

# From Sparse Signals to Sparse Residuals for Robust Sensing

Vassilis Kekatos, *Member, IEEE*, and Georgios B. Giannakis, *Fellow, IEEE*

**Abstract**—One of the key challenges in sensor networks is the extraction of information by fusing data from a multitude of distinct, but possibly unreliable sensors. Recovering information from the maximum number of dependable sensors while specifying the unreliable ones is critical for robust sensing. This sensing task is formulated here as that of finding the maximum number of feasible subsystems of linear equations and proved to be NP-hard. Useful links are established with compressive sampling, which aims at recovering vectors that are sparse. In contrast, the signals here are not sparse, but give rise to sparse residuals. Capitalizing on this form of sparsity, four sensing schemes with complementary strengths are developed. The first scheme is a convex relaxation of the original problem expressed as a second-order cone program (SOCP). It is shown that when the involved sensing matrices are Gaussian and the reliable measurements are sufficiently many, the SOCP can recover the optimal solution with overwhelming probability. The second scheme is obtained by replacing the initial objective function with a concave one. The third and fourth schemes are tailored for noisy sensor data. The noisy case is cast as a combinatorial problem that is subsequently surrogated by a (weighted) SOCP. Interestingly, the derived cost functions fall into the framework of robust multivariate linear regression, while an efficient block-coordinate descent algorithm is developed for their minimization. The robust sensing capabilities of all schemes are verified by simulated tests.

**Index Terms**—Compressive sampling, convex relaxation, coordinate descent, multivariate regression, robust methods, sensor networks.

## I. INTRODUCTION

RECENT advances in sensor technology have made it feasible to deploy a network of inexpensive sensors for carrying out synergistically even sophisticated inference tasks. In applications such as environmental monitoring, surveillance of critical infrastructure, agriculture, or medical imaging, the typical concept of operation involves a large and possibly heterogeneous set of sensors locally observing the signal of interest and transmitting their measurements to a higher-layer agent (fusion

center). This so-termed layered sensing apparatus entails three operational conditions:

- c1) Each node's measurement vector comprising either a collection of scalar observations across time, or a snapshot of different sensor readings, is typically assumed to be linearly related to the unknown variable(s). Such a *linear* model can arise when the sensing system is viewed as a linear filter with known impulse response. Even when the underlying model is nonlinear, the observations are approximately modeled as adhering to a (multivariate) linear regression;
- c2) Either because readings are costly to sense and transmit, due to delay or stationarity constraints, or simply because dimensionality reduction is invoked to cope with the "curse of dimensionality," the linear model is often-times *underdetermined*, i.e., the dimension of the unknown vector is larger than that of each sensor's vector observation; and
- c3) Not all sensors are *reliable* because failures in the sensing devices, fades of the sensor-agent communication link, physical obstruction of the scene of interest and (un)intentional interference, all can severely deteriorate the consistency and reliability of sensor data.

Conditions c1)–c3) suggest that the fusion center should not simply aggregate all sensor measurements, but instead identify and discard unreliable sensors before estimating the unknown vector based on reliable sensor data. This task is henceforth referred to as *robust sensing* (RS) and provides context of the present paper. Discerning the unreliable sensors not only promises higher estimation accuracy, but also enables corrective actions to re-establish a sensor's reliability, by e.g., remotely directing the sensor to the area of interest, or, increasing its sensitivity. Even though the related problem of outlier detection in sensor networks has been studied extensively (see, e.g., [33] for a recent survey), the RS setup and the approaches described here have not been considered before.

The first contribution of this work is to formulate the RS task as an optimization problem based on the sensor data and show it to be NP-hard (Section II). The second one consists of two (sub)-optimum RS solvers (Section III). The first solver is expressed as a second-order cone program (SOCP) through a convex relaxation of the original NP-hard problem. The idea of convex relaxation has been employed in the emerging area of compressive sampling (CS) [8], [9], [28]. CS asserts that a sparse vector (i.e., one having many zero entries) can be recovered with overwhelming probability as the vector with minimum  $\ell_1$ -norm satisfying an underdetermined system of linear

Manuscript received May 11, 2010; revised November 01, 2010 and March 27, 2011; accepted March 31, 2011. Date of publication April 11, 2011; date of current version June 15, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jean-Yves Tournier. This work was supported by the European Community's Seventh Framework Programme (FP7/2008 under Grant 234914) and by NSF Grant CCF-1016605. Part of this work was presented at the Eleventh IEEE International Workshop on Signal Processing Advances in Wireless Communications, Marrakech, Morocco, June 2010.

The authors are with the Department of Electrical and Computer Engineering and the Digital Technology Center, University of Minnesota, MN 55455 USA (e-mail: kekatos@umn.edu; georgios@umn.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.2011.2141661

equations; a setup known as basis pursuit (BP) [8], [9], [29]. CS has been generalized to block-sparse signals, where the unknown vector comprises predetermined subsets of variables that are (non) zero as a group [5], [11], [26], [27]. Block sparsity emerges also in the RS formulation herein, not in the unknown vector though, but in the per-sensor residual error vectors. The relation between recovering block-sparse signals and the developed RS solver nicely generalizes the equivalence of BP with  $\ell_1$ -error regression from the scalar to the vector case. As an alternative to convex relaxation, the  $\ell_0$ -(pseudo)norm of the wanted vector can be replaced by a concave approximation to further promote sparsity [7], [12]. This constitutes the second RS solver, which surrogates the original objective by a concave function and minimizes it through a sequence of weighted SOCPs.

The third contribution consists in analyzing the performance (identifiability) of the convex relaxation approach to recover the unknown vector and successfully select the reliable sensors in the noise-free case (Section IV). The analysis hinges on a set of necessary and sufficient conditions on the involved matrix range space, which appear also in the context of [27]. Here a lower bound expressed in closed form is established on the probability of success when the design matrix is drawn from the Gaussian ensemble; see also [24]. It is shown that whenever there is sufficient majority of reliable sensors and quantifiably enough per-sensor measurements, the solution of the SOCP is exact with overwhelming probability.

In real-world applications, sensor readings are contaminated by additive noise due to quantization, communication noise and/or unmodeled dynamics. Besides identifiability, the aforementioned schemes are thus appropriate only for the high signal-to-noise ratio (SNR) regime. When the sparse vector in CS is observed in noise, its recovery is based on methods such as the Lasso [28], or the group Lasso for vectors that are block-sparse [32]. Different from CS, the approach here views the unreliable sensors as outliers, thus placing the sensing in the presence of noise (RSN) task under a robust multivariate linear regression framework [2], [4]. The fourth contribution of this work (Section V) is initially formulating RSN as a combinatorial optimization problem that is subsequently surrogated by a convex approximation. Interestingly, the novel cost function turns out to be a block version of Huber's function [17]. The resultant optimization problem is transformed to a group Lasso-type SOCP and a computationally attractive block-coordinate descent algorithm is developed. An alternative RSN solver is also offered after replacing the previously derived convex problem with a nonconvex one. The simulated tests presented in Section VI corroborate the proposed schemes and the paper is concluded in Section VII.

*Notation:* Lowercase (upper-case) boldface letters are reserved for column vectors (matrices) and calligraphic letters for sets;  $(\cdot)^T$  denotes transposition;  $\mathcal{N}(\mathbf{m}, \mathbf{\Sigma})$  stands for the multivariate Gaussian probability density with mean  $\mathbf{m}$  and covariance matrix  $\mathbf{\Sigma}$ , while  $\mathbb{E}[\cdot]$  denotes the expectation operator. The notation  $\|\mathbf{x}\|_p := (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}}$  for  $p = 1(2)$  stands for the  $\ell_1(\ell_2)$ -norm in  $\mathbb{R}^n$  and  $\|\mathbf{x}\|_0$  the  $\ell_0$ -(pseudo)norm which equals the number of nonzero entries of  $\mathbf{x}$ .

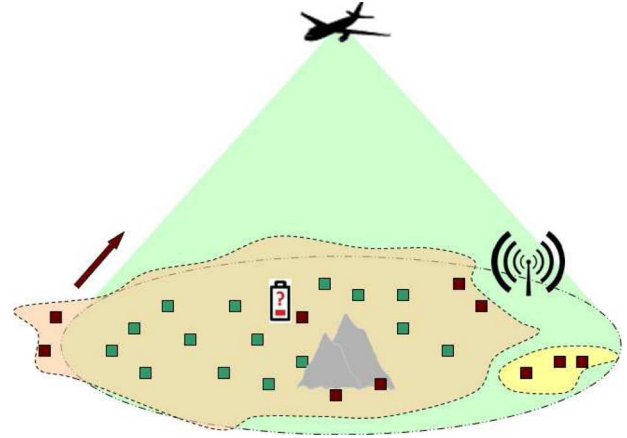


Fig. 1. A wireless sensor network linked with a fusion center. (Un)reliable sensors are color coded as (red) green.

## II. PRELIMINARIES AND PROBLEM STATEMENT

Consider an agent, e.g., an unmanned aerial vehicle, collecting data vectors  $\{\mathbf{b}_i\}_{i=1}^k$  of size  $m_i \times 1$  and corresponding  $m_i \times n$  regression matrices  $\{\mathbf{A}_i\}_{i=1}^k$  from  $k$  sensors. The goal is to find an unknown vector  $\mathbf{x} \in \mathbb{R}^n$ , possibly satisfying the linear subsystems of equations  $\mathbf{b}_i = \mathbf{A}_i \mathbf{x}$  for some  $i \in \{1, \dots, k\}$ . This goal is challenging since the unknown vector  $\mathbf{x}$  satisfies only an *unknown* subset of sensors. The RS problem can be compactly stated as follows.

**Problem Statement 1 (Robust Sensing (RS)):** *Given  $k$  vector-matrix pairs,  $\{\mathbf{b}_i, \mathbf{A}_i\}_{i=1}^k$ , where  $\mathbf{b}_i \in \mathbb{R}^{m_i}$  and  $\mathbf{A}_i \in \mathbb{R}^{m_i \times n}$ , find a vector  $\mathbf{x} \in \mathbb{R}^n$  that maximizes the number of feasible linear subsystems  $\{\mathbf{b}_i = \mathbf{A}_i \mathbf{x}\}$ .*

Vector  $\mathbf{x}$  could model a scene (lexicographically ordered image) of interest viewed by multiple and possibly heterogeneous, e.g., Infrared, SAR, or, Lidar imaging systems. Matrices  $\mathbf{A}_i$  may capture variable fields of view, different perspectives and resolutions in some (e.g., wavelet) domain, or, calibration parameters of the respective sensors. Alternatively, in an environmental monitoring application,  $\mathbf{x}$  could represent the unknown parameters of a chemical/biological compound diffusion field described by the Green's function captured by the matrices  $\{\mathbf{A}_i\}_{i=1}^k$  and measured by a wireless sensor network deployed over a region of interest. In such sensing applications, a sensor may be reckoned unreliable or irrelevant due to obstruction, fading propagation effects, device failures, jamming, or, even because it collects data corresponding to an irrelevant  $\mathbf{x}' \neq \mathbf{x}$ ; see Fig. 1.

The RS task is different for over- and underdetermined linear subsystems. Assume that all  $\mathbf{A}_i$ 's are full rank, i.e.,  $\text{rank}(\mathbf{A}_i) = \min\{m_i, n\}$  for all  $i$ .<sup>1</sup> Then, suppose that the  $i$ th linear subsystem is overdetermined ( $m_i > n$ ). This subsystem is either infeasible and can be ignored, or, it admits a unique solution  $\tilde{\mathbf{x}}_i$ . In the latter case, it can be easily checked whether  $\tilde{\mathbf{x}}_i$  satisfies any other subsystem. The solution  $\tilde{\mathbf{x}}_i$  together with the total

<sup>1</sup>This is without loss of generality (w.l.o.g.), because every sensor with  $\text{rank}(\mathbf{A}_i) < \min\{m_i, n\}$  will be either infeasible, or, it can be transformed to an underdetermined subsystem with full row rank.

number of subsystems it satisfies are retained and the method proceeds similarly with all other overdetermined subsystems. However, checking the underdetermined subsystems ( $m_i < n$ ) is more challenging, since each one of them admits infinitely many solutions. Recognizing that overdetermined subsystems can be easily handled, this paper focuses on the RS task when  $m_i < n$  for all  $i$ . Note that underdeterminacy may arise naturally because of stringent power, bandwidth, delay, or stationarity constraints. Given that the  $\mathbf{b}_i$ 's ( $\mathbf{A}_i$ 's) can be padded with zero entries (rows) to match the dimension  $\max_i m_i$ , it will be henceforth assumed w.l.o.g.  $m_i = m < n$  for all  $i$ .

Before proceeding, it is useful to introduce some parameters. The set of all subsystem indexes is denoted by  $\mathcal{I} := \{1, \dots, k\}$ , whereas the pair  $(\mathcal{S}, \bar{\mathcal{S}})$  denotes a partition of  $\mathcal{I}$  into the subset  $\mathcal{S}$  and its complement  $\bar{\mathcal{S}}$  ( $\mathcal{S} \cup \bar{\mathcal{S}} = \mathcal{I}$ ,  $\mathcal{S} \cap \bar{\mathcal{S}} = \emptyset$ ). Consider now the  $|\mathcal{S}|m \times n$  matrix  $\mathbf{A}_{\mathcal{S}}$  constructed by concatenating the matrices  $\{\mathbf{A}_i\}_{i \in \mathcal{S}}$  and likewise for the vector  $\mathbf{b}_{\mathcal{S}}$ . The aggregate regression matrix and data vector are defined as  $\mathbf{A}^T := [\mathbf{A}_1^T \dots \mathbf{A}_k^T]$  and  $\mathbf{b}^T := [\mathbf{b}_1^T \dots \mathbf{b}_k^T]$ , respectively.

Upon introducing an auxiliary vector  $\mathbf{t} \in \mathbb{R}^k$ , the RS problem can be rigorously posed as

$$\begin{array}{ll} \min_{\mathbf{x}, \mathbf{t}} & \|\mathbf{t}\|_0 \\ \text{s.t.} & \|\mathbf{b}_i - \mathbf{A}_i \mathbf{x}\|_2 \leq t_i, \quad i = 1, \dots, k. \end{array} \quad (\text{P0})$$

If the  $i$ th subsystem is deemed feasible, then  $t_i = 0$ ; otherwise,  $t_i$  is strictly positive and the cost  $\|\mathbf{t}\|_0$  increases. In a nutshell, (P0) minimizes the number of infeasible linear subsystems and hence solves RS. Note also that the constraints are satisfied as equalities at the optimum. Thus, if the optimum  $\mathbf{x}$  is given, the optimum  $\mathbf{t}$  is readily available. This implies that the solution pair  $(\mathbf{x}, \mathbf{t})$  is identified solely by  $\mathbf{x}$ , which will be henceforth called the *solution* of (P0).

