality $|K| = k \leq n$, $C\|\mathbf{w}_K\|_1 = \|\mathbf{w}_{\overline{K}}\|_1$ for some number $C \geq 1$, then

$$\frac{k}{n} \geq \frac{(B + 1 - C^2) - \sqrt{(B + 1 - C^2)^2 - 4B}}{2B}, \tag{IV.1}$$

where $B = \frac{(C+1)^2}{\alpha^2}$.

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

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

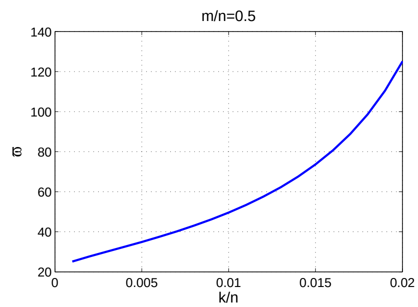

Fig. 2: ϖ versus $\frac{k}{n}$

This is equivalent to

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

Solving this inequality for $\frac{k}{n}$, we get (IV.1). ■

So for a sparsity ratio $\frac{k}{n}$, this corollary can be used to find a lower bound on C satisfying $\frac{\|\mathbf{w}_K\|_1}{\|\mathbf{w}\|_1} = \frac{1}{C+1}$. Combining these results on computing $\sigma_{\min}$, α and C, we can then compute the bound $\frac{2(C+1)}{\sigma_{\min}\alpha(C-1)}\sqrt{n} = \varpi$ in Theorem 2.1. For example, when $\delta = \frac{m}{n} = \frac{1}{2}$, we plot the bound ϖ as a function of $\frac{k}{n}$ in Fig. 2.

V. Sparse Error Correction from Nonlinear Measurements

In applications, measurement outcome can be 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. In Subsection V-D, we give an example to illustrate our algorithm and analysis.

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 $\mathbf{y}$, $h(\cdot)$, $\mathbf{x}$, H, 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

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

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 $\|\mathbf{e}\|_0 \leq k$. Suppose that instead an 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 $\mathbf{x}$ can not be a solution to (V.1) and is certainly not a unique solution to (V.1). ■

Theorem 5.2: Let $\mathbf{y}$, $h(\cdot)$, $\mathbf{x}$, H, and $\mathbf{e}$ be specified as above; and 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

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

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

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

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

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 $\|(h(\mathbf{x}) - h(\mathbf{x}^*))_K\|_1 \geq \|(h(\mathbf{x}) - h(\mathbf{x}^*))_{\overline{K}}\|_1$, 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 $\overline{K}$. Then

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

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

However, direct ℓ_0 and ℓ_1 minimization may be computationally costly because ℓ_0 norma 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 $\mathbf{y}$, $h(\cdot)$, $\mathbf{x}$, H, $\mathbf{e}$ 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,

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

where $\hat{\mathbf{x}}$ is the optimal solution $\mathbf{x}$. Even though 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,

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

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

$$\mathbf{x}^k = \mathbf{x}^{k-1} + \Delta\mathbf{x}^k. \tag{V.5}$$

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. When there is no additive noise, the algorithm is exactly the same as the state estimation algorithm from [12].

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 (Theorem 5.4) for the iterative algorithm to converge to the true state in the absence of additive noise. Secondly, we give the performance bounds (Theorem 5.5) for the iterative sparse recovery algorithm when there is additive noise.

Theorem 5.3 (Necessary Condition for Recovering True State): Let $\mathbf{y}$, $h(\cdot)$, $\mathbf{x}$, H, 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 $\mathbf{x}^* \neq 0$, $\|(H^{local}\mathbf{x}^*)_K\|_1 \leq \|(H^{local}\mathbf{x}^*)_{\overline{K}}\|_1$, 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 the true state $\mathbf{x}$. In the following theorem, we give a sufficient condition for the algorithm to converge to the true state when there is no additive noise.

Theorem 5.4 (Guarantee 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 is full rank and satisfies that for every $\mathbf{z}$ in the range of H, $C\|\mathbf{z}_K\|_1 \leq \|\mathbf{z}_{\overline{K}}\|_1$, where K is the support of the error vector $\mathbf{e}$. Moreover, for a fixed constant $\beta < 1$, we assume that

$$\frac{2(C+1)}{C-1} \frac{\sigma^1_{max}(H^{true} - H^{local})}{\sigma^1_{min}(H^{local})} \leq \beta, \tag{V.6}$$

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 $h(\mathbf{x}_2) - h(\mathbf{x}_1) = H^{true}(\mathbf{x}_2 - \mathbf{x}_1)$, $\sigma^1_{max}(A)$ is the induced ℓ_1 matrix norm for A, and $\sigma^1_{min}(A)$ for a matrix A is defined as $\sigma^1_{min}(A) = \min\{\|A\mathbf{z}\|_1 : \text{ with } \|\mathbf{z}\|_1 = 1\}$.