Even though the constraints in (P0) are convex, the problem is nonconvex. A greedy approach to solving it would be to assume there are  $s$  feasible subsystems and let  $s$  range from  $k$  down to 1. For each value of  $s$ , one can check feasibility of the linear systems  $\mathbf{b}_{\mathcal{S}} = \mathbf{A}_{\mathcal{S}} \mathbf{x}$  for each of the  $\binom{k}{s}$  subsets  $\mathcal{S} \subset \mathcal{I}$  having cardinality  $|\mathcal{S}| = s$ , until a feasible subset is found. But this approach incurs combinatorial complexity and can be computationally feasible only for small-size problems. In fact, it is not difficult to establish the following result.

**Proposition 1:** *The RS problem is NP-hard.*

*Proof:* Consider first the following problem of maximizing the number of consistent linear equations (MCLE): “Given a system of linear equations  $\mathbf{C}\mathbf{x} = \mathbf{d}$ , where  $\mathbf{C} \in \mathbb{R}^{k \times n}$  and  $\mathbf{d} \in \mathbb{R}^k$ , find a vector  $\mathbf{x} \in \mathbb{R}^n$  satisfying as many equations as possible.” The MCLE problem is known to be NP-hard [3, Th. 1]. Consider an instance of the MCLE problem. Choose an integer  $m \geq 2$  and define the instance of RS with parameters  $(\mathbf{b}, \mathbf{A})$  selected as  $b_{(i-1)m+1} = d_i$  and  $A_{(i-1)m+1, j} = C_{i, j}$  for  $i = 1, \dots, k$  and  $j = 1, \dots, n$ ; and 0 for their remaining entries. Solving an MCLE problem is hence equivalent to solving an instance of an RS. This simple reduction of MCLE to RS establishes the proposition. ■

In search of sub-optimum yet computationally affordable solvers of (P0), one could adopt the least-squares (LS) approach, which amounts to

$$\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2. \quad (1)$$

Alternatively, one could consider minimizing the  $\ell_1$ -norm of the error, namely

$$\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_1. \quad (2)$$

Unfortunately, both approaches handle separately every linear equation and thus ignore the underlying per-sensor linear subsystem. In addition, they cannot reliably identify the unreliable sensors.

### III. RS SOLVERS

#### A. A Convex Relaxation Solver

It is known that if the infinity norm satisfies  $\|\mathbf{t}\|_{\infty} := \max_i |t_i| \leq 1$ , then the  $\ell_1$ -norm  $\|\mathbf{t}\|_1$  is the convex envelope (the largest convex underapproximant) of  $\|\mathbf{t}\|_0$ ; see e.g., [6, p. 119]. This property is used also in CS [29] and prompts one to relax the NP-hard problem (P0) to

$$\begin{array}{ll} \min_{\mathbf{x}, \mathbf{t}} & \|\mathbf{t}\|_1 \\ \text{s.t.} & \|\mathbf{b}_i - \mathbf{A}_i \mathbf{x}\|_2 \leq t_i, \quad i = 1, \dots, k. \end{array} \quad (3)$$

Note though that  $\mathbf{x}$  here does not have to be sparse. The problem in (3) is an SOCP and can be efficiently solved by several existing algorithms [6]. Invoking the implicit constraint  $\mathbf{t} \geq \mathbf{0}$  and the definition of the  $\ell_1$ -norm  $\|\mathbf{t}\|_1 := \sum_{i=1}^k |t_i|$ , the problem (3) is equivalent to

$$\min_{\mathbf{x}} \sum_{i=1}^k \|\mathbf{b}_i - \mathbf{A}_i \mathbf{x}\|_2 \quad (\text{P1})$$

which is still an SOCP, albeit unconstrained.

The cost in (P1) is the sum of the  $\ell_2$ -norms of the residual vectors associated with the linear subsystems, which is continuous, but not differentiable. In the optimization circles, (P1) is known as the minimization of the sum of (Euclidean) norms problem [6, Sec. 6.4]. It emerges also when solving problems related to Steiner trees, optimal location and image restoration model constraints; see, e.g., [20, Sec. 2.2] and references therein. Algorithmically, (P1) is tackled either by generic SOCP solvers, or, by interior-point algorithms customized to its specific form [20].

Having relaxed the RS problem (P0) to its closest convex approximation (P1) which is tractable, it is of interest to reflect on various links and interpretations that (P1) can afford, postponing its performance analysis to Section IV.

*Remark 1 ((P1) Versus LS):* Clearly, the LS problem in (1) can be rewritten as

$$\min_{\mathbf{x}, \mathbf{t}} \{ \|\mathbf{t}\|_2 : \|\mathbf{b}_i - \mathbf{A}_i \mathbf{x}\|_2 \leq t_i, \quad i = 1, \dots, k \}$$

which is again a convex approximation of (P0), though, as mentioned earlier, not the closest one.

*Remark 2 ((P1) Versus Block-sparse Signal Reconstruction):* To establish this connection, assume that  $\text{null}(\mathbf{A}^T)$  is nonempty. Let  $\mathbf{r}_i := \mathbf{b}_i - \mathbf{A}_i \mathbf{x}$  denote the residual error vectors and  $\mathbf{r}^T := [\mathbf{r}_1^T \cdots \mathbf{r}_k^T]$ . Upon defining matrix  $\mathbf{C}$  such that its null space is spanned by  $\text{range}(\mathbf{A})$ , i.e.,  $\mathbf{C}\mathbf{A} = \mathbf{0}$  and  $\mathbf{d} := \mathbf{C}\mathbf{b}$ , the problem (P1) can be rendered equivalent to

$$\min_{\mathbf{r}} \sum_{i=1}^k \|\mathbf{r}_i\|_2 \quad (4a)$$

$$\text{s.t. } \mathbf{C}\mathbf{r} = \mathbf{d} \quad (4b)$$

which emerges when reconstructing a *block-sparse* vector  $\mathbf{r}$  satisfying the underdetermined system in (4b) [5], [11], [26], [27]. To establish the equivalence, write (P1) as  $\min_{\mathbf{r}} \sum_{i=1}^k \|\mathbf{r}_i\|_2$  subject to  $\mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{x}$ . Premultiplying both sides of the last equality by  $\mathbf{C}$ , one arrives at (4). The same equality couples the minimizers of the two problems: if  $\mathbf{r}_0$  solves (4) and  $\mathbf{A}^\dagger$  is the pseudo-inverse of  $\mathbf{A}$ , then  $\mathbf{A}^\dagger(\mathbf{b} - \mathbf{r}_0)$  solves (P1). The optimization in (4) relies on the prior information that  $\mathbf{r}$  is block sparse. For the RS problem, the vector of interest  $\mathbf{x}$  is not (block) sparse; but the residual error vector is block sparse.

*Remark 3 ((P1) Versus  $\ell_1$ -error Regression):* In the degenerate case  $m = 1$ , where every subsystem reduces to a single equation, (P1) reduces to the  $\ell_1$ -error minimization problem (2), which is known to be robust to outliers [22, Ch. 4], [6], [8]. Under the conditions stated in Remark 2, the unconstrained  $\ell_1$ -error regression problem is equivalent to the constrained optimization (cf. (4))

$$\min_{\mathbf{r}} \|\mathbf{r}\|_1 \quad (5)$$

$$\text{s.t. } \mathbf{C}\mathbf{r} = \mathbf{d}.$$

The problem in (5) is widely known in the CS literature as basis pursuit (BP); for a thorough treatment on this pair of problems see also [8].

### B. A Concave Surrogate for RS

Instead of substituting the cost  $\|\mathbf{t}\|_0$  of (P0) by its closest convex approximation, namely  $\|\mathbf{t}\|_1$ , letting the surrogate function be nonconvex can yield tighter approximations. For example, the  $\ell_0$ -norm of a vector  $\mathbf{x} \in \mathbb{R}^n$  was surrogated in [7] by the logarithm of the geometric mean of its elements, or, by  $\sum_{i=1}^n \log |x_i|$ . In rank minimization problems, apart from the nuclear norm relaxation, minimizing the logarithm of the determinant of the unknown matrix has been proposed as an alternative surrogate; see [12, Sec. 5.2]. Building on this line of thought, consider surrogating (P0) by

$$\boxed{\begin{array}{ll} \min_{\mathbf{x}, \mathbf{t}} & \sum_{i=1}^k \log(t_i + \delta) \\ \text{s.t.} & \|\mathbf{b}_i - \mathbf{A}_i \mathbf{x}\|_2 \leq t_i, \quad i = 1, \dots, k \end{array}} \quad (P2)$$

where  $\delta > 0$  is a sufficiently small constant preventing the cost from tending to  $-\infty$ . The cost in (6) is concave, but since it is smooth w.r.t.  $\mathbf{t} \in \mathbb{R}_+^k$ , iterative linearization may be utilized to obtain a local minimum [7], [12]. Specifically, let  $(\mathbf{x}^{(l)}, \mathbf{t}^{(l)})$  de-

note a tentative solution at the  $l$ th iteration. Due to the concavity of the logarithm, the first-order approximation of  $\log(t_i + \delta)$  around  $t_i^{(l-1)} + \delta$  yields

$$\log(t_i + \delta) \leq \log(t_i^{(0)} + \delta) + \frac{1}{t_i^{(0)} + \delta} (t_i - t_i^{(0)}). \quad (6)$$

Thinking along the majorization-minimization approach [18], one can instead of minimizing the original cost on the left-hand side, minimize the majorizing cost on the right-hand side of (6) and iterate. Specifically, the minimization in (P2) can be iteratively driven to a local minimum [12] as

$$(\mathbf{x}^{(l)}, \mathbf{t}^{(l)}) := \arg \min_{\mathbf{x}, \mathbf{t}} \sum_{i=1}^k \frac{t_i}{t_i^{(l-1)} + \delta}$$

$$\text{s.t. } \|\mathbf{b}_i - \mathbf{A}_i \mathbf{x}\|_2 \leq t_i$$

for  $i = 1, \dots, k$ , or equivalently,

$$\mathbf{x}^{(l)} := \arg \min_{\mathbf{x}} \sum_{i=1}^k \frac{\|\mathbf{b}_i - \mathbf{A}_i \mathbf{x}\|_2}{\|\mathbf{b}_i - \mathbf{A}_i \mathbf{x}^{(l-1)}\|_2 + \delta}. \quad (7)$$

The iterative scheme can be terminated as soon as the relative error  $\|\mathbf{x}^{(l)} - \mathbf{x}^{(l-1)}\|_2 / \|\mathbf{x}^{(l-1)}\|_2$  becomes smaller than some  $\epsilon$  chosen equal to say  $10^{-6}$ . The cost in (7) has the form of a weighted version of (P1), where each of the error norms is weighted by  $w_i^{(l)} = (\|\mathbf{b}_i - \mathbf{A}_i \mathbf{x}^{(l-1)}\|_2 + \delta)^{-1}$ . When the residual error of a subsystem is small, then the error of this system is weighted more during the minimization of the next iteration. A good initialization point for the iteration in (7) is the solution of (P1) that is equivalent to one iteration of (7) with all weights chosen equal. The simulated tests in Section VI will indicate that (7) can provide higher probability of identifying reliable sensors than (P1).

## IV. UNIQUENESS AND IDENTIFIABILITY

Let  $s$  denote the minimum cost of (P0). Then, there exists at least one unknown  $\mathbf{x}_0 \in \mathbb{R}^n$  such that  $\mathbf{b}_{\mathcal{S}_0} = \mathbf{A}_{\mathcal{S}_0} \mathbf{x}_0$  for an unknown subset of sensors  $\mathcal{S}_0$  with  $|\mathcal{S}_0| = s$ . The sensors in  $\mathcal{S}_0$  will be referred to as *reliable* or *consistent* with respect to (w.r.t.)  $\mathbf{x}_0$ . Also, let  $\beta := s/k$  denote the number of consistent sensors over the total number of sensors; and  $\gamma := \frac{n}{km}$  the ratio of the size of the unknown vector over the total number of measurements.

Whether (P0) has a unique minimizer and hence an underlying  $\mathbf{x}_0$  can be uniquely recovered by (P0), is considered next. The first thing to note at the outset is that when the consistent sensors w.r.t.  $\mathbf{x}_0$  are outnumbered by the unreliable ones, uniquely recovering  $\mathbf{x}_0$  is not guaranteed. This is because with  $s \leq \frac{k}{2}$ , there may exist an  $\mathbf{x}_1 \neq \mathbf{x}_0$  and an  $\mathcal{S}_1 \subset \mathcal{I}$  with  $|\mathcal{S}_1| = |\mathcal{S}_0| = s$  and  $\mathcal{S}_1 \cap \mathcal{S}_0 = \emptyset$  such that  $\mathbf{b}_{\mathcal{S}_1} = \mathbf{A}_{\mathcal{S}_1} \mathbf{x}_1$ ; thus,  $\mathbf{x}_0$  and  $\mathbf{x}_1$  are both minimizers of (P0). It is henceforth assumed that  $s > \frac{k}{2}$  or  $\beta \in (\frac{1}{2}, 1]$ . Under this assumption, uniqueness of the (P0) minimizer is further characterized in the following lemma.

**Lemma 1:** *Let vector  $\mathbf{x}_0$  be a minimizer of (P0) satisfying  $s > k/2$  out of the  $k$  subsystems. This minimizer is unique if and only if*

$$\text{rank}(\mathbf{A}_{\mathcal{S}_c}) = n \quad (8)$$

for every  $\mathcal{S}_c \subset \mathcal{I}$  with cardinality  $|\mathcal{S}_c| = 2s - k$ .

*Proof:* Vector  $\mathbf{x}_0$  is not the unique minimizer of (P0) if and only if there exists at least one  $\mathbf{x}_1 \neq \mathbf{x}_0$  such that  $\mathbf{b}_{\mathcal{S}_1} = \mathbf{A}_{\mathcal{S}_1} \mathbf{x}_1$  for an  $\mathcal{S}_1 \subset \mathcal{I}$  with  $|\mathcal{S}_1| = |\mathcal{S}_0| = s$ . Given that  $s > \frac{k}{2}$ , the two subsets cannot be disjoint; hence, they must have a nonempty intersection  $\mathcal{S}_c := \mathcal{S}_0 \cap \mathcal{S}_1$  with cardinality  $2s - k \leq |\mathcal{S}_c| \leq s$ . The subsystems belonging to  $\mathcal{S}_c$  are satisfied by both solutions; that is,  $\mathbf{b}_{\mathcal{S}_c} = \mathbf{A}_{\mathcal{S}_c} \mathbf{x}_0 = \mathbf{A}_{\mathcal{S}_c} \mathbf{x}_1$ , which is equivalent to the existence of a nonzero  $\mathbf{z} \in \mathbb{R}^n$  such that  $\mathbf{A}_{\mathcal{S}_c} \mathbf{z} = \mathbf{0}$  or  $\text{rank}(\mathbf{A}_{\mathcal{S}_c}) < n$ . Multiple minimizers of (P0) can thus be avoided if and only if  $\text{rank}(\mathbf{A}_{\mathcal{S}_c}) = n$ . Note that whenever  $\text{rank}(\mathbf{A}_{\mathcal{S}_c}) = n$  for every  $\mathcal{S}_c$  with  $|\mathcal{S}_c| = 2s - k$ , it holds for every  $\mathcal{S}_c$  of larger cardinality as well. ■

Lemma 1 reveals two interesting points on uniquely recovering  $\mathbf{x}_0$  by (P0). First, the reliable sensors should not only outnumber the unreliable ones, i.e.,  $\beta > \frac{1}{2}$ ; condition (8) implies additionally that  $(2s-k)m \geq n$ , or  $\beta \geq \frac{(\gamma+1)}{2}$ . Second, because  $\beta \leq 1$ , the inequality  $\beta \geq \frac{(\gamma+1)}{2}$  implies  $\gamma \leq 1$  or  $km \geq n$ , requiring the total number of equations to be at least equal to the number of unknowns.

Uniqueness of the (P0) minimizer is also implied by the conditions stated in the next lemma. These conditions will be used in the next subsection.

**Lemma 2:** *If for any nonzero  $\mathbf{v} \in \text{range}(\mathbf{A})$  and any partition  $(\mathcal{S}, \bar{\mathcal{S}})$  of  $\mathcal{I}$  with  $|\mathcal{S}| = s > \frac{k}{2}$  it holds that*

$$\sum_{i \in \mathcal{S}} \|\mathbf{v}_i\|_2 > \sum_{i \in \bar{\mathcal{S}}} \|\mathbf{v}_i\|_2 \quad (9)$$

where  $\mathbf{v}_i$  is the  $i$ th  $m \times 1$  block subvector of  $\mathbf{v}$ , then (8) is satisfied.

*Proof:* Arguing by contradiction, suppose that (9) holds, whereas (8) does not hold; or, in other words there exists an  $\mathcal{S}_c \subset \mathcal{I}$  with  $|\mathcal{S}_c| = 2s - k \leq s$  and  $\text{rank}(\mathbf{A}_{\mathcal{S}_c}) < n$ . Consequently, there exists a nonzero vector  $\mathbf{u} \in \mathbb{R}^n$  such that  $\mathbf{A}_{\mathcal{S}_c} \mathbf{u} = \mathbf{0}$ . Next, partition  $\mathcal{I}$  into three collectively exhaustive and mutually exclusive subsets  $\mathcal{S}_c, \mathcal{S}_1$  and  $\mathcal{S}_2$ , with  $|\mathcal{S}_1| = |\mathcal{S}_2| = k - s$ . Define also  $\mathbf{v} := \mathbf{A} \mathbf{u}$  for which  $\mathbf{v}_{\mathcal{S}_c} = \mathbf{0}$  by the definition of  $\mathbf{u}$ .

Consider first (9) with  $\mathcal{S} = \mathcal{S}_c \cup \mathcal{S}_1$  and  $\bar{\mathcal{S}} = \mathcal{S}_2$ , to get

$$\sum_{i \in \mathcal{S}_1} \|\mathbf{v}_i\|_2 > \sum_{i \in \mathcal{S}_2} \|\mathbf{v}_i\|_2.$$

Apply (9) again for  $\mathcal{S} = \mathcal{S}_c \cup \mathcal{S}_2$  and  $\bar{\mathcal{S}} = \mathcal{S}_1$ , to arrive at

$$\sum_{i \in \mathcal{S}_2} \|\mathbf{v}_i\|_2 > \sum_{i \in \mathcal{S}_1} \|\mathbf{v}_i\|_2$$

which clearly contradicts the previous inequality and completes the proof. ■

Having introduced the convex relaxation (P1) of (P0), the next critical question is whether the solution of the former coincides with the solution of the latter. Even though the NP-hardness of (P0) forejudges that this cannot hold in general, the ensuing results show that for random Gaussian matrices  $\mathbf{A}$  and under reasonable assumptions on the problem dimensions, equivalence of (P1) and (P0) occurs with probability exponentially decaying in  $n$ . The analysis starts by characterizing this equivalence using a set of necessary and sufficient conditions.

### A. Necessary and Sufficient Conditions

The conditions under which the convex optimization problem (P1) yields the same solution as the NP-hard problem (P0) are provided in the following theorem. Using the equivalence between (P1) and (4) under the conditions of Remark 2, this theorem is related to [27, Th. 2], which in turn, generalizes results from [10] to the block-sparse signal case.

**Theorem 1 (Range Space Conditions):** *Every  $\mathbf{x}_0$  minimizing (P0) by satisfying  $s > \frac{k}{2}$  out of the  $k$  subsystems is the unique minimizer of (P1) if and only if*

$$\sum_{i \in \mathcal{S}} \|\mathbf{v}_i\|_2 > \sum_{i \in \bar{\mathcal{S}}} \|\mathbf{v}_i\|_2 \quad (10)$$

for any nonzero  $\mathbf{v} \in \text{range}(\mathbf{A})$  and for any partition  $(\mathcal{S}, \bar{\mathcal{S}})$  of  $\mathcal{I}$  with  $|\mathcal{S}| = s$ .

*Proof:* See the Appendix. ■

In words, Theorem 1 asserts that for every nonzero  $\mathbf{v} \in \text{range}(\mathbf{A})$ , the sum of the  $s$  smallest  $\|\mathbf{v}_i\|_2$  components should be larger than the sum of the remaining  $(k - s)$  components. It is worth mentioning that the range space conditions are impossible to check in practice; but they are useful in establishing identifiability, as it will be the case for the probabilistic characterization of the (P0)–(P1) equivalence when  $\mathbf{A}$  is random (cf. Section IV-C).

Another set of (P0)–(P1) equivalence conditions can be derived from the block restricted isometry properties of matrix  $\mathbf{C}$  as defined in Remark 2; see [5], [11]. However, these conditions are only sufficient.

**Remark 4:** Conditions (10) do not depend on  $\mathbf{b}$ , but only on  $\text{range}(\mathbf{A})$ . Thus, whenever  $\mathbf{A}$  satisfies (10), any matrix  $\mathbf{A}' := \mathbf{A} \mathbf{G}$  for any nonsingular  $\mathbf{G} \in \mathbb{R}^{n \times n}$  satisfies (10) as well.

**Remark 5:** Sufficiency of the conditions in (10) remains valid even if some additional constraints of the generic form  $\mathbf{x} \in \mathcal{C}$  are present in the original problem (P0). In certain applications for instance, the unknown  $\mathbf{x}$  may be non-negative so that  $\mathcal{C} = \mathbb{R}_+^n$ ; or, there may be *a priori* information of the form  $\mathcal{C} = \{\mathbf{x} : \|\mathbf{x} - \mathbf{x}_c\|_2 \leq R\}$ , dictating the unknown vector to lie in a ball of radius  $R$  around a known center  $\mathbf{x}_c \in \mathbb{R}^n$ . Even though the extra constraints generally reduce the feasible sets of (P0) and (P1), the conditions remain sufficient. Hence, the probabilistic bound to be developed in Section IV-B remains valid even when extra constraints are imposed.

### B. Probability Bound

As commented earlier, the conditions in Lemma 2 are practically infeasible to check for a given sensing matrix  $\mathbf{A}$ . However, similar to CS [8], it will be possible to prove that the conditions in (9) hold with overwhelming probability [8], i.e., probability decaying exponentially in  $n$  when  $\gamma$  and  $k$  are fixed, assuming  $\mathbf{A}$  has i.i.d. Gaussian entries. The main result, summarized in Theorem 2, is based on the following lemma.

**Lemma 3 (Deviation Inequality):** *Consider  $\mathbf{x} \sim \mathcal{N}(\mathbf{0}_p, \mathbf{I}_p)$  and a Lipschitz continuous function  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  with Lipschitz constant  $L$ . Then for any  $t \geq 0$ , it holds that*

$$\Pr(f(\mathbf{x}) - \mathbb{E}[f(\mathbf{x})] \leq -t) \leq \exp\left(-\frac{t^2}{2L^2}\right). \quad (11)$$

This deviation inequality is a special case of more general concentration results [19, Sec. 1.1]. It provides exponentially decreasing bounds on the tail distribution for any sufficiently smooth function  $f(\mathbf{x})$  of a multivariate Gaussian  $\mathbf{x}$ , thus generalizing the Chernoff bound to nonlinear functions.

Capitalizing on Lemma 3, the next theorem extends the results of [27, Th. 4] and its refined version [26, Th. 3]. Focusing on the Gaussian case and following a different line of proof, neat closed-form expressions will emerge not only for the values of  $\beta$  and  $\gamma$ , for which the probabilistic bound is valid, but also for the bound itself. The proof is based partly on the methodology of [25], where the minimum nuclear norm relaxation of the rank minimization problem is analyzed under linear constraints on the unknown matrix. In contrast, related probabilistic analysis in [11] and [5] is based on a generalization of the restricted isometry property of  $\mathbf{A}$  that serves only as a sufficient condition for the exactness of the convex relaxation; see also [8].

**Theorem 2:** *Let vector  $\mathbf{x}_0$  be a minimizer of (P0) satisfying  $s > \frac{k}{2}$  out of the  $k$  subsystems and assume that the entries of  $\mathbf{A} \in \mathbb{R}^{km \times n}$  are independently drawn from  $\mathcal{N}(0, 1)$ . If*

$$\beta > \frac{\sqrt{\gamma} + 1}{2} \quad (12)$$

then whenever  $m \geq \frac{\beta \log(\frac{e}{\beta})}{(1-\alpha)c_0(\beta, \gamma)\gamma}$ , the vector  $\mathbf{x}_0$  is the unique minimizer of (P1) with probability exceeding  $1 - e^{-\alpha c_0(\beta, \gamma)n + o_n(n)}$ , where  $c_0(\beta, \gamma) := \frac{1}{2} \left( \frac{2\beta-1}{\sqrt{\gamma}} - 1 \right)^2$  and  $\alpha \in (0, 1)$ .

*Proof:* To lower bound the probability of success for the (P1) problem, it suffices to upper bound the probability that the conditions in (10) fail, an event denoted by  $\mathcal{E}$ . Let  $\{\mathcal{S}_j\}$  be all the  $N := \binom{k}{s}$  subsets of  $\mathcal{I}$  having cardinality  $s$ . Moreover, let  $\mathcal{E}_j$  denote the event of having the conditions in (10) failing for the partition  $(\mathcal{S}_j, \bar{\mathcal{S}}_j)$

$$\mathcal{E}_j := \left\{ \exists \mathbf{v} \in \text{range}(\mathbf{A}) \setminus \{0\} : \sum_{i \in \mathcal{S}_j} \|\mathbf{v}_i\|_2 \leq \sum_{i \in \bar{\mathcal{S}}_j} \|\mathbf{v}_i\|_2 \right\} \quad (13)$$

for  $j = 1, \dots, N$ . The probability of failure can be expressed as  $\Pr(\mathcal{E}) = \Pr\left(\bigcup_j \mathcal{E}_j\right)$ . The events  $\{\mathcal{E}_j\}_{j=1}^N$  are not independent, but  $\Pr(\mathcal{E})$  can be bounded as

$$\Pr(\mathcal{E}) \stackrel{(a)}{\leq} \sum_{j=1}^N \Pr(\mathcal{E}_j) \stackrel{(b)}{=} \binom{k}{s} \Pr(\mathcal{E}_j) \stackrel{(c)}{\leq} e^{s(1-\log \beta)} \Pr(\mathcal{E}_j) \quad (14)$$

where inequality (a) comes from the union bound; (b) is due to the symmetry of the distribution of  $\mathbf{A}$  which implies that all the  $\mathcal{E}_j$ 's are equiprobable; and (c) is the standard upper bound of the binomial coefficient  $\binom{k}{s} \leq \left(\frac{ke}{s}\right)^s$ . Based on (14), the goal now is to upper bound the probability  $\Pr(\mathcal{E}_j)$ . For notational simplicity, the partition corresponding to  $\mathcal{E}_j$  will be denoted by  $(\mathcal{S}, \bar{\mathcal{S}})$  instead of  $(\mathcal{S}_j, \bar{\mathcal{S}}_j)$ .

Given that  $\mathbf{v} \in \text{range}(\mathbf{A}) \setminus \{0\}$ , there exists a nonzero  $\mathbf{u} \in \mathbb{R}^n$  such that  $\mathbf{v}_i = \mathbf{A}_i \mathbf{u}$  for  $i = 1, \dots, k$ . To render the in-

equality in (13) scale-invariant, one can study only the cases for which  $\|\mathbf{u}\|_2 = 1$ ; hence,  $\Pr(\mathcal{E}_j)$  is equal to

$$\Pr\left(\exists \mathbf{u} \text{ with } \|\mathbf{u}\|_2 = 1 : \sum_{i \in \mathcal{S}} \|\mathbf{A}_i \mathbf{u}\|_2 \leq \sum_{i \in \bar{\mathcal{S}}} \|\mathbf{A}_i \mathbf{u}\|_2\right)$$

which is also equal to  $\Pr(f(\mathbf{A}) \leq 0)$  for

$$f(\mathbf{A}) := \inf_{\|\mathbf{u}\|_2=1} \left\{ \sum_{i \in \mathcal{S}} \|\mathbf{A}_i \mathbf{u}\|_2 - \sum_{i \in \bar{\mathcal{S}}} \|\mathbf{A}_i \mathbf{u}\|_2 \right\} \quad (15)$$

due to the fact that if there exists a unit  $\ell_2$ -norm  $\mathbf{u}$  satisfying  $\sum_{i \in \mathcal{S}} \|\mathbf{A}_i \mathbf{u}\|_2 \leq \sum_{i \in \bar{\mathcal{S}}} \|\mathbf{A}_i \mathbf{u}\|_2$ , then the minimum value of (15) should be nonpositive. The function  $f(\mathbf{A})$  possesses convenient properties which facilitate the application of Lemma 3. Specifically, it is shown in the Appendix that  $f(\mathbf{A})$  is Lipschitz continuous with constant  $L \leq \sqrt{k}$  (cf. Lemma 4); and the expected value of the function is lower bounded (cf. Lemma 5), that is  $\mathbb{E}[f(\mathbf{A})] \geq \mu = \left(\frac{2\beta-1}{\sqrt{\gamma}} - 1\right) \sqrt{kn} (1 + o_n(1))$ . Hence, for every  $t \geq 0$ , Lemma 3 implies that

$$\Pr\left(f(\mathbf{A}) \leq \mu - t\right) \leq \Pr\left(f(\mathbf{A}) \leq \mathbb{E}[f(\mathbf{A})] - t\right) \leq e^{-\frac{t^2}{2k}}. \quad (16)$$

Upon focusing on  $\mu$  and ignoring the  $o_n(1)$  term, whenever  $\beta > \frac{(\sqrt{\gamma}+1)}{2}$  so that  $\mu > 0$  and setting  $t = \mu$  in (16), yields the bound

$$\Pr(\mathcal{E}_j) \leq \Pr\left(f(\mathbf{A}) \leq 0\right) \leq e^{-c_0(\beta, \gamma)n + o_n(n)} \quad (17)$$

where  $c_0(\beta, \gamma) := \frac{(2\beta-1-1)^2}{2}$ .