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.

Proof: We know that

$$\mathbf{y} = H^{true}\Delta\mathbf{x}^* + h(\mathbf{x}^k) + \mathbf{e}, \tag{V.7}$$

where H^{true} is an $n \times m$ matrix 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

$$\min_{\Delta\mathbf{x}} \quad \|\mathbf{y} - h(\mathbf{x}^k) - H^{local}\Delta\mathbf{x}\|_1. \tag{V.8}$$

Plugging (V.7) into (V.8), this is equivalent to

$$\min_{\Delta\mathbf{x}} \quad \|H^{true}\Delta\mathbf{x}^* + \mathbf{e} - H^{local}\Delta\mathbf{x}\|_1, \tag{V.9}$$

which we can further write as

$$\min_{\Delta\mathbf{x}} \quad \|H^{local}\Delta\mathbf{x}^* + (H^{true} - H^{local})\Delta\mathbf{x}^* + \mathbf{e} - H^{local}\Delta\mathbf{x}\|_1. \tag{V.10}$$

We denote $(H^{true} - H^{local})\Delta\mathbf{x}^*$ as $\mathbf{w}$, which is the measurement gap generated by using the local Jacobian matrix H^{local} instead of H^{true}. Suppose that the solution is $\Delta\mathbf{x} = \Delta\mathbf{x}^* - error$. Since we are looking for the solution which minimizes the objective ℓ_1 norm, and $\Delta\mathbf{x} = \Delta\mathbf{x}^*$ is feasible for the optimization problem (V.8), we have

$$\|H^{local} \times error + \mathbf{w} + \mathbf{e}\|_1 \leq \|\mathbf{w} + \mathbf{e}\|_1. \tag{V.11}$$

By triangular inequality, we have

$$\|\mathbf{e}\|_1 + \frac{C-1}{C+1}\|H^{local} \times error\|_1 - \|\mathbf{w}\|_1 \leq \|\mathbf{e}\|_1 + \|\mathbf{w}\|_1. \tag{V.12}$$

So

$$\|H^{local} \times error\|_1 \leq \frac{2(C+1)}{C-1}\|\mathbf{w}\|_1. \tag{V.13}$$

Since $error = \Delta\mathbf{x}^* - \Delta\mathbf{x}$, $(\mathbf{x} - \mathbf{x}^k) = \Delta\mathbf{x}^*$, 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

$$\frac{\|\mathbf{x} - \mathbf{x}^{k+1}\|_1}{\|\mathbf{x} - \mathbf{x}^k\|_1} \leq \frac{2(C+1)}{C-1}\frac{\sigma^1_{max}(H^{true} - H^{local})}{\sigma^1_{min}(H^{local})}, \tag{V.14}$$

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

So as long as

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

for some constant $\beta < 1$, the algorithm converges to the true state $\mathbf{x}$ and the estimation error eventually decreases to 0. ■

While the algorithm can converge to the true state when there is no additive noise, the following theorem gives the performance bound for the iterative sparse recovery algorithm when there is additive noise.

Theorem 5.5 (Guarantee with Additive noise): Let $\mathbf{y}$, $h(\cdot)$, $\mathbf{x}$, H, $\mathbf{e}$, and n be specified as above; and $\mathbf{y} = h(\mathbf{x}) + \mathbf{e} + \mathbf{v}$ with $\|\mathbf{v}\|_2 \leq \epsilon$. Suppose that at every point $\mathbf{x}$, the local Jacobian matrix H is full rank and satisfies that for every $\mathbf{z}$ in the range of H, $C\|\mathbf{z}_K\|_1 \leq \|\mathbf{z}_{\overline{K}}\|_1$, where K is the support of the error vector $\mathbf{e}$. Moreover, for a fixed constant $\beta < 1$, we assume that

$$\frac{2(C+1)}{C-1}\frac{\sigma^1_{max}(H^{true} - H^{local})}{\sigma^1_{min}(H^{local})} \leq \beta \tag{V.16}$$

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 $h(\mathbf{x}_2) - h(\mathbf{x}_1) = H^{true}(\mathbf{x}_2 - \mathbf{x}_1)$, $\sigma^1_{max}(A)$ is the induced ℓ_1 matrix norm for A, and $\sigma^1_{min}(A)$ for a matrix A is defined as $\sigma^1_{min}(A) = \min\{\|A\mathbf{z}\|_1 : \text{ with } \|\mathbf{z}\|_1 = 1\}$.