Substituting the bound (17) into (14), it follows that

$$\begin{aligned} \Pr(\mathcal{E}) &\leq \exp\left(-\left(c_0(\beta, \gamma) - \frac{s \log\left(\frac{e}{\beta}\right)}{n}\right)n + o_n(n)\right) \\ &\leq \exp(-c_1(\beta, \gamma)n + o_n(n)). \end{aligned} \quad (18)$$

For every  $\beta > \frac{(\sqrt{\gamma}+1)}{2}$ , choose  $c_1(\beta, \gamma) = \alpha c_0(\beta, \gamma)$  and define  $c_2(\beta, \gamma) := ((1-\alpha)c_0(\beta, \gamma))^{-1}$  for any  $\alpha \in (0, 1)$ . Then, whenever  $m \geq c_2(\beta, \gamma) \frac{\beta}{\gamma} \log\left(\frac{e}{\beta}\right)$ , the bound in (18) is non-trivial. ■

*Remark 6:* As a sanity test, the condition  $\beta > \frac{\sqrt{\gamma}+1}{2}$  posed by Theorem 2 coincides with that in [26, Th. 3] after the appropriate mapping of dimensions. However, in Theorem 2, both the values of  $m$  over which the bound holds, as well as the bound itself are explicitly defined.

*Remark 7:* As expected, the condition  $\beta > \frac{\sqrt{\gamma}+1}{2}$  is clearly stronger than the condition  $\beta > \frac{\gamma+1}{2}$  implied by the uniqueness of the (P0) solution in Lemma 1.

## V. ROBUSTNESS TO NOISE

In a more realistic sensing scenario, the acquired measurements are corrupted by additive noise. If  $\mathcal{S}_0$  denotes the unknown subset of reliable sensors, the pertinent model is

$$\mathbf{b}_{\mathcal{S}_0} = \mathbf{A}_{\mathcal{S}_0} \mathbf{x}_0 + \mathbf{n}_{\mathcal{S}_0} \quad (19)$$

where  $\mathbf{n}_{S_0}$  stands for zero-mean noise assumed independent across sensors. Vector  $\mathbf{n}_{S_0}$  models ambient noise, finite precision, analog-to-digital conversion and quantization effects, communication noise, or even, the inadequacy of linear regression to fully capture the measured data  $\mathbf{b}_{S_0}$ .

In this noisy case, the unknown  $\mathbf{x}_0$  does not exactly satisfy the linear subsystems in  $S_0$ . In an attempt to exploit the link between (P1) and (4) when noise is present, one may be tempted to apply the group-Lasso regularization, which was originally proposed for recovering block-sparse vectors in a linear regression setup [32]. However, this approach is not applicable because  $\mathbf{r}$  is not block sparse when noise is present. In fact, solvers of the noise-free setups (P1) and (P2) are useful for analyzing uniqueness and identifiability issues. In addition, (P1) and (P2) solvers are practically suitable for high-SNR sensing applications. This motivates the ensuing framework which is suitable for RS in the presence of noise. Without additional prior information on the model describing the unreliable sensors, the noisy counterpart of the RS problem can be stated as follows.

**Problem Statement 2 (Robust Sensing in Noise (RSN)):**  
 Given  $\{\mathbf{b}_i, \mathbf{A}_i\}_{i \in \mathcal{I}}$  where  $\mathbf{b}_i \in \mathbb{R}^m$  and  $\mathbf{A}_i \in \mathbb{R}^{m \times n}$ , for which an unknown subset  $S_0 \subset \mathcal{I}$  of known cardinality  $s$  follows the model in (19), estimate the unknown  $\mathbf{x}_0$  by minimizing the least-squares error over any  $S \subset \mathcal{I}$  with  $|S| = s$ .

The aforementioned problem statement lends itself naturally to the following optimization problem:

$$\min_{\mathbf{x}} \min_{|S|=s} \|\mathbf{b}_S - \mathbf{A}_S \mathbf{x}\|_2^2. \quad (20)$$

The function of  $\mathbf{x}$  defined by the inner minimization is the pointwise minimum over finitely many convex functions and as such, it is nonconvex. Solving (20) incurs combinatorial complexity since one has to solve all the  $\binom{k}{s}$  LS problems before solving the outer minimization.

An optimization problem related to that in (20) is the following:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \sum_{i=1}^k h(\mathbf{b}_i - \mathbf{A}_i \mathbf{x}) \quad (21a) \\ \text{s.t.} \quad & h(\mathbf{r}_i) := \begin{cases} \frac{1}{2} \|\mathbf{r}_i\|_2^2, & \|\mathbf{r}_i\|_2 \leq \lambda \\ \frac{1}{2} \lambda^2, & \|\mathbf{r}_i\|_2 > \lambda \end{cases}, \lambda \geq 0. \quad (21b) \end{aligned}$$

Functional  $h(\mathbf{r}_i)$  amounts to the LS cost for residuals smaller than the threshold  $\lambda$  and ignores sensors attaining larger residuals. In the scalar case (cf.  $m = 1$ ), problem (21) has been considered in [6, Sec. 6.1.2]. Problems (20) and (21) are related as follows: suppose that for a specific  $\lambda$  the solution of (21) is  $\mathbf{x}^*$  for which there are  $s^*$  residuals satisfying the upper branch of (21b). Then it can be readily shown that  $\mathbf{x}^*$  is a solution of (20) for  $s = s^*$ . Unfortunately though,  $h(\mathbf{r}_i)$  is nonconvex as well. The problem in (21) can be surrogated by replacing  $h(\mathbf{r}_i)$  by its closest convex approximation, which is pursued in the next subsection by establishing a neat link between the RSN problem at hand and robust estimation methods [17, Ch. 7], [22, Ch. 4].

### A. RSN and Robust Linear Multivariate Regression

Building on Remark 3 of Section III-A, the unreliable sensors can be viewed as giving rise to outlier-corrupted equations in a linear regression setting. Robust linear regression has been extensively studied over the past decades [17], [22]. When  $m = 1$ , the RSN problem can be solved by Huber's M-estimator

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \sum_{i=1}^k \rho(b_i - \mathbf{a}_i^T \mathbf{x}) \quad (22a)$$

$$\text{where } \rho(r) := \begin{cases} \frac{1}{2} r^2, & |r| \leq \tau \\ \tau |r| - \frac{\tau^2}{2}, & |r| > \tau \end{cases} \quad (22b)$$

is the Huber function for  $\tau > 0$ . The problem in (22) is convex and can be cast as an SOCP [6, p. 190], [15], [21]. Regarding the cutoff parameter  $\tau$ , when the outliers' distribution is known a priori, its value is available in closed form so that Huber's M-estimator is asymptotically optimal; see [17, Sec. 4.5]. Alternatively, assuming that the noise is standard Gaussian,  $\tau$  is usually set to  $\tau = 1.34$  such that the estimator in (22) is 95% asymptotically efficient at the normal distribution [22, p. 26]. To render Huber's M-estimator invariant to any noise variance  $\sigma^2$ , one has to multiply  $\tau$  by  $\sigma$  in (22b). If  $\sigma$  is unknown, a robust estimate of it is commonly used instead [22, Sec. 4.4].

The case  $m > 1$ , which is of interest here falls under the realm of robust multivariate linear regression [2], [4]. The novel approach to tackle it will be to postulate a model accommodating inconsistent sensors, approximate the meaningful cost of (21) by a convex one and solve it using an efficient globally convergent algorithm.

Consider modeling the unreliable sensors using the auxiliary outlier vectors  $\{\mathbf{u}_i \in \mathbb{R}^m\}_{i=1}^k$ . Vector  $\mathbf{u}_i = \mathbf{0}$  if the  $i$ th sensor is reliable; and  $\mathbf{u}_i \neq \mathbf{0}$  deterministically, otherwise. Model (19) can now be extended to incorporate the unreliable sensors as

$$\mathbf{b}_i = \mathbf{A}_i \mathbf{x} + \mathbf{u}_i + \mathbf{n}_i, \quad i = 1, \dots, k. \quad (23)$$

Since some  $\mathbf{u}_i$ 's are zero, the aggregate outlier vector  $\mathbf{u}^T := [\mathbf{u}_1^T \cdots \mathbf{u}_k^T]$  is block sparse. Hence, using the aggregate model  $\mathbf{b} = \mathbf{A} \mathbf{x} + \mathbf{u} + \mathbf{n}$ , the novel RSN solver amounts to

$$\boxed{\min_{\mathbf{x}, \mathbf{u}} \frac{1}{2} \|\mathbf{b} - \mathbf{A} \mathbf{x} - \mathbf{u}\|_2^2 + \lambda \sum_{i=1}^k \|\mathbf{u}_i\|_2} \quad (P3)$$

where  $\lambda > 0$  is an appropriately chosen tuning parameter. Among the two optimization variables of (P3), only the outlier vector  $\mathbf{u}$  is block sparse. For  $m = 1$ , (P3) reduces to the cost proposed in [15] and shown to be equivalent to (22). Even when the initial matrix of interest  $\mathbf{A}$  is tall, (P3) always entails the fat matrix  $[\mathbf{A} \quad \mathbf{I}_{km}] \in \mathbb{R}^{km \times (n+km)}$ . The second part is a regularization term, reminiscent of the group Lasso penalty function [32], which is known to promote block sparsity in the  $\mathbf{u}$  vector. The latter will be explicitly accounted for in the forthcoming analysis.

### B. Solving (P3)

To better understand (P3) and develop an efficient solver, it is prudent to explore the form of its minimizer(s). Let  $[(\mathbf{x}^*)^T (\mathbf{u}^*)^T]^T$  denote a minimizer of (P3) and define the associated residual vector  $\mathbf{r}^* := \mathbf{b} - \mathbf{A}\mathbf{x}^*$ . Given  $\mathbf{x}^*$ , the vectors  $\{\mathbf{u}_i^*\}_{i=1}^k$  in (P3) can be found separately as the minimizers of

$$\begin{aligned} \min_{\mathbf{u}_i} \quad & \phi(\mathbf{u}_i) \\ \text{s.t.} \quad & \phi(\mathbf{u}_i) := \frac{1}{2} \|\mathbf{r}_i^* - \mathbf{u}_i\|_2^2 + \lambda \|\mathbf{u}_i\|_2, \quad i = 1, \dots, k. \end{aligned} \quad (24)$$

Although  $\phi(\mathbf{u}_i)$  is not everywhere differentiable, its subdifferential  $\partial\phi(\mathbf{u}_i)$  can be defined [6]. For  $\mathbf{u}_i \neq \mathbf{0}$ , where  $\phi(\mathbf{u}_i)$  is differentiable, the subdifferential is simply  $\mathbf{u}_i \left(1 + \frac{\lambda}{\|\mathbf{u}_i\|_2}\right) - \mathbf{r}_i^*$ . Otherwise, by definition and after using (34),  $\partial\phi(\mathbf{u}_i)$  can be shown to be the set  $\{\lambda\mathbf{g}_i - \mathbf{r}_i^* \mid \|\mathbf{g}_i\|_2 \leq 1\}$ . Compactly,

$$\partial\phi(\mathbf{u}_i) := \begin{cases} \mathbf{u}_i \left(1 + \frac{\lambda}{\|\mathbf{u}_i\|_2}\right) - \mathbf{r}_i^*, & \mathbf{u}_i \neq \mathbf{0} \\ \{\lambda\mathbf{g}_i - \mathbf{r}_i^* \mid \|\mathbf{g}_i\|_2 \leq 1\}, & \mathbf{u}_i = \mathbf{0}. \end{cases} \quad (25)$$

Vector  $\mathbf{u}_i^*$  is a minimizer of (24) if and only if  $\mathbf{0} \in \partial\phi(\mathbf{u}_i^*)$ . Based on (25), two cases are considered.

First, if  $\mathbf{u}_i^* \neq \mathbf{0}$ , the condition  $\mathbf{0} \in \partial\phi(\mathbf{u}_i^*)$  yields

$$\mathbf{u}_i^* \left(1 + \frac{\lambda}{\|\mathbf{u}_i^*\|_2}\right) = \mathbf{r}_i^* \quad (26)$$

which means that  $\mathbf{u}_i^*$  is a positively scaled version of  $\mathbf{r}_i^*$ . Considering the  $\ell_2$ -norm in both sides of (26), it follows that  $\|\mathbf{u}_i^*\|_2 = \|\mathbf{r}_i^*\|_2 - \lambda$ . Plugging  $\|\mathbf{u}_i^*\|_2$  back into (26), yields  $\mathbf{u}_i^* = \mathbf{r}_i^* \left(1 - \frac{\lambda}{\|\mathbf{r}_i^*\|_2}\right)$ . Since  $\|\mathbf{u}_i^*\|_2 > 0$ , this holds if and only if  $\|\mathbf{r}_i^*\|_2 > \lambda$ .

Second, for the minimizer to be  $\mathbf{u}_i^* = \mathbf{0}$ , there should be a  $\mathbf{g}_i^*$  for which  $\|\mathbf{g}_i^*\|_2 \leq 1$  and  $\lambda\mathbf{g}_i^* = \mathbf{r}_i^*$ , or equivalently,  $\|\mathbf{r}_i^*\|_2 \leq \lambda$ . The latter proves that (P3) indeed admits a block-sparse minimizer  $\mathbf{u}^*$ .