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

$$\begin{aligned} \min_{\Delta \mathbf{x}^{k+1}, \mathbf{z}} \quad & \|\Delta \mathbf{y}^{k+1} - H^{local} \Delta \mathbf{x}^{k+1} - \mathbf{z}\|_1, \\ \text{subject to} \quad & \|\mathbf{z}\|_2 \leq \epsilon \end{aligned} \tag{V.17}$$

satisfies

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

As $k \to \infty$, with $\frac{2(C+1)}{C-1} \frac{\sigma^1_{max}(H^{true} - H^{local})}{\sigma^1_{min}(H^{local})} \leq \beta < 1$,

$$\|\mathbf{x} - \mathbf{x}^{k+1}\|_1 \leq \frac{2(C+1)}{(1-\beta)(C-1)\sigma^1_{min}(H^{local})} \times 2\sqrt{n}\epsilon.$$

Proof: The proof follows the same line of reasoning in proving Theorem 5.4 and Theorem 2.1. In fact,

$$\mathbf{y} = H^{true} \Delta \mathbf{x}^* + h(\mathbf{x}^k) + \mathbf{e} + \mathbf{v}, \tag{V.18}$$

where H^{true} is an $n \times m$ matrix 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

$$\begin{aligned} \min_{\Delta \mathbf{x}, \mathbf{z}} \quad & \|\Delta \mathbf{y} - H^{local} \Delta \mathbf{x} - \mathbf{z}\|_1, \\ \text{subject to} \quad & \|\mathbf{z}\|_2 \leq \epsilon. \end{aligned} \tag{V.19}$$

Plugging (V.18) into (V.19), we are really solving

$$\begin{aligned} \min_{\Delta \mathbf{x}, \mathbf{z}} \quad & \|H^{true} \Delta \mathbf{x}^* + \mathbf{e} + \mathbf{v} - H^{local} \Delta \mathbf{x} - \mathbf{z}\|_1, \\ \text{subject to} \quad & \|\mathbf{z}\|_2 \leq \epsilon. \end{aligned} \tag{V.20}$$

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.20) is equivalent to

$$\begin{aligned} \min_{\Delta\mathbf{x},\mathbf{z}} \quad & \|H^{local}(\Delta\mathbf{x}^* - \Delta\mathbf{x}) + \mathbf{w} + \mathbf{e} + \mathbf{v} - \mathbf{z}\|_1, \\ \text{subject to} \quad & \|\mathbf{z}\|_2 \leq \epsilon. \end{aligned} \tag{V.21}$$

Suppose that the solution to (V.17) is $\Delta\mathbf{x} = \Delta\mathbf{x}^* - error$. 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

$$\|H^{local} \times error + \mathbf{w} + \mathbf{e} + \mathbf{v} - \mathbf{z}\|_1 \leq \|\mathbf{w} + \mathbf{e}\|_1. \tag{V.22}$$

By triangular inequality and the property of H^{local}, using the same line of reasoning as in the proof of Theorem 2.1, we have

$$\begin{aligned} \|\mathbf{e}\|_1 \quad + \quad & \frac{C-1}{C+1}\|H^{local} \times error\|_1 - \|\mathbf{w}\|_1 - \|\mathbf{v}\|_1 - \|\mathbf{z}\|_1 \\ \leq \quad & \|\mathbf{e}\|_1 + \|\mathbf{w}\|_1. \end{aligned} \tag{V.23}$$

So

$$\|H^{local} \times error\|_1 \leq \frac{2(C+1)}{C-1}(\|\mathbf{w}\|_1 + \|\mathbf{v}\|_1 + \|\mathbf{z}\|_1). \tag{V.24}$$

Since $\|\mathbf{v}\|_1$ and $\|\mathbf{z}\|_1$ are both no bigger than $2\sqrt{n}\epsilon$, using the same reasoning as in the proof of Theorem 5.4, we have

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

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

So as long as

$$\frac{2(C+1)}{C-1}\frac{\sigma^1_{max}(H^{true} - H^{local})}{\sigma^1_{min}(H^{local})} \leq \beta, \tag{V.25}$$

for some fixed constant $\beta < 1$, the error upper bound converges to $\frac{2(C+1)}{(1-\beta)(C-1)\sigma^1_{min}(H^{local})} \times 2\sqrt{n}\epsilon$. ■

D. An Example of Sparse Recovery from Nonlinear Measurements

Now we give an example of sparse recovery from nonlinear measurements. For simplicity, we make the measurements corrupted with sparse bad data but not with additive noise. $h(\cdot)$ is a 12-dimensional vector as a mapping of two variables x and y, which is given in (V-D). We index the 12 measurements sequentially from top to bottom as 1,2,..., 12.

$$h(x,y) = \begin{pmatrix} (x+y)\sin(x+y) \\ (x+y)\cos(x+y) \\ (x-y)\sin(x-y) \\ (x-y)\cos(x-y) \\ (x+y)\sin(x-y) \\ (x-y)\sin(x+y) \\ (x+y)\cos(x-y) \\ (x-y)\cos(x+y) \\ x\sin(x) \\ x\cos(x) \\ y\sin(y) \\ y\cos(y) \end{pmatrix}$$

The following theorem shows that this set of nonlinear measurements are able to correct 1 corrupted entry in the measurements.