Substituting  $\mathbf{u}_i^*$  into (24), yields  $\phi(\mathbf{u}_i^*) = \frac{\|\mathbf{r}_i^*\|_2^2}{2}$ , when  $\|\mathbf{r}_i^*\|_2 \leq \lambda$ ; and  $\phi(\mathbf{u}_i^*) = \lambda\|\mathbf{r}_i^*\|_2 - \frac{\lambda^2}{2}$ , otherwise. Having minimized (P3) over the  $\mathbf{u}_i^*$ 's, the minimizer  $\mathbf{x}^*$  can now be found as

$$\min_{\mathbf{x}} \quad \sum_{i=1}^k \rho_v(\mathbf{b}_i - \mathbf{A}_i\mathbf{x}) \quad (27a)$$

$$\text{s.t.} \quad \rho_v(\mathbf{r}_i) := \begin{cases} \frac{1}{2} \|\mathbf{r}_i\|_2^2, & \|\mathbf{r}_i\|_2 \leq \lambda \\ \lambda \|\mathbf{r}_i\|_2 - \frac{\lambda^2}{2}, & \|\mathbf{r}_i\|_2 > \lambda \end{cases} \quad (27b)$$

where  $\rho_v(\mathbf{r}_i)$  is a vector-generalized Huber function. It is now evident that (P3) is equivalent to (27), which rather surprisingly turns out to be a generalization of Huber's M-estimator (22) to the vector case. The sensors capable of achieving a lower  $\|\mathbf{r}_i\|_2$  value and are more likely to be reliable, appear in (27) under the conventional LS criterion. But the sensors having  $\|\mathbf{r}_i\|_2 > \lambda$ , contribute  $\left(\lambda\|\mathbf{r}_i\|_2 - \frac{\lambda^2}{2}\right) < \frac{\|\mathbf{r}_i\|_2^2}{2}$  to the cost and are deemed "less important" in specifying  $\mathbf{x}$ . For the latter set of sensors,  $\mathbf{u}_i^* \neq \mathbf{0}$  holds too. Thus, (P3) not only estimates the unknown vector  $\mathbf{x}$ , but also reveals the sensors most likely to be unreliable in the presence of noise.

Regarding the cutoff parameter  $\lambda$  in (P3) and (27b), it is worth noting that when  $\lambda \rightarrow 0^+$ , the costs of (27) and (P3) tend to the

cost of (P1). Consequently, for  $\lambda \rightarrow 0^+$  the data of all sensors are declared to contain outliers; and according to the previous analysis,  $(\mathbf{b}_i - \mathbf{A}_i\mathbf{x}^*) \rightarrow \mathbf{u}_i^* \neq \mathbf{0}$  for all  $i$ . This suggests that the solution of (P1) does not provide zero residuals anymore. On the other hand, as  $\lambda \rightarrow \infty$ , the same costs reduce to the LS criterion and all sensors are classified as reliable, or  $\mathbf{u}_i^* = \mathbf{0}$  for all  $i$ .

A heuristic rule of thumb for practically selecting  $\lambda$  is setting it to  $\tau\sqrt{m}$ , where  $\tau$  is the equivalent parameter for the scalar case and has been selected according to the techniques mentioned after (22). If the number of reliable sensors is roughly known (e.g., based on prior operation of the network), an alternative approach is solving (P3) for a grid of  $\lambda$  values and selecting the one identifying the prescribed number of outliers. Note that solving (P3) for several values of  $\lambda$  can be efficiently performed either through the group-LARS algorithm [32], or, by using the block coordinate descent algorithm of the next subsection with what is called "warm startup" [14]. The latter initializes the tentative solutions of (P3) for a grid value of  $\lambda$  with the solution derived for the previous grid value of  $\lambda$ . The computational efficiency of such an approach has been numerically verified for the Lasso problem [14], [28].

*Remark 8:* In Problem Statement 2, the noise term was assumed to be independent across sensors. Specifications such as the geographical distribution of sensors may impose correlation across different sensor readings. In this case, if the covariance matrix  $\Sigma$  of the aggregate noise vector  $\mathbf{n}^T := [\mathbf{n}_1^T \cdots \mathbf{n}_k^T]$  is known, a standard preprocessing step is to prewhiten the data as  $\mathbf{b}' := \Sigma^{-1/2}\mathbf{b}$  and  $\mathbf{A}' := \Sigma^{-1/2}\mathbf{A}$ . Prewhitening "spreads" the influence of unreliable sensors across the entries of  $\mathbf{b}'$ . As a result, the LS and  $\ell_1$ -error regression estimators and even the robust Huber M-estimator are not applicable; see also [15] for similar observations in the scalar case ( $m = 1$ ). On the contrary, given that  $\mathbf{u}$  remains block sparse, the (P3) estimator can successfully handle a colored noise setup by simply modifying its cost to  $\frac{\|\mathbf{b}' - \mathbf{A}'\mathbf{x} - \Sigma^{-1/2}\mathbf{u}\|_2^2}{2} + \lambda \sum_{i=1}^k \|\mathbf{u}_i\|_2$ .

### C. A Block Coordinate Descent Algorithm

As mentioned earlier, (P3) is convex. It can be cast as an SOCP and solved by standard, interior point-based solvers. An alternative solver of (P3) exploiting the problem structure and offering computational advantages is block coordinate descent, which has been successfully applied to related optimization problems [13], [31]. The core idea behind this solver is to partition the optimization variable into blocks and minimize iteratively the cost w.r.t. one block variable while keeping the rest fixed.

To apply block coordinate descent to the RSN problem at hand, consider minimizing the cost separately w.r.t.  $\mathbf{x}$  and  $\mathbf{u}$ . Each iteration involves two steps: In the first step, the objective is minimized w.r.t.  $\mathbf{x}$ , while keeping  $\mathbf{u}$  fixed, whereas in the second step the roles are interchanged. Specifically, let  $\mathbf{x}^{(l-1)}$  and  $\mathbf{u}^{(l-1)}$  denote the tentative solutions at the  $(l-1)$ th iteration. During the first step of the  $l$ th iteration, fix  $\mathbf{u} = \mathbf{u}^{(l-1)}$  and find  $\mathbf{x}^{(l)}$  as the minimizer of the resultant quadratic; that is,

$$\mathbf{x}^{(l)} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T(\mathbf{b} - \mathbf{u}^{(l-1)}). \quad (28)$$



In the second step, fix  $\mathbf{x} = \mathbf{x}^{(l)}$  and find the  $\mathbf{u}_i^{(l)}$ 's as the minimizers of the per-sensor optimization problems

$$\min_{\mathbf{u}_i} \frac{1}{2} \|\mathbf{r}_i^{(l)} - \mathbf{u}_i\|_2^2 + \lambda \|\mathbf{u}_i\|_2 \quad (29)$$

where  $\mathbf{r}_i^{(l)} := \mathbf{b}_i - \mathbf{A}_i \mathbf{x}^{(l)}$  for  $i = 1, \dots, k$ . As per (24), the solutions of (29) are provided neatly in closed form<sup>2</sup> as

$$\mathbf{u}_i^{(l)} = \begin{cases} \mathbf{0}, & \|\mathbf{r}_i^{(l)}\|_2 \leq \lambda \\ \mathbf{r}_i^{(l)} \left(1 - \frac{\lambda}{\|\mathbf{r}_i^{(l)}\|_2}\right), & \|\mathbf{r}_i^{(l)}\|_2 > \lambda. \end{cases} \quad (30)$$

The solution in (30) does not require  $\mathbf{x}^{(l)}$ , but only  $\mathbf{r}^{(l)}$ . Combining (28) and (29), it follows that

$$\mathbf{r}^{(l)} = \mathbf{P}_A^\perp \mathbf{b} + \mathbf{P}_A \mathbf{u}^{(l-1)} \quad (31)$$

where  $\mathbf{P}_A := \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$  and  $\mathbf{P}_A^\perp := \mathbf{I} - \mathbf{P}_A$ .

Summarizing, the iterations entail: (a) updating the residuals based on (31); and (b) applying the thresholding rule in (30). As matrix  $\mathbf{P}_A$  and vector  $\mathbf{P}_A^\perp \mathbf{b}$  can be computed offline, the most computationally demanding operation is the matrix-vector product in step (a). Since  $km \geq n$ , this product would better be implemented as  $(\mathbf{A}(\mathbf{A}\mathbf{A}^T)^{-1})(\mathbf{A}^T \mathbf{u})$  in  $O(kmn)$  operations. The developed algorithm has overall complexity  $O(kmn)$  per iteration. The presence of zero blocks in  $\mathbf{u}$  can be further exploited to save computations. Numerical simulations demonstrate that the overall complexity of this block-coordinate approach is much lower than the complexity of the interior point-based algorithms.

Due to the specific form of (P3), convergence of the block coordinate descent iteration follows readily from the results of [30]. The algorithm can be initialized at  $\mathbf{u}^{(0)} = \mathbf{0}$ , so that  $\mathbf{x}^{(1)}$  is the conventional LS solution. It is terminated when the relative error  $\frac{\|\mathbf{u}^{(l)} - \mathbf{u}^{(l-1)}\|_2}{\|\mathbf{u}^{(l)}\|_2}$  becomes smaller than a predefined threshold, e.g.,  $\epsilon = 10^{-6}$ . Upon termination, the output is the solution vector  $\hat{\mathbf{u}}$ , which reveals the sensors affected by outliers, whereas the solution  $\hat{\mathbf{x}}$  can be obtained directly from (28).

#### D. A Nonconvex Surrogate for RSN

In the context of robust linear regression, Huber's M-estimator is just one choice from the class of robust estimators defined as the minimizers of (22) for appropriately chosen  $\rho$  functions. It has been argued that estimators corresponding to nonconvex  $\rho$  functions, such as the bi-square (Tukey's), or Hampel's estimators, yield improved robustness-efficiency trade-offs in practice [22, p. 99]. Similarly in the multivariate case, convex M-estimators [4] are practically replaced by nonconvex M- or S-estimators appropriately initialized [2].

Alternatively, it is of interest to explore a nonconvex surrogate of (P3) paralleling that of Subsection Section III-B. Recall that the RSN solver in (P3) seeks  $\mathbf{x}$  and  $\mathbf{u}$  based on fewer observations than unknowns, but taking advantage of  $\mathbf{u}$ 's block sparsity. To further promote block sparsity in  $\mathbf{u}$ , the  $\|\mathbf{u}_i\|_2$  terms in (P3)

can be replaced by  $\log(\|\mathbf{u}_i\|_2 + \delta)$  for a small positive  $\delta$ , to end up with the nonconvex problem

$$\min_{\mathbf{x}, \mathbf{u}} \frac{1}{2} \|\mathbf{b} - \mathbf{A}\mathbf{x} - \mathbf{u}\|_2^2 + \lambda \sum_{i=1}^k \log(\|\mathbf{u}_i\|_2 + \delta). \quad (P4)$$

Following the majorization-minimization rationale presented in Section III-B, (P4) can be driven to a stationary point [18] using the iterations

$$\begin{aligned} (\mathbf{x}^{(l)}, \mathbf{u}^{(l)}) &:= \arg \min_{\mathbf{x}, \mathbf{u}} \frac{1}{2} \|\mathbf{b} - \mathbf{A}\mathbf{x} - \mathbf{u}\|_2^2 + \lambda \sum_{i=1}^k w_i^{(l)} \|\mathbf{u}_i\|_2, \\ w_i^{(l)} &:= \left( \|\mathbf{u}_i^{(l-1)}\|_2 + \delta \right)^{-1}, \quad i = 1, \dots, k. \end{aligned} \quad (32)$$

The optimization per iteration of (32) is a weighted version of (P3) and thus can be efficiently solved using the steps (28) and (30) after replacing  $\lambda$  in (30) with  $\lambda_i^{(l)} := \lambda w_i^{(l)}$  for all  $i$  at the  $l$ th iteration. The iterations can be initialized with the (P3) solution which corresponds to setting all weights to unity. The simulations of Section VI will demonstrate that the (P4) solver outperforms that of (P3) in terms of the mean-square error (MSE) even after a single iteration. Note that as with (32), single-iteration methods based on nonconvex surrogates of the (group) Lasso cost function have been proposed with well documented properties [23], [34].

## VI. SIMULATED TESTS

### A. Checking the Weak Bound

Among the results of Section IV, the one that can be numerically validated is the weak bound of (17). This bound is termed weak because it refers to the occurrence of a single event  $\mathcal{E}_s$ , namely, to a single partition  $(\mathcal{S}, \bar{\mathcal{S}})$  with  $\mathcal{S} = s$ . According to this bound, if  $\beta$  and  $\gamma$  are kept fixed and as long as  $\beta > \frac{(\sqrt{\gamma+1})}{2}$ , the probability  $\Pr(\mathcal{E}_s)$  is arbitrarily small for large  $n$ .

To validate this result, the entries of  $\mathbf{A}$  are drawn independently from  $\mathcal{N}(0, 1)$  and the unknown vector is modeled as  $\mathbf{x}_0 \sim n^{-\frac{1}{2}} \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$ . Given that  $\Pr(\mathcal{E}_s)$  is invariant to the permutations of the subsystems, the partition  $(\mathcal{S}_0, \bar{\mathcal{S}}_0)$  with  $\mathcal{S}_0 = \{1, \dots, s\}$  is simply selected. The output of the consistent subsystems is  $\mathbf{b}_{\mathcal{S}_0} = \mathbf{A}_{\mathcal{S}_0} \mathbf{x}_0$ ; whereas for the inconsistent ones  $\mathbf{b}_{\bar{\mathcal{S}}_0} = \mathbf{w}$  is simulated with  $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{(k-s)m})$ . Notice that due to the selected normalization, the observation vectors have equal variance, i.e.,  $\mathbb{E}[\|\mathbf{b}_i\|_2^2] = m$  for all  $i \in \mathcal{I}$ . For several  $(n, m)$  pairs, ten values of  $\gamma$  are selected uniformly over the interval  $(0, 1, 1]$  that correspond to ten values of  $k$ . And for every  $\gamma(k)$ , the number of consistent subsystems  $s$  is chosen such that  $\beta(s, k) = \frac{s}{k} \in [0.5, 1]$ . For each pair  $(\gamma(k), \beta(s, k))$ , the probability of (P1) identifying uniquely the (P0) solution is empirically evaluated through 100 Monte Carlo runs. For each experiment, the solution of (P1) is deemed successful whenever  $\hat{\mathbf{x}}$  satisfies  $\|\hat{\mathbf{x}} - \mathbf{x}_0\|_\infty \leq 10^{-4}$ .

The results are depicted in Fig. 2. Every pair  $(\gamma(k), \beta(s, k))$  corresponds to a circle whose face intensity indicates the probability of recovery as explained in the caption. The east and south-east parts of Fig. 2(a)–(e) are not as crowded, since for  $\gamma$  close to 1, the integer  $k$  becomes small, which implies that there

<sup>2</sup>This is not the case for the colored noise scenario discussed in Remark 8, where the vectors  $\{\mathbf{u}_i\}$  can then be jointly found by any group Lasso algorithm instead [32].