Lemma 5.6: If $(x_1, y_1) \neq (x_2, y_2)$, $\|h(x_1, y_1) - h(x_2, y_2)\|_0 \geq 3$, and so any state can be correctly recovered when at most one error is present in the measurements.

Proof: Suppose that $x_1 - y_1 = x_2 - y_2$ and $x_1 + y_1 \neq x_2 + y_2$. We further consider two cases. In the first case, $x_1 - y_1 = x_2 - y_2 = 0$, then for index 7, at least one of indices 1 and 2, and at least one of indices 9, 10, 11 and 12, $h(x_1, y_1)$ and $h(x_2, y_2)$ are different. In the second case, $x_1 - y_1 = x_2 - y_2 \neq 0$, then for at least one of indices 1 and 2, $h(x_1, y_1)$ and $h(x_2, y_2)$ are different; for at least one of indices 5 and 7, $h(x_1, y_1)$ and $h(x_2, y_2)$ are different; and for at least one of indices 9, 10, 11 and 12, $h(x_1, y_1)$ and $h(x_2, y_2)$ are different.

Suppose that $x_1 + y_1 = x_2 + y_2$ and $x_1 - y_1 \neq x_2 - y_2$. By symmetry to the previous scenario "$x_1 - y_1 = x_2 - y_2$ and $x_1 + y_1 \neq x_2 + y_2$", we have $\|h(x_1, y_1) - h(x_2, y_2)\|_0 \geq 3$.

Now we suppose that $x_1 - y_1 \neq x_2 - y_2$ and $x_1 + y_1 \neq x_2 + y_2$, then then for at least one of indices 1 and 2, $h(x_1, y_1)$ and $h(x_2, y_2)$ are different; for at least one of indices 3 and 4, $h(x_1, y_1)$ and $h(x_2, y_2)$

are different; and for at least one of indices 9, 10, 11 and 12, $h(x_1, y_1)$ and $h(x_2, y_2)$ are different.

Summarizing all these scenarios, if $(x_1, y_1) \neq (x_2, y_2)$, $\|h(x_1, y_1) - h(x_2, y_2)\|_0 \geq 3$. ■

So this system of nonlinear measurements can guarantee correcting 1 bad data entry. But can we efficiently find the true state from bad data using the iterative sparse recovery algorithm in Subsection V-B? To proceed, we first give the Jacobian matrix for $h(\cdot)$ in (V.6).

Suppose the true state is $(x, y) = (0.2, 0.45)$, and suppose there is one bad data entry in the measurements, where we let $\mathbf{e} = (0, 0, 0, 0, 0, 1.7783, 0, 0, 0, 0, 0, 0)$. Suppose that the iterative sparse recovery algorithm starts with the initial state $(x_0, y_0) = (0.1, 0.2)$. Then by definition, at the initial point $(x_0, y_0) = (0.1, 0.2)$, the local Jacobian matrix is

$$H^{local} = \begin{pmatrix} 0.5821 & 0.5821 \\ 0.8667 & 0.8667 \\ -0.1993 & 0.1993 \\ 0.9850 & -0.9850 \\ 0.1987 & -0.3983 \\ 0.2000 & -0.3911 \\ 1.0250 & 0.9651 \\ 0.9849 & -0.9258 \\ 0.1993 & 0 \\ 0.9850 & 0 \\ 0 & 0.3947 \\ 0 & 0.9403 \end{pmatrix} \tag{V.26}$$

Then by using the mean value theorem in two variables for the twelve functions in $h(\cdot)$, we can calculate a H^{true} as

$$H^{true} = \begin{pmatrix} 0.8707 & 0.8707 \\ 0.6596 & 0.6596 \\ -0.3459 & 0.3459 \\ 0.9515 & -0.9515 \\ 0.2932 & -0.6407 \\ 0.2984 & -0.6063 \\ 1.0709 & 0.8968 \\ 0.9683 & -0.8013 \\ 0.2975 & 0 \\ 0.9651 & 0 \\ 0 & 0.6239 \\ 0 & 0.8367 \end{pmatrix} \tag{V.27}$$

For the small example with two variables, we can calculate $\sigma^1_{max}(H^{true}-H^{local}) = 1.6590$, $\sigma^1_{min}(H^{local}) = 3.9284$, and $C = 13.5501$. So

$$\frac{2(C+1)}{C-1}\frac{\sigma^1_{max}(H^{true}-H^{local})}{\sigma^1_{min}(H^{local})} = 0.9792 < 1, \tag{V.28}$$

which satisfies the condition appearing in Theorem 5.4.

In fact, in the first iteration, the algorithm outputs $(x_1, y_1) = (0.1980, 0.4392)$ and $\|(x_1, y_1)-(x, y)\|_2 = 0.0110$. After the second iteration, we already get a very good estimation $(x_2, y_2) = (0.2000, 0.4500)$ and $\|(x_2, y_2) - (x, y)\|_2 = 2.2549 \times 10^{-5}$. The solution does converge to the true state.

We note that the convergence condition in Theorem 5.4 is conservative. Sometimes even if the initial starting point is far from the true state and the convergence condition fails, the algorithm can still converge. For example, now suppose that the true state is at $(x, y) = [0.4, 1.2]$ and the iterative sparse recovery algorithm still initializes with $(x_0, y_0) = (0.1, 0.2)$. Suppose that the bad data vector is still $\mathbf{e} = (0, 0, 0, 0, 0, 1.7783, 0, 0, 0, 0, 0, 0)$. Then again by using the mean value theorem, we calculate a

$$J(x,y) = \begin{pmatrix} \sin(x+y)+(x+y)\cos(x+y) & \sin(x+y)+(x+y)\cos(x+y) \\ \cos(x+y)-(x+y)\sin(x+y) & \cos(x+y)-(x+y)\sin(x+y) \\ \sin(x-y)+(x-y)\cos(x-y) & -\sin(x-y)-(x-y)\cos(x-y) \\ \cos(x-y)-(x-y)\sin(x-y) & -\cos(x-y)+(x-y)\sin(x-y) \\ \sin(x-y)+(x+y)\cos(x-y) & \sin(x-y)-(x+y)\cos(x-y) \\ \sin(x+y)+(x-y)\cos(x+y) & -\sin(x+y)+(x-y)\cos(x+y) \\ \cos(x-y)-(x+y)\sin(x-y) & \cos(x-y)+(x+y)\sin(x-y) \\ \cos(x+y)-(x-y)\sin(x+y) & -\cos(x+y)-(x-y)\sin(x+y) \\ \sin(x)+x\cos(x) & 0 \\ \cos(x)-x\sin(x) & 0 \\ 0 & \sin(y)+y\cos(y) \\ 0 & \cos(y)-y\sin(y) \end{pmatrix} \tag{V.6}$$

H^{true} as

$$H^{true} = \begin{pmatrix} 1.1621 & 1.1621 \\ -0.2566 & -0.2566 \\ -0.8055 & 0.8055 \\ 0.6543 & -0.6543 \\ 0.4119 & -1.2413 \\ 0.4068 & -0.8921 \\ 1.3597 & 0.4083 \\ 0.9437 & -0.1640 \\ 0.4860 & 0 \\ 0.8964 & 0 \\ 0 & 1.0786 \\ 0 & 0.2385 \end{pmatrix} \tag{V.29}$$

In the first iteration, we get a new estimation of the state $(x_1, y_1) = (0.3730, 0.7558)$ and $\|(x_1, y_1) - (x, y)\|_2 = 0.4450$. After the second iteration, we get a new estimation $(x_2, y_2) = (0.3995, 1.1468)$ and $\|(x_2, y_2) - (x, y)\|_2 = 0.0532$. After the third iteration, we get a new estimation $(x_3, y_3) = (0.400, 1.2003)$ and $\|(x_3, y_3) - (x, y)\|_2 = 2.96 \times 10^{-4}$. The algorithm converges to the true state even though in the first step, $\sigma^1_{max}(H^{true} - H^{local}) = 6.6885$, $\sigma^1_{min}(H^{local}) = 3.9284$, $C = 13.5501$ and

$$\frac{2(C+1)}{C-1}\frac{\sigma^1_{max}(H^{true} - H^{local})}{\sigma^1_{min}(H^{local})} = 3.9478 > 1. \tag{V.30}$$

VI. Numerical Results

In our simulation, we apply (II.6) 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. $N(0, 1)$ entries. We also generate a vector $\mathbf{x} \in R^m$ with i.i.d Gaussian entries and normalize it to $\|\mathbf{x}\|_2 = 1$.

We fix the noise level and consider the estimation performance when the number of erroneous measurements changes. We add to each measurement of $H\mathbf{x}$ with a Gaussian noise independently drawn from $N(0, 0.5^2)$. Let ρ denote the percentage of erroneous measurements. Given ρ, we randomly choose ρn measurements, and each such measurement is added with a Gaussian error independently drawn from $N(0, 5^2)$. We apply (II.6) to estimate $\mathbf{x}$ using different choices of λ. Let $\mathbf{x}^*$ denote the estimation of $\mathbf{x}$, and the estimation error is represented by $||\mathbf{x}^* - \mathbf{x}||_2$. We use (II.6) instead of (II.1) in simulation, since the recovering algorithm has no prior knowledge of the noise vector, and solving an unconstrained optimization problem is more computationally efficient than solving a constrained one.