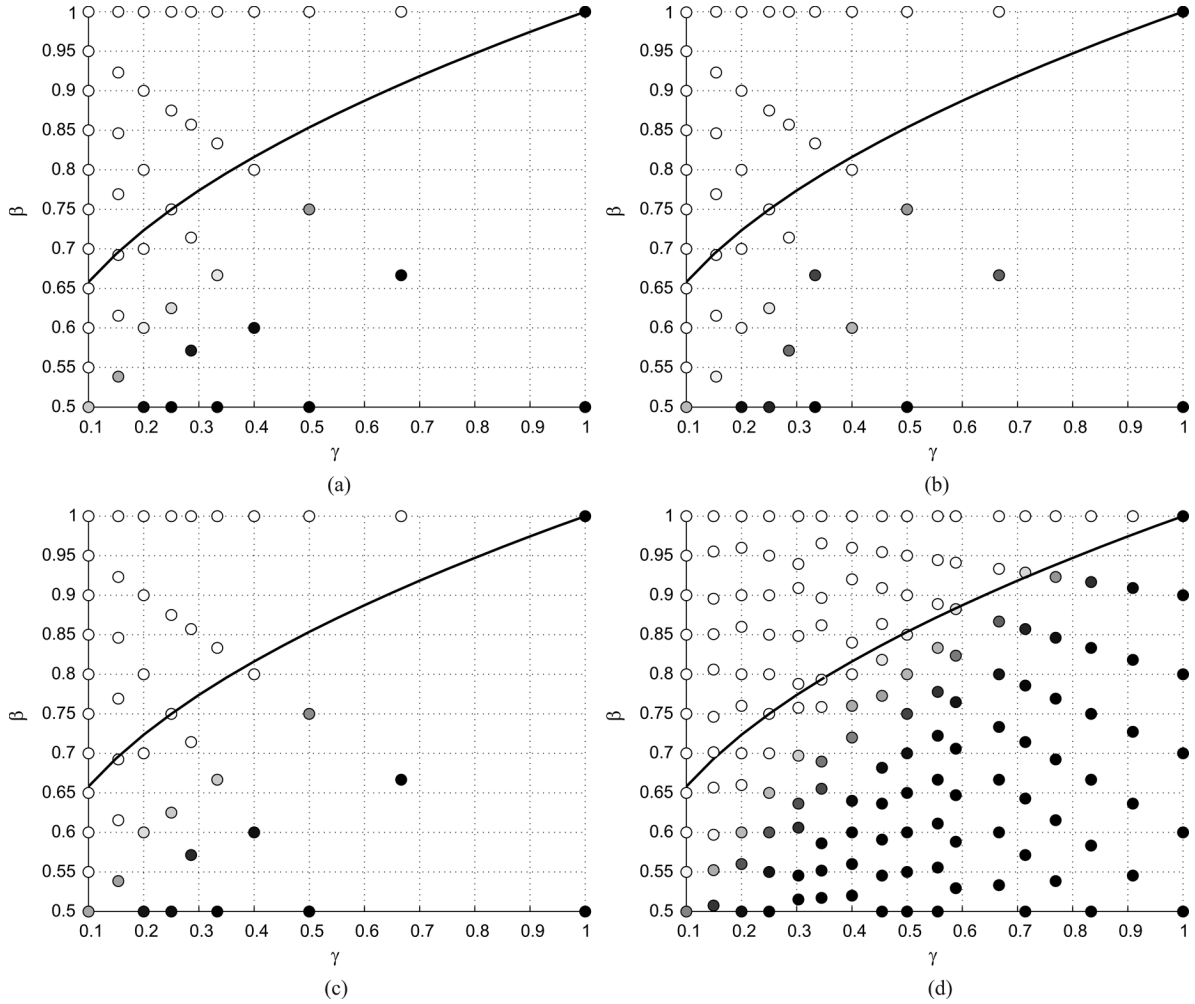


Fig. 2. Empirical probability of success for (P1) and the weak bound of (17) (solid black line). Empty circles correspond to quadruplets  $(n, m, k, s)$  with perfect empirical recovery and solid black circles to problem setups having failed in all experiments: (a)  $n = 40, m = 20$  (b)  $n = 40, m = 10$  (c)  $n = 20, m = 10$  (d)  $n = 20, m = 2$ .

are not many choices for an integer  $s \in [\frac{k}{2}, k]$ . The condition for highly probable recovery in the weak sense,  $\beta = \frac{\sqrt{\gamma+1}}{2}$ , is also shown as a black solid curve. According to the weak bound (17), the circles above this curve correspond to dimension setups with high probability of success for large  $n$ . The empirically evaluated probabilities validate the result even for moderate values of  $n$ .

### B. Test Cases for RS

The RS solvers developed are numerically compared in this subsection. The setup involves a network of  $k = 16$  sensors collecting observation vectors of size  $m = 4$  and an unknown vector of size  $n = 20$ . Quantities  $\mathbf{x}_0$ ,  $\mathbf{A}$  and  $\mathbf{b}$ , all follow the model of the previous experiment and the number of consistent sensors ranges from 8 to 16. The comparison includes i) the LS solution of (1); ii) the  $\ell_1$ -error regression solution of (2); iii) the (P1) solver; and iv) the (P2) solver obtained after one iteration of (7). In addition, a genie-aided LS (GA-LS) solver knowing *a priori* the reliable sensors,  $\hat{\mathbf{x}}_{GA-LS} := (\mathbf{A}_{\mathcal{S}_0}^T \mathbf{A}_{\mathcal{S}_0})^{-1} \mathbf{A}_{\mathcal{S}_0}^T \mathbf{b}_{\mathcal{S}_0}$ , is implemented to serve as a benchmark. The parameter  $\delta$  in (7) is set to  $10^{-4}$ , whereas the simulation results were insensitive to the range of values from  $10^{-2}$  to  $10^{-8}$ .

The sensor detection probability is empirically estimated through 1000 Monte Carlo experiments. An estimate  $\hat{\mathbf{x}}$  is considered to have successfully classified the sensors whenever the residual  $\|\mathbf{b}_i - \mathbf{A}_i \hat{\mathbf{x}}\|_\infty$  is smaller than or equal to  $10^{-4}$  for  $i \in \mathcal{S}_0$  and larger than  $10^{-4}$  for  $i \in \bar{\mathcal{S}}_0$ . As evidenced by Table I, the LS solution fails to identify the reliable subset. In contrast, the novel (P1) scheme shows a clear advantage over the  $\ell_1$ -error regression solution, while the empirical detection probability further improves for the (P2) method, even after a single iteration.

### C. Test Cases for RSN

To evaluate the developed RSN solvers, the unknown vector was fixed at  $\mathbf{x}_0 = \frac{1}{\sqrt{n}}$ , while the reliable sensors followed the model  $\mathbf{b}_{\mathcal{S}_0} = \mathbf{A}_{\mathcal{S}_0} \mathbf{x}_0 + \mathbf{n}_{\mathcal{S}_0}$ , with  $\mathbf{n}_{\mathcal{S}_0} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{sm})$  and known  $\sigma$ . A plausible figure of merit in this scenario is the MSE,  $\mathbb{E}[\|\mathbf{x}_0 - \hat{\mathbf{x}}\|_2^2]$ , which was empirically estimated by averaging over 1000 Monte Carlo experiments.

Comparisons included i) the LS estimator; ii) the GA-LS estimator; iii) the  $\ell_1$ -error estimator of (2); iv) the conventional (scalar) Huber's M-estimator of (P3); v) the (P1) solver; vi) the one-iteration solution of (P2); vii) the (P3) solver; and viii) the

TABLE I  
EMPIRICAL PROBABILITY OF SUCCESSFUL SENSOR CLASSIFICATION (%) FOR THE RS TASK WITH  $(n, m, k) = (20, 4, 16)$

Method	Number of consistent sensors $s$				
	8	10	12	14	16
GA-LS	100.0	100.0	100.0	100.0	100.0
LS	50.0	37.5	25.0	12.5	100.0
$l_1$	51.4	46.3	94.6	100.0	100.0
(P1)	53.5	67.4	99.6	100.0	100.0
(P2)	81.5	99.3	100.0	100.0	100.0

one-iteration solution of (P4). The value of  $\delta$  parameters in (P2) and (P4) turned out to be not critical and were set to  $10^{-4}$ . The cutoff parameter  $\tau$  for the Huber’s M-estimator was selected as  $1.34\sigma$ , whereas  $\lambda$  in both (P3) and (P4) was set to  $1.34\sigma\sqrt{m}$ . It is worth noting that the average number of iterations for the block-coordinate descent algorithm of Section V-C was between 16 (for SNR = 10 dB) and 30 (for SNR = 25 dB), while its execution time was 1000 times lower than that of a standard SOCP solver.

In Fig. 3(a), the MSE achieved by each method is plotted versus the number of consistent sensors  $s$  for SNR = 10 dB. The curves show that the block-sparsity ignorant LS,  $l_1$  and Huber’s estimators are generally outperformed by the novel schemes. The (P1) and (P2) solvers, originally designed for the RS task, still exhibit reasonable performance that worsens as  $s \rightarrow k$ . The (P3) estimator shows a slight improvement; but its solution serves as a good initialization point for the one-iteration estimates of (P4). Note that the derived RSN solvers combine robustness with efficiency in the absence of outliers.

To test the effect of correlated sensor measurements, the following experiment was performed. The reliable sensors were modeled again as  $\mathbf{b}_{S_0} = \mathbf{A}_{S_0} \mathbf{x}_0 + \mathbf{n}_{S_0}$ , the unreliable ones as  $\mathbf{b}_{\bar{S}_0} = \mathbf{n}_w + \mathbf{n}_{\bar{S}_0}$  where  $\mathbf{n}_w \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{(k-s)m})$ , while  $[\mathbf{n}_{S_0}^T \ \mathbf{n}_{\bar{S}_0}^T]^T \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$  and  $\mathbf{\Sigma}$  is a symmetric Toeplitz matrix with first column  $[1 \ 0.9 \ 0.9^2 \ \dots \ 0.9^{km-1}]^T$ . The two RSN solvers were modified according to Remark 8. Fig. 3(b) shows the MSE curves obtained at SNR = 10 dB. In this correlated noise setup, the superiority of RSN solvers is even more prominent.

Correctly classifying the sensors as reliable/unreliable is critical. Once a method has completed this classification task, the estimation of  $\mathbf{x}_0$  can be performed based solely on the sensors classified as reliable. Assuming successful classification, the MSE performance of GA-LS can be attained. The probability of correct sensor classification was evaluated in another simulation setup that differs from the previous ones in the following ways: problem dimensions were  $(n, m, k) = (80, 8, 32)$ ; the reliable sensors followed the linear white Gaussian model at SNR = 5 dB;  $\mathbf{b}_{\bar{S}_0}$  had entries independently drawn from the zero mean Laplacian distribution with variance  $(\sigma^2 + 1)$ ; and  $\tau$  and  $\lambda$  parameters were set to  $\sigma$  and  $\sigma\sqrt{m}$ , respectively. The solvers i)–iii) and iv)–v) do not provide a classification mechanism, hence, a sensor was deemed reliable when its residual  $l_2$ -norm was smaller than  $10^{-4}$ . The Huber’s estimator iv) can identify outlying scalar measurements and a sensor was considered correctly classified when all its measurements were correctly classified. For (P3) and (P4), the identification followed

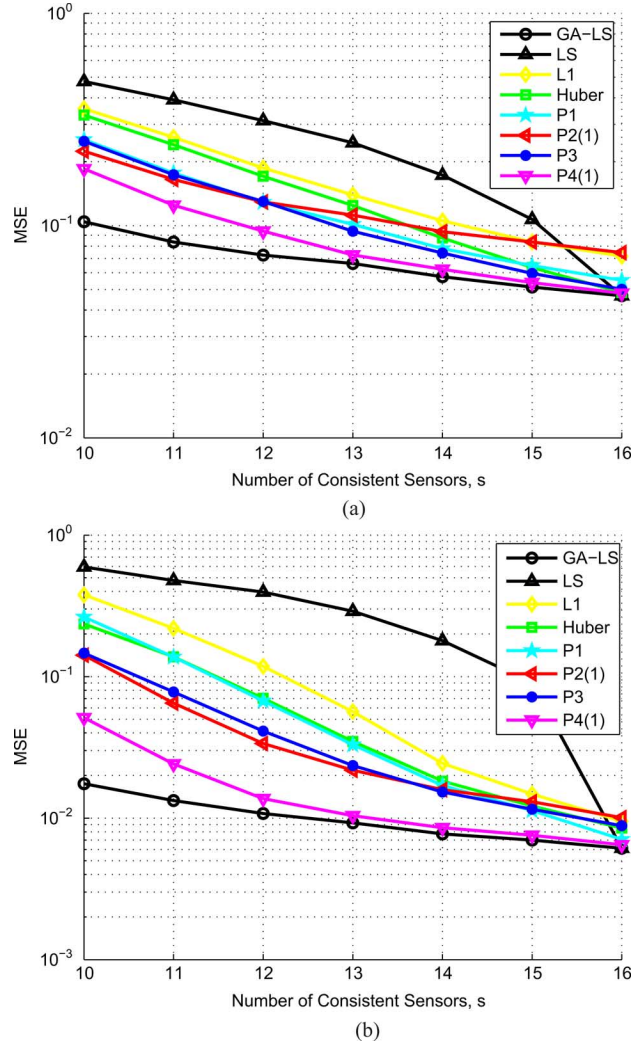


Fig. 3. MSE performance for RSN with  $(n, m, k) = (20, 4, 16)$ . (a) White noise. (b) Colored noise.

TABLE II  
EMPIRICAL PROBABILITY OF SUCCESSFUL SENSOR CLASSIFICATION (%) FOR THE RSN TASK WITH  $(n, m, k) = (80, 8, 32)$

Method	Number of consistent sensors $s$				
	16	20	24	28	32
GA-LS	50.0	37.5	25.0	12.5	0.0
LS	50.0	37.5	25.0	12.5	0.0
$l_1$	50.0	37.5	25.0	12.5	0.0
Huber’s	53.2	43.9	36.0	27.9	20.1
(P1)	50.1	37.6	25.1	12.6	0.1
(P2)	55.0	44.1	31.8	18.5	5.3
(P3)	68.7	73.9	79.6	83.5	84.4
(P4)	72.6	82.8	90.7	96.1	99.1

naturally from the  $\mathbf{u}_i$  vectors. The results are listed in Table II. The majority of methods fail to identify the reliable sensors and yield an empirical probability close to  $(1 - \frac{s}{k})$ , which is the ratio of unreliable sensors. The improvement offered by Huber’s estimator is marginal, while (P3) and in particular (P4) outperform all others.

## VII. CONCLUSION

Contemporary approaches to compressive sampling and variable selection in linear regression problems exploit (block) sparsity present in the signal of interest. The fresh perspective offered in this work broadens the scope of sparsity-exploiting algorithms to settings where model mismatch induced by unreliable sensors or outliers gives rise to (block) sparse residuals, even when the signal of interest is not sparse. This perspective links compressive sampling and sparse linear regression with two important problems: i) finding the maximum number of feasible subsystems of linear equations and ii) robust multivariate linear regression. Capitalizing on these links, robust sensing algorithms were developed to reveal unreliable sensors and recover the signal of interest based on reliable sensors. In the absence of noise, necessary and sufficient conditions were provided for exact recovery (identifiability). Their probabilistic characterization showed that they hold with overwhelming probability when the regression matrix is Gaussian distributed. In the presence of noise, the RS task was reformulated to a combinatorial problem that was subsequently surrogated by (non)convex costs. The two subsystem-aware robust estimators derived can be solved by an efficient block coordinate descent algorithm. The simulated tests demonstrated that all proposed schemes succeed in the task for which they have been designed for.

## APPENDIX

*Proof of Theorem 1:* The sufficiency of the conditions in (10) is shown first. Recall from Lemma 2 that the conditions in (10) imply that  $\mathbf{x}_0$  is the unique minimizer of (P0). Let  $\mathcal{S}$  denote the set of reliable w.r.t.  $\mathbf{x}_0$  sensors with  $|\mathcal{S}| = s > \frac{k}{2}$  for which  $\mathbf{b}_{\mathcal{S}} = \mathbf{A}_{\mathcal{S}}\mathbf{x}_0$ . Vector  $\mathbf{x}_0$  is the unique minimizer of (P1) too if and only if the vector  $(\mathbf{x}_0 - \mathbf{u})$  for any nonzero  $\mathbf{u} \in \mathbb{R}^n$  yields a strictly larger (P1) cost than  $\mathbf{x}_0$  does. Indeed, letting  $\mathbf{v} := \mathbf{A}\mathbf{u}$ , the cost attained by  $(\mathbf{x}_0 - \mathbf{u})$  is

$$\begin{aligned} & \sum_{i=1}^k \|\mathbf{b}_i - \mathbf{A}_i\mathbf{x}_0 + \mathbf{v}_i\|_2 \\ &= \sum_{i \in \mathcal{S}} \|\mathbf{b}_i - \mathbf{A}_i\mathbf{x}_0 + \mathbf{v}_i\|_2 + \sum_{i \in \bar{\mathcal{S}}} \|\mathbf{b}_i - \mathbf{A}_i\mathbf{x}_0 + \mathbf{v}_i\|_2 \\ &\stackrel{(a)}{=} \sum_{i \in \mathcal{S}} \|\mathbf{v}_i\|_2 + \sum_{i \in \bar{\mathcal{S}}} \|\mathbf{b}_i - \mathbf{A}_i\mathbf{x}_0 + \mathbf{v}_i\|_2 \\ &\stackrel{(b)}{\geq} \sum_{i \in \mathcal{S}} \|\mathbf{v}_i\|_2 + \sum_{i \in \bar{\mathcal{S}}} \|\mathbf{b}_i - \mathbf{A}_i\mathbf{x}_0\|_2 - \sum_{i \in \bar{\mathcal{S}}} \|\mathbf{v}_i\|_2 \\ &\stackrel{(c)}{>} \sum_{i=1}^k \|\mathbf{b}_i - \mathbf{A}_i\mathbf{x}_0\|_2 \end{aligned}$$

where equality (a) uses that  $\mathbf{b}_{\mathcal{S}} = \mathbf{A}_{\mathcal{S}}\mathbf{x}_0$ , inequality (b) stems from the reverse triangle inequality and inequality (c) is due to the assumed conditions of the theorem and again the fact that  $\mathbf{b}_{\mathcal{S}} = \mathbf{A}_{\mathcal{S}}\mathbf{x}_0$ .

Necessity is shown by proving the contrapositive. Specifically, it must be shown that if there exists a  $\mathbf{v} \in \text{range}(\mathbf{A})$  and an  $(\mathcal{S}, \bar{\mathcal{S}})$  partition of  $\mathcal{I}$  with  $|\mathcal{S}| = s$  for which  $\sum_{i \in \mathcal{S}} \|\mathbf{v}_i\|_2 \leq \sum_{i \in \bar{\mathcal{S}}} \|\mathbf{v}_i\|_2$ , then there exists an  $\mathbf{x}_0$  that attains a minimum (P0)

cost of  $s$ , but is not the unique minimizer of (P1). Suppose that  $\mathbf{b}_{\mathcal{S}} := \mathbf{A}_{\mathcal{S}}\mathbf{x}_0$  and  $\mathbf{b}_{\bar{\mathcal{S}}} := \frac{\mathbf{A}_{\bar{\mathcal{S}}}\mathbf{x}_0}{2}$  for an  $(\mathcal{S}, \bar{\mathcal{S}})$  partition with  $|\mathcal{S}| = s > \frac{k}{2}$ . Vector  $\mathbf{x}_0$  obviously minimizes (P0), whereas  $\frac{\mathbf{x}_0}{2}$  does not since  $|\bar{\mathcal{S}}| < |\mathcal{S}|$ . Assume  $\mathbf{v} := \mathbf{A}\mathbf{x}_0 \in \text{range}(\mathbf{A})$  and  $\sum_{i \in \mathcal{S}} \|\mathbf{v}_i\|_2 \leq \sum_{i \in \bar{\mathcal{S}}} \|\mathbf{v}_i\|_2$ . It is easy to check that the (P0) costs attained by  $\frac{\mathbf{x}_0}{2}$  and  $\mathbf{x}_0$  are respectively  $\sum_{i \in \mathcal{S}} \frac{\|\mathbf{v}_i\|_2}{2}$  and  $\sum_{i \in \bar{\mathcal{S}}} \frac{\|\mathbf{v}_i\|_2}{2}$ . Hence, it has been shown that  $\frac{\mathbf{x}_0}{2}$  attains a (P1) cost not greater than that of  $\mathbf{x}_0$ , i.e.,  $\mathbf{x}_0$  is not the unique minimizer of (P1). This concludes the proof. ■

**Lemma 4 (Lipschitz Continuity of  $f(\mathbf{A})$ ):** *The function  $f(\mathbf{A})$  defined in (15) is Lipschitz continuous with Lipschitz constant at most  $\sqrt{k}$ .*

*Proof:* Let  $\mathbf{A}, \mathbf{A}' \in \mathbb{R}^{km \times n}$  and  $\mathbf{w}, \mathbf{w}' \in \mathbb{R}^n$  be the minimizing arguments of  $f(\mathbf{A})$  and  $f(\mathbf{A}')$ , respectively. The difference  $f(\mathbf{A}) - f(\mathbf{A}')$  is equal to

$$\begin{aligned} & \left( \sum_{i \in \mathcal{S}} \|\mathbf{A}_i\mathbf{w}\|_2 - \sum_{i \in \bar{\mathcal{S}}} \|\mathbf{A}_i\mathbf{w}\|_2 \right) - \left( \sum_{i \in \mathcal{S}} \|\mathbf{A}'_i\mathbf{w}'\|_2 - \sum_{i \in \bar{\mathcal{S}}} \|\mathbf{A}'_i\mathbf{w}'\|_2 \right) \\ &\stackrel{(a)}{\leq} \left( \sum_{i \in \mathcal{S}} \|\mathbf{A}_i\mathbf{w}'\|_2 - \sum_{i \in \bar{\mathcal{S}}} \|\mathbf{A}_i\mathbf{w}'\|_2 \right) - \left( \sum_{i \in \mathcal{S}} \|\mathbf{A}'_i\mathbf{w}'\|_2 - \sum_{i \in \bar{\mathcal{S}}} \|\mathbf{A}'_i\mathbf{w}'\|_2 \right) \\ &\stackrel{(b)}{\leq} \sum_{i \in \mathcal{S}} \|(\mathbf{A}_i - \mathbf{A}'_i)\mathbf{w}'\|_2 + \sum_{i \in \bar{\mathcal{S}}} \|(\mathbf{A}_i - \mathbf{A}'_i)\mathbf{w}'\|_2 \\ &\stackrel{(c)}{\leq} \sup_{\|\mathbf{u}\|_2=1} \sum_{i=1}^k \|\tilde{\mathbf{A}}_i\mathbf{u}\|_2 \end{aligned}$$

where inequality (a) holds because  $\mathbf{w}$  is by definition the minimizer of  $f(\mathbf{A})$ ; (b) follows from the reverse triangle inequality applied on each subset; (c) holds trivially for  $\|\mathbf{w}'\|_2 = 1$ ; and  $\tilde{\mathbf{A}}_i := \mathbf{A}_i - \mathbf{A}'_i$ .

Now, define the function appearing in the right-hand side of the last inequality as

$$g(\tilde{\mathbf{A}}) := \sup_{\|\mathbf{u}\|_2=1} \sum_{i=1}^k \|\tilde{\mathbf{A}}_i\mathbf{u}\|_2 \quad (33)$$

so that  $f(\mathbf{A}) - f(\mathbf{A}') \leq g(\tilde{\mathbf{A}})$ . Since  $f(\mathbf{A}') - f(\mathbf{A}) \leq g(-\tilde{\mathbf{A}}) = g(\tilde{\mathbf{A}})$ , it holds that  $|f(\mathbf{A}) - f(\mathbf{A}')| \leq g(\tilde{\mathbf{A}})$ . Given that  $g(\mathbf{0}) = 0$ , if  $g(\mathbf{A})$  is Lipschitz continuous with constant at most  $L$ , i.e.,  $|g(\mathbf{A})| \leq L\|\mathbf{A}\|_F$ , where  $\|\mathbf{A}\|_F$  is the Frobenius norm of  $\mathbf{A}$ , then  $f(\mathbf{A})$  is also Lipschitz continuous and its constant is at most  $L$ . Thus, it suffices to show that  $g(\mathbf{A})$  is Lipschitz continuous and its constant is upper bounded by  $\sqrt{k}$ .

To proceed, recall first that the  $\ell_2$ -norm of a vector  $\mathbf{x} \in \mathbb{R}^n$  can be written as [6, p. 637]

$$\|\mathbf{x}\|_2 = \sup\{\mathbf{x}^T\mathbf{y} : \|\mathbf{y}\|_2 \leq 1\}. \quad (34)$$

Using (34),  $g(\mathbf{A})$  can alternatively be expressed as

$$g(\mathbf{A}) = \sup_{\|\mathbf{u}\|_2=1} \sup_{\|\mathbf{v}_i\|_2 \leq 1} \sum_{i=1}^k \mathbf{v}_i^T \mathbf{A}_i \mathbf{u} \quad (35)$$

which is a supremum over infinitely many linear functions of  $\mathbf{A}$  and as such it is convex. Recall that if a function  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  is convex with a subgradient  $s(\mathbf{x})$  for which  $\|\sup_{\mathbf{x}} s(\mathbf{x})\|_2$  is finite, then  $f$  is Lipschitz with constant  $L \leq \|\sup_{\mathbf{x}} s(\mathbf{x})\|_2$ .

This claim can be proved by the definitions of the subgradient and the Lipschitz constant. Thus, it suffices to find a subgradient of  $g(\mathbf{A})$  and upper bound its norm.

If  $\mathbf{u}_*$  and  $\{\mathbf{v}_{*,i}\}_{i=1}^k$  are the maximizers of  $g(\mathbf{A})$ , then a subgradient is given by the matrix  $\mathbf{G}(\mathbf{A}) = [\mathbf{u}_* \mathbf{v}_{*,1}^T \cdots \mathbf{u}_* \mathbf{v}_{*,k}^T]$  with  $\|\mathbf{u}_*\|_2 = 1$  and  $\|\mathbf{v}_{*,i}\|_2 \leq 1$  for  $i = 1, \dots, k$ . The norm of this subgradient is

$$\begin{aligned} \|\mathbf{G}(\mathbf{A})\|_F &= \sqrt{\sum_{i=1}^k \|\mathbf{u}_* \mathbf{v}_{*,i}^T\|_F^2} = \sqrt{\sum_{i=1}^k \|\mathbf{u}_*\|_2^2 \|\mathbf{v}_{*,i}\|_2^2} \\ &= \sqrt{\sum_{i=1}^k \|\mathbf{v}_{*,i}\|_2^2} \leq \sqrt{k}. \end{aligned}$$

The bound is independent of  $\mathbf{A}$  and the proof is complete.  $\blacksquare$

**Lemma 5 (Expected Value Lower Bound):** *For the random matrix  $\mathbf{A} \in \mathbb{R}^{km \times n}$  with entries drawn independently from  $\mathcal{N}(0, 1)$ , it holds that  $\mathbb{E}[f(\mathbf{A})] \geq \mu$  with  $\mu := \left(\frac{2\beta-1}{\sqrt{\gamma}} - 1\right) \sqrt{kn} (1 + o_n(1))$ .*

*Proof:* Consider rewriting  $f(\mathbf{A})$  using (34) as

$$f(\mathbf{A}) = \inf_{\|\mathbf{u}\|_2=1} \sup_{\|\mathbf{v}_i\|_2 \leq 1} \inf_{\|\mathbf{z}_i\|_2 \leq 1} \sum_{i \in \mathcal{S}} \mathbf{v}_i^T \mathbf{A}_i \mathbf{u} + \sum_{i \in \bar{\mathcal{S}}} \mathbf{z}_i^T \mathbf{A}_i \mathbf{u}. \quad (36)$$

Next, introduce auxiliary random vectors  $\mathbf{y} \in \mathbb{R}^n$ ,  $\mathbf{s}_i \in \mathbb{R}^m$ ,  $\mathbf{t}_i \in \mathbb{R}^m$  for  $i = 1, \dots, k$  and  $w \in \mathbb{R}$  having their entries drawn independently from  $\mathcal{N}(0, 1)$  and define the functionals

$$h_y(\mathbf{u}, \mathbf{v}_i, \mathbf{z}_i) := \sum_{i \in \mathcal{S}} \mathbf{v}_i^T \mathbf{A}_i \mathbf{u} + \sum_{i \in \bar{\mathcal{S}}} \mathbf{z}_i^T \mathbf{A}_i \mathbf{u} + \sqrt{k} w \quad (37a)$$

$$h_x(\mathbf{u}, \mathbf{v}_i, \mathbf{z}_i) := \sum_{i \in \mathcal{S}} \mathbf{v}_i^T \mathbf{s}_i + \sum_{i \in \bar{\mathcal{S}}} \mathbf{z}_i^T \mathbf{t}_i + \sqrt{k} \mathbf{u}^T \mathbf{y}, \text{ for } \quad (37b)$$

$$\|\mathbf{u}\|_2 = 1, \{ \|\mathbf{v}_i\|_2 \leq 1 \}_{i=1}^k \text{ and } \{ \|\mathbf{z}_i\|_2 \leq 1 \}_{i=1}^k. \quad (38)$$