Fig. 3 shows how the estimation error changes as ρ increases, where each result is averaged over one hundred and fifty runs. As discussed earlier, when λ is large, like $\lambda = 18$ in this example, (II.6) approaches ℓ_1-minimization; when λ is close to zero, like $\lambda = 0.05$ here, (II.6) approaches ℓ_1-minimization; when $\lambda = 8$, (II.6) can be viewed as a weighted version of ℓ_1 and ℓ_2 minimization. When ρ is zero or close to one, the measurements only contain i.i.d. Gaussian noises, thus, among the three choices of λ, the estimation error is relatively small when $\lambda = 0.05$. When ρ is away from zero and one, the measurements contain both noise and sparse errors, then a weighted version of ℓ_1 and ℓ_2 minimization (represented by the case $\lambda = 8$) outperforms both ℓ_1-minimization (approximated by the case $\lambda = 18$) and ℓ_2-minimization (approximated by the case $\lambda = 0.05$) in terms of a small estimation error.

We next consider the recovery performance when the number of erroneous measurements is fixed. We randomly choose twelve measurements and add to each such measurement an independent Gaussian error from $N(0, 5^2)$. Then, we add an independent Gaussian noise from $N(0, \sigma^2)$ to each one of the n measurements. Fig. 4 shows how the estimation error $||\mathbf{x}^* - \mathbf{x}||_2$ changes as σ increases with different choices of λ. When σ is close to zero, the effect of sparse errors are dominating, thus ℓ_1-minimization (approximated by the case $\lambda = 18$) has the best recovery performance. When σ is large, the effect of i.i.d. Gaussian noises are dominating, thus ℓ_2-minimization (approximated by the case $\lambda = 0.05$) has the

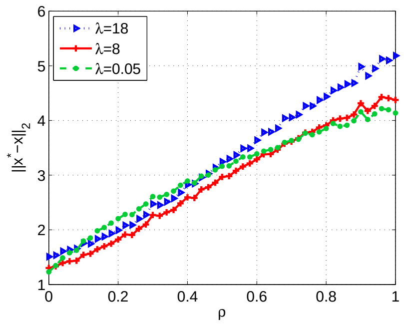

Fig. 3: Estimation error versus ρ for Gaussian measurements with fixed noise level

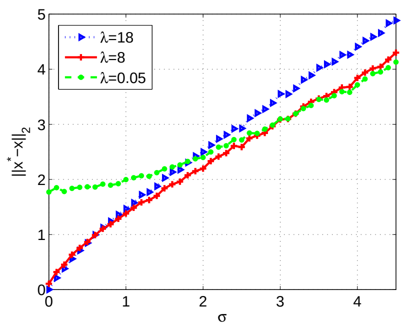

Fig. 4: Estimation error versus σ for Gaussian measurements with fixed percentage of errors

best recovery performance. In between, a weighted version of ℓ_1 and ℓ_2 minimization (represented by the case $\lambda = 8$) has the best performance.

For a given σ, we also apply (II.6) with λ from 0.05 to 12.05 (step size 0.2), and pick the best λ^* with which the estimation error is minimized. For each σ, the result is averaged over three hundred runs. Fig. 5 shows the curve of λ^* against σ. When the percentage of measurements with bad data is fixed, ($\rho = 12/100 = 0.12$ here,) λ^* decreases as the noise level increases.

Power System: We also consider estimating the state of a power system from available measurements and known system configuration. The state variables are the voltage magnitudes and the voltage angles at each bus. The measurements can be the real and reactive power injections at each bus, and the real and reactive power flows on the lines. All the measurements are corrupted with noise, and a small fraction of the measurements contains errors. We would like to estimate the state variables from the corrupted measurements.

The relationship between the measurements and the state variables for a k'-bus system can be stated

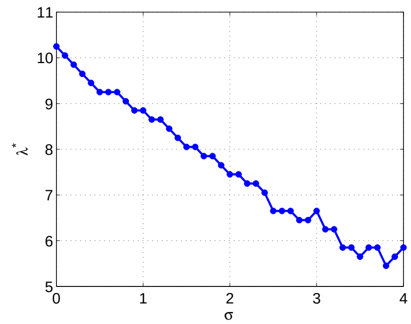

Fig. 5: λ^* versus σ for Gaussian measurements

as follows [12]:

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

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

$$P_{ij} = E_i E_j Y_{ij} \cos(\theta_{ij} + \delta_i - \delta_j) - E_i^2 Y_{ij} \cos\theta_{ij} + E_i^2 Y_{si} \cos\theta_{si} \quad i \neq j, \quad \text{(VI.3)}$$