Consider now the triplets  $(\mathbf{u}, \mathbf{v}_i, \mathbf{z}_i)$  and  $(\mathbf{u}', \mathbf{v}'_i, \mathbf{z}'_i)$ . By using the i.i.d. property of the random variables appearing in the functionals, it holds that

$$\mathbb{E}[h_y(\mathbf{u}, \mathbf{v}_i, \mathbf{z}_i) h_y(\mathbf{u}', \mathbf{v}'_i, \mathbf{z}'_i)] = \mathbf{u}^T \mathbf{u}' \left( \sum_{i \in \mathcal{S}} \mathbf{v}_i^T \mathbf{v}'_i + \sum_{i \in \bar{\mathcal{S}}} \mathbf{z}_i^T \mathbf{z}'_i \right) + k$$

$$\mathbb{E}[h_x(\mathbf{u}, \mathbf{v}_i, \mathbf{z}_i) h_x(\mathbf{u}', \mathbf{v}'_i, \mathbf{z}'_i)] = \sum_{i \in \mathcal{S}} \mathbf{v}_i^T \mathbf{v}'_i + \sum_{i \in \bar{\mathcal{S}}} \mathbf{z}_i^T \mathbf{z}'_i + k \mathbf{u}^T \mathbf{u}'$$

while the difference of the two expectations above is  $(\mathbf{u}^T \mathbf{u}' - 1) (\sum_{i \in \mathcal{S}} \mathbf{v}_i^T \mathbf{v}'_i + \sum_{i \in \bar{\mathcal{S}}} \mathbf{z}_i^T \mathbf{z}'_i - k)$ . By exploiting the properties of vectors  $\mathbf{u}$ ,  $\mathbf{v}_i$  and  $\mathbf{z}_i$  in (38), it follows that

$$\mathbb{E}[h_x(\mathbf{u}, \mathbf{v}_i, \mathbf{z}_i) h_x(\mathbf{u}', \mathbf{v}'_i, \mathbf{z}'_i)] = \mathbb{E}[h_y(\mathbf{u}, \mathbf{v}_i, \mathbf{z}_i) h_y(\mathbf{u}', \mathbf{v}'_i, \mathbf{z}'_i)] \quad (39a)$$

$$\mathbb{E}[h_x(\mathbf{u}, \mathbf{v}_i, \mathbf{z}_i) h_x(\mathbf{u}', \mathbf{v}'_i, \mathbf{z}'_i)] \leq \mathbb{E}[h_y(\mathbf{u}, \mathbf{v}_i, \mathbf{z}_i) h_y(\mathbf{u}', \mathbf{v}'_i, \mathbf{z}'_i)]. \quad (39b)$$

To proceed, the following lemma is needed [16, Cor. 10].

**Lemma 6 [16]:** *Let  $\{X_{ijk}\}$  and  $\{Y_{ijk}\}$  be two zero-mean Gaussian processes indexed by  $(i, j, k)$  for  $i = 1, \dots, n$ ,  $j =$*

*1, \dots, m and  $k = 1, \dots, p$ , which satisfy the following conditions:*

- c1)  $\mathbb{E}[X_{ijk}^2] = \mathbb{E}[Y_{ijk}^2]$  for all  $(i, j, k)$ .
- c2) For any two triplets  $\alpha = (i, j, k)$  and  $\alpha' = (i', j', k')$ ,  $\mathbb{E}[X_\alpha X_{\alpha'}] \geq \mathbb{E}[Y_\alpha Y_{\alpha'}]$  if  $i = i'$  and  $j \neq j'$  and  $\mathbb{E}[X_\alpha X_{\alpha'}] \leq \mathbb{E}[Y_\alpha Y_{\alpha'}]$  in all other cases.

Under c1) and c2), it holds that

$$\mathbb{E} \left[ \max_i \min_j \max_k X_{ijk} \right] \geq \mathbb{E} \left[ \max_i \min_j \max_k Y_{ijk} \right]. \quad (40)$$

Even though the indexes  $(i, j, k)$  are denumerable, by using the compactness argument of [25, Pr.1.], the comparison in (40) extends to minimizations/maximizations over compact sets as well. Mapping the  $X_{ijk}(Y_{ijk})$  variables of Lemma 6 to  $-h_x(\mathbf{u}, \mathbf{v}_i, \mathbf{z}_i)(-h_y(\mathbf{u}, \mathbf{v}_i, \mathbf{z}_i))$ , it can be verified that the conditions of the lemma are met (cf. (39)) and upon using (40) deduce that

$$\begin{aligned} \mathbb{E} \left[ \sup_{\|\mathbf{u}\|_2=1} \inf_{\|\mathbf{z}_i\|_2 \leq 1} \sup_{\|\mathbf{v}_i\|_2 \leq 1} -h_x(\mathbf{u}, \mathbf{v}_i, \mathbf{z}_i) \right] \\ \geq \mathbb{E} \left[ \sup_{\|\mathbf{u}\|_2=1} \inf_{\|\mathbf{z}_i\|_2 \leq 1} \sup_{\|\mathbf{v}_i\|_2 \leq 1} -h_y(\mathbf{u}, \mathbf{v}_i, \mathbf{z}_i) \right]. \end{aligned}$$

Given that  $\sup_x -f(x) = -\inf_x f(x)$ , the previous inequality is equivalent to

$$\begin{aligned} \mathbb{E} \left[ \inf_{\|\mathbf{u}\|_2=1} \inf_{\|\mathbf{z}_i\|_2 \leq 1} \sup_{\|\mathbf{v}_i\|_2 \leq 1} h_x(\mathbf{u}, \mathbf{v}_i, \mathbf{z}_i) \right] \\ \leq \mathbb{E} \left[ \inf_{\|\mathbf{u}\|_2=1} \inf_{\|\mathbf{z}_i\|_2 \leq 1} \sup_{\|\mathbf{v}_i\|_2 \leq 1} h_y(\mathbf{u}, \mathbf{v}_i, \mathbf{z}_i) \right] \stackrel{(a)}{=} \mathbb{E}[f(\mathbf{A})] \end{aligned}$$

where (a) comes from (36) and the fact that the  $w$  in (37a) is zero mean. Thus, it has been established that

$$\mathbb{E}[f(\mathbf{A})] \geq \mathbb{E} \left[ \inf_{\|\mathbf{u}\|_2=1} \inf_{\|\mathbf{z}_i\|_2 \leq 1} \sup_{\|\mathbf{v}_i\|_2 \leq 1} h_x(\mathbf{u}, \mathbf{v}_i, \mathbf{z}_i) \right].$$

Using the definition of  $h_x(\mathbf{u}, \mathbf{v}_i, \mathbf{z}_i)$  and exploiting the separability of the optimization, as well as the properties in (38), one arrives at

$$\mathbb{E}[f(\mathbf{A})] \geq s \mathbb{E}[\|\mathbf{s}_i\|_2] - (k-s) \mathbb{E}[\|\mathbf{t}_i\|_2] - \sqrt{k} \mathbb{E}[\|\mathbf{y}\|_2]. \quad (41)$$

Recall that if  $\mathbf{x} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{I}_n)$ , then  $\|\mathbf{x}\|_2$  is chi-distributed with  $n$  degrees of freedom and mean value

$$\mathbb{E}[\|\mathbf{x}\|_2] = \frac{\sqrt{2\pi}}{B\left(\frac{n}{2}, \frac{1}{2}\right)} \quad (42)$$

where  $B(\cdot, \cdot)$  denotes the Beta function. Applying (42) three times in (41) yields

$$\mathbb{E}[f(\mathbf{A})] \geq \mu = (2s - k) \frac{\sqrt{2\pi}}{B\left(\frac{m}{2}, \frac{1}{2}\right)} - \sqrt{k} \frac{\sqrt{2\pi}}{B\left(\frac{n}{2}, \frac{1}{2}\right)}.$$

Using the standard approximation  $\frac{\sqrt{2\pi}}{B\left(\frac{n}{2}, \frac{1}{2}\right)} = \sqrt{n}(1 + o_n(1))$  [1, Formulas 6.1.46 and 6.2.2] and for fixed  $\gamma = \frac{n}{km}$  and  $k$ , it also

holds that  $\frac{\sqrt{2\pi}}{B(\frac{\alpha}{2}, \frac{1}{2})} = \frac{\sqrt{n}}{\sqrt{\gamma k}}(1 + o_n(1))$ . Thus, the bound  $\mu$  can be compactly expressed as  $\mu = \left(\frac{2\beta-1}{\sqrt{\gamma}} - 1\right) \sqrt{kn}(1 + o_n(1))$ , which concludes the proof. ■

#### ACKNOWLEDGMENT

The authors wish to thank the anonymous reviewers for their helpful feedback which improved exposition of this paper.

#### REFERENCES

- [1] *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, M. Abramowitz and I. A. Stegun, Eds.. New York: Dover, 1972.
- [2] S. V. Aelst and G. Willems, "Multivariate regression S-estimators for robust estimation and inference," *Statistica Sinica*, vol. 15, pp. 981–1001, 2005.
- [3] E. Amaldi and V. Kann, "The complexity and approximability of finding maximum feasible subsystems of linear relations," *Theoretic. Comput. Sci.*, vol. 147, pp. 181–210, 1995.
- [4] Z. D. Bai, C. R. Rao, and Y. Wu, "M-estimation of multivariate linear regression parameters under a convex discrepancy function," *Statistica Sinica*, vol. 2, pp. 237–254, 1992.
- [5] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.
- [6] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York: Cambridge Univ. Press, 2004.
- [7] E. J. Candès, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted  $\ell_1$  minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877–905, Dec. 2008.
- [8] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
- [9] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, pp. 33–61, 1998.
- [10] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decompositions," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.
- [11] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.
- [12] M. Fazel, "Matrix rank minimization with applications," Ph.D. dissertation, Electr. Eng. Dept., Stanford Univ., Stanford, CA, 2002.
- [13] J. Friedman, T. Hastie, H. Hofling, and R. Tibshirani, "Pathwise coordinate optimization," *Ann. Appl. Stat.*, vol. 1, pp. 302–332, Dec. 2007.
- [14] J. Friedman, T. Hastie, and R. Tibshirani, "Regularized paths for generalized linear models via coordinate descent," *J. Stat. Softw.*, vol. 33, no. 1, pp. 1–22, 2010.
- [15] J.-J. Fuchs, "An inverse problem approach to robust regression," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Phoenix, AZ, 1999, pp. 1809–1812.
- [16] Y. Gordon, "Gaussian processes and almost spherical sections of convex bodies," *Ann. Probab.*, vol. 16, no. 1, pp. 180–188, 1988.
- [17] P. J. Huber and E. M. Ronchetti, *Robust Statistics*. New York: Wiley, 2009.
- [18] K. Lange, D. Hunter, and I. Yang, "Optimization transfer using surrogate objective functions with discussion," *J. Comput. Graph. Stat.*, vol. 9, pp. 1–59, 2000.
- [19] M. Ledoux and M. Talagrand, *Probability in Banach Spaces*. Berlin, Germany: Springer-Verlag, 1991.
- [20] M. S. Lobo, L. Vandenberghe, H. Le Bret, and S. Boyd, "Applications of second-order cone programming," *Linear Algebra Appl.*, vol. 284, pp. 193–228, 1998.
- [21] O. L. Mangasarian and D. R. Musicant, "Robust linear and support vector regression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 9, pp. 950–955, Sep. 2000.
- [22] R. A. Maronna, R. D. Martin, and V. J. Yohai, *Robust Statistics: Theory and Methods*. New York: Wiley, 2006.
- [23] Y. Nardi and A. Rinaldo, "On the asymptotic properties of the group Lasso estimator for linear models," *Electron. J. Stat.*, vol. 2, pp. 605–633, 2008.

- [24] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum rank solutions to linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.
- [25] B. Recht, W. Xu, and B. Hassibi, "Null space conditions and thresholds for rank minimization," *Math. Program. B.*, vol. 127, pp. 175–211, 2011.
- [26] M. Stojnic, "Strong thresholds for  $\frac{\ell_2}{\ell_1}$ -optimization in block-sparse compressed sensing," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Taipei, Taiwan, R.O.C., 2009, pp. 3025–3028.
- [27] M. Stojnic, F. Parvaresh, and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3075–3085, Aug. 2009.
- [28] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. Roy. Stat. Soc. B.*, vol. 58, no. 1, pp. 267–288, 1996.
- [29] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 1030–1051, Mar. 2006.
- [30] P. Tseng, "Convergence of block coordinate descent method for nondifferentiable minimization," *J. Optim. Theory Appl.*, vol. 109, pp. 475–494, Jun. 2001.
- [31] T. T. Wu and K. Lange, "Coordinate descent algorithms for Lasso penalized regression," *Ann. Appl. Stat.*, vol. 2, pp. 224–244, Mar. 2008.
- [32] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Stat. Soc. B.*, vol. 68, no. 1, pp. 49–67, 2006.
- [33] Y. Zhang, N. Meratnia, and P. Havinga, "Outlier detection techniques for wireless sensor networks: A survey," *IEEE Commun. Surveys Tut.*, vol. 12, no. 2, pp. 159–170, 2010.
- [34] H. Zou and R. Li, "One-step sparse estimates in nonconcave penalized likelihood models," *Ann. Stat.*, vol. 36, no. 4, pp. 1509–1533, 2008.



**Vassilis Kekatos** (M'10) was born in Athens, Greece, in 1978. He received the Diploma, M.Sc., and Ph.D. degrees in computer engineering and computer science from the University of Patras, Greece, in 2001, 2003, and 2007, respectively.

Since 2009, he has been a Marie Curie Fellow and he is currently a Postdoctoral Associate with the Department of Electrical and Computer Engineering of the University of Minnesota. His research interests lie in the areas of statistical signal processing with emphasis on compressive sampling and wireless com-

munications.



**Georgios B. Giannakis** (F'97) received the Diploma degree in electrical engineering from the National Technical University of Athens, Greece, in 1981 and the M.Sc. degree in electrical engineering, the M.Sc. degree in mathematics, and the Ph.D. degree in electrical engineering from the University of Southern California (USC) in 1983, 1986, and 1986, respectively.

Since 1999, he has been a Professor with the University of Minnesota, where he now holds an ADC Chair in Wireless Telecommunications in the Electric and Computer Engineering Department and serves as Director of the Digital Technology Center. His general interests span the areas of communications, networking and statistical signal processing subjects on which he has published more than 300 journal papers, 500 conference papers, two edited books, and two research monographs. Current research focuses on compressive sensing, cognitive radios, network coding, cross-layer designs, mobile ad hoc networks, wireless sensor, power, and social networks.

Dr. Giannakis is the (co-)inventor of 20 patents issued, and the (co-)recipient of seven paper awards from the IEEE Signal Processing (SP) and Communications Societies, including the G. Marconi Prize Paper Award in Wireless Communications. He also received Technical Achievement Awards from the SP Society (2000), from EURASIP (2005), a Young Faculty Teaching Award, and the G. W. Taylor Award for Distinguished Research from the University of Minnesota. He is a Fellow of EURASIP and has served the IEEE in a number of posts, including that of a Distinguished Lecturer for the IEEE-SP Society.