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

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 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,

$$\mathbf{y} = h(\mathbf{x}) + \mathbf{v} + \mathbf{e}, \quad \text{(VI.5)}$$

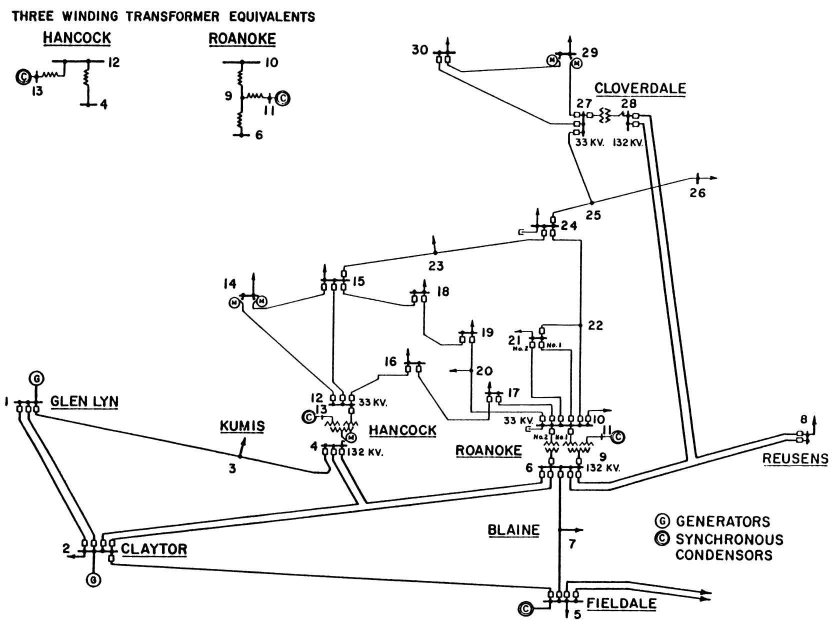

Fig. 6: IEEE 30-bus test system

where $h(\cdot)$ denotes n nonlinear functions defined in (VI.1) to (VI.4).

We use the iterative algorithm introduced in Subsection V-B to recover $\mathbf{x}$ from $\mathbf{y}$. We start with the initial state $\mathbf{x}^0$ where $x_i^0 = 1$ for all $i \in \{1, ..., n\}$, and $x_i^0 = 0$ for all $i \in \{n+1, ..., 2n-1\}$. Since we assume no knowledge of the magnitude of $\mathbf{v}$ and unconstrained problem is generally more computationally efficient than a constrained one, in the kth iteration, instead of solving (V.4), we solve the following unconstrained convex optimization problem

$$\min_{\Delta\mathbf{x}, \mathbf{z}} \quad \|\Delta\mathbf{y}^k - H^{local}\Delta\mathbf{x} - \mathbf{z}\|_1 + \lambda\|\mathbf{z}\|_2, \tag{VI.6}$$

where H^{local} is the Jacobian matrix of h evaluated at $\mathbf{x}^{k-1}$. Let $\Delta\mathbf{x}^k$ denote the optimal solution of $\Delta\mathbf{x}$ to (VI.6), then the state estimation is updated by

$$\mathbf{x}^k = \mathbf{x}^{k-1} + \Delta\mathbf{x}^k. \tag{VI.7}$$

We repeat the process until $\Delta\mathbf{x}^k$ is close to 0, or the number of iteration reaches a specified value.

We evaluate the performance on the IEEE 30-bus test system. Fig. 6 shows the structure of the test system. Then the state vector contains 59 variables. 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. evaluate how the estimation performance changes as the noise level increases when the number of

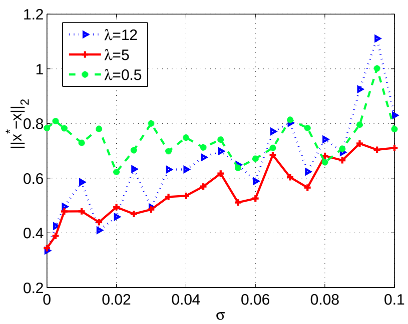

Fig. 7: Estimation error versus σ with fixed percentage of errors in power system

erroneous measurements is fixed. ρ is fixed to be 0.06, and we randomly choose a set T with cardinality $|T| = \rho n = 6$. Each measurement with its index in T contains a Gaussian error independently drawn from $N(0, 0.7^2)$. Each measurement also contains a Gaussian noise independently drawn from $N(0, \sigma^2)$. For a fixed noise level σ, we use the above mentioned iterative procedure to recover the state vector $\mathbf{x}$. The result is averaged over two hundred runs. Fig. 7 shows the estimation error $\|\mathbf{x}^* - \mathbf{x}\|_2$ against σ when $\rho = 0.06$. Between $\lambda = 12$ (approximating ℓ_1-minimization) and $\lambda = 0.5$ (approximating ℓ_2-minimization), the former one has a better recovery performance when the noise level σ is small, and the latter one has a better performance when σ is large. Moreover, the recovery performance when $\lambda = 5$ in general outperforms that when λ is either large ($\lambda = 12$) or small ($\lambda = 0.05$).

VII. Conclusion

In this paper, we studied sparse recovery from nonlinear measurements with applications in state estimation for power networks from nonlinear measurements corrupted with bad data. An iterative mixed ℓ_1 and ℓ_2 convex programming was proposed for state estimation by locally linearizing the nonlinear measurements. By studying the almost Euclidean property for a linear subspace, we gave a new state estimation error bound when the measurements are linear and the measurements are corrupted with both bad data and by additive noise. When the measurements are nonlinear and corrupted with bad data, we gave conditions under which the solution of the iterative algorithm converges to the true state even though local linearizing of measurements may not be accurate. We numerically evaluated the iterative convex programming approach performance in bad data detection for nonlinear electrical power networks. As a byproduct, we provided sharp bounds on the almost Euclidean property of a linear subspace, using the "escape-through-a-mesh" theorem from geometric functional analysis.

ACKNOWLEDGMENT

The research is supported by NSF under CCF-0835706, ONR under N00014-11-1-0131, and DTRA under HDTRA1-11-1-0030.

REFERENCES

[1] T. Blumensath, “Compressed Sensing with Nonlinear Observations,” *http://eprints.soton.ac.uk/164753/*

[2] Anjan Bose and Kevin Clements, “Real-Time Modeling of Power Networks,” *Proceedings of IEEE*, 75(12), 1607-1622, 1987

[3] Broussolle, “State estimation in power systems: Detecting bad data through the sparse inverse matrix method,” *IEEE Trans. Power App. Syst.*, vol. PAS-94, pp. 329-337, Mar./Apr. 1975.

[4] E. Candès and P. Randall, “Highly robust error correction by convex programming,” *IEEE Transactions on Information Theory*, vol. 54, pp. 2829-2840.

[5] E. Candès and T. Tao, “Decoding by linear programming”, *IEEE Trans. on Information Theory*, 51(12), pp. 4203 - 4215, December 2005.

[6] A. Yu. Garnaev and E. D. Gluskin,“The widths of a Euclidean ball,” (Russian), Dokl. Akad. Nauk SSSR 277 (1984), no. 5, 10481052. (English translation: Soviet Math. Dokl. 30 (1984), no. 1, 200204.)

[7] S. Geman, “A limit theorem for the norm of random matrices,” *Annals of Probability*, vol. 8, No. 2, pp.252-261, 1980.

[8] 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*, 1317,1988.

[9] B.S. Kasin, “The widths of certain finite-dimensional sets and classes of smooth functions,” (Russian), Izv. Akad. Nauk SSSR Ser. Mat. 41, 334-351 (1977). (English translation: Math. USSR-Izv. 11 (1977), no. 2, 317-333 (1978)).

[10] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, ”On malicious data attacks on power system state estimation,” *Proceedings of UPEC*,Cardiff, Wales, UK, 2010.

[11] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, “Malicious data attacks on smart grid state estimation: attack strategies and countermeasures,” *Proceedings of IEEE 2010 SmartGridComm*, 2010.

[12] W. Kotiuga and M. Vidyasagar, “Bad data rejection properties of weighted least absolute value techniques applied to static state estimation,” *IEEE Trans. Power App. Syst.*, vol. PAS-101, pp. 844-853, Apr. 1982.

[13] 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.

[14] A. Monticelli and A. Garcia, “Reliable bad data processing for real time state estimation,” *IEEE Trans. Power App. Syst.*, vol. PAS-102, pp.1126-1139, May 1983.

[15] M. Rudelson and R. Vershynin, “On sparse reconstruction from Fourier and Gaussian measurements”, *Comm. on Pure and Applied Math.*, 61(8), 2007.

[16] J. Silverstein, “The smallest eigenvalue of a large dimensional Wishart matrix,” *Annals of Probability*, vol. 13, pp. 1364-1368, 1985

[17] M. Stojnic, “Various thresholds for ℓ_1-optimization in Compressed Sensing,” *http://arxiv.org/abs/0907.3666*

[18] W. Xu, M. Wang, and A. Tang, “On state estimation with bad data detection,” *Proceedings of IEEE Conference on Decision and Control*, 2011.

[19] Y. Zhang, “A simple proof for recoverability of ℓ_1-minimization: go over or under?” Technical Report, *http://www.caam.rice.edu/$\sim$yzhang/reports/tr0509.pdf*.