Robust Layered Sensing: From Sparse Signals to Sparse Residuals

Vassilis Kekatos and Georgios B. Giannakis
Dept. of Electrical and Computer Engr.
University of Minnesota
Minneapolis,
MN 55455, USA
Emails: {kekatos,georgios}@umn.edu.

***Abstract*—One of the key challenges in sensing networks is the extraction of information by fusing data from a multitude of possibly unreliable sensors. Robust sensing, viewed here as the simultaneous recovery of the wanted information-bearing signal vector together with the subset of (un)reliable sensors, is a problem whose optimum solution incurs combinatorial complexity. The present paper relaxes this problem to its closest convex approximation that turns out to yield a vector-generalization of Huber's scalar criterion for robust linear regression. The novel generalization is shown equivalent to a second-order cone program (SOCP), and exploits the block-sparsity inherent to a suitable model of the residuals. A computationally efficient solver is developed using a block-coordinate descent algorithm, and is tested with simulations.**

I. Introduction

Recent advances in sensor technology have made it feasible to deploy a network of sensors for carrying out synergistically even sophisticated inference tasks. In applications such as environmental monitoring, the typical concept of operation involves a large set of sensors locally observing the signal of interest, and transmitting their measurements to a higher-layer 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 linearly related to the unknown variable(s). Such a *linear* model can arise when the sensing system is viewed as a linear filter, or, when the measured field is linearly represented over a fixed basis;

(c2) Due to stringent power, delay, bandwidth, or model constraints, the linear model can be also *under-determined*, i.e., the dimension of the unknown vector exceeds that of each sensor's vector observation; and

(c3) Not all sensors are *reliable* due to failures in the sensing devices, fades of the sensor-to-fusion-center communication link, physical obstruction of the scene of interest, and (un)intentional interference; see Fig. 1.

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. Even though the related problem of outlier detection in sensor networks has been studied extensively (see e.g., [16] for a recent survey), the RS setup and the approaches described here have not been considered before. In [9], RS was studied under the asymptotically high SNR regime, identifiability conditions were developed, and it was shown that for Gaussian sensing matrices a convex relaxation of the NP-hard problem involved recovers the solution with overwhelming probability.

Work was supported by the European Community's Seventh Framework Programme (FP7/2008 under grant agreement No. 234914); and by NSF grant CCF-1016605.

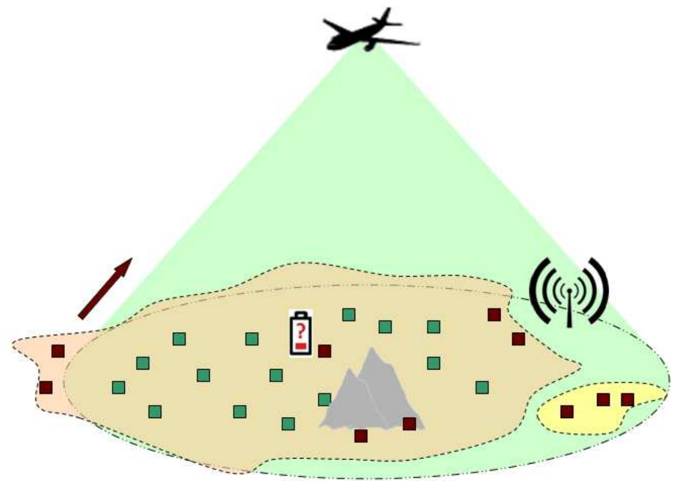

Fig. 1. A wireless sensor network linked with a fusion center. (Un)reliable sensors are color coded as (red) green.

The approach here considers a practical noisy setup and views the unreliable sensors as outliers, thus placing the RS task under a robust multivariate linear regression framework [1]. By proper modeling of the unreliable sensors, the RS task is formulated as a combinatorial optimization problem and subsequently surrogated by (non-)convex approximants. Interestingly, the novel cost functions turns out to be a block version of Huber's function [8]. The resultant optimization problem is transformed to a group Lasso-type cost [15], and a computationally attractive block-coordinate descent algorithm is developed. The simulations corroborate the results.

II. Problem Statement and Preliminaries

To concretely formulate the problem under conditions (c1)-(c3), consider a network of k sensors. Each sensor acquires a measurement vector $\{\mathbf{b}_i \in \mathbb{R}^m\}_{i=1}^k$ through the known sensing matrix $\{\mathbf{A}_i \in \mathbb{R}^{m\times n}\}_{i=1}^k$. However, among the set of sensors $\mathcal{I} := \{1,\ldots,k\}$, only s of them are reliable and adhere to the per-sensor linear regression model

$$\mathbf{b}_i = \mathbf{A}_i\mathbf{x}_o + \mathbf{v}_i,\ i \in \mathcal{S}_o \tag{1}$$

where $\mathcal{S}_o$ denotes the unknown subset of reliable sensors $\mathcal{S}_o \subset \mathcal{I}$ with cardinality $|\mathcal{S}_o| = s$, and $\mathbf{v}_i$ stands for independent identically distributed zero-mean noise of variance σ^2 that captures quantization effects, communication noise, and/or unmodeled dynamics. With reference to (1), the task of *robust sensing* (RS) amounts to estimating the unknown $\mathbf{x}_o$ as the solution of

$$\min_{\mathbf{x}} \min_{|\mathcal{S}|=s} \|\mathbf{b}_{\mathcal{S}} - \mathbf{A}_{\mathcal{S}}\mathbf{x}\|_2^2 \tag{P_0}$$

where $\mathbf{A}_{\mathcal{S}}$ is the $|\mathcal{S}|m\times n$ matrix constructed by concatenating $\{\mathbf{A}_i\}_{i\in\mathcal{S}}$, and likewise for $\mathbf{b}_{\mathcal{S}}$ for any $\mathcal{S} \subset \mathcal{I}$. Unfortunately, solving the problem (P_0) incurs combinatorial complexity, since one has to find all $\binom{k}{s}$ solutions of (P_0) using e.g., least-squares (LS).

Looking for practical solvers, the LS estimator aiming to minimize $\|\mathbf{b}_{\mathcal{I}} - \mathbf{A}_{\mathcal{I}}\mathbf{x}\|_2^2$ is not appropriate, since the unreliable sensors will adversely influence the solution. When the sensing matrices $\mathbf{A}_i$'s are tall ($m \geq n$), a heuristic approach would be to find the per sensor LS solutions, i.e., $(\mathbf{A}_i^T\mathbf{A}_i)^{-1}\mathbf{A}_i^T\mathbf{b}_i$, and try identifying the outlying ones. The problem becomes even more challenging for $m < n$, and recall that under-determinacy can arise naturally as explained in operating condition (c2).

One could possibly try ℓ_1-error regression, which seeks to $\min_{\mathbf{x}} \|\mathbf{b}_{\mathcal{I}} - \mathbf{A}_{\mathcal{I}}\mathbf{x}\|_1$ and is known to be outlier resistant, or an M-estimator from robust linear regression [8]. However, these two approaches do not account for the block structure inherent to the problem. Indeed, as outliers constitute quantities violating (1), they are not single measurements, but a group of m measurements.

III. A Convex Relaxation

Thus, the RS task falls under the realm of robust multivariate linear regression [1]. The novel approach developed here starts by recognizing that model (1) is not valid for the unreliable sensors. Hence, consider modeling the unreliable sensors using the auxiliary outlier vectors $\{\mathbf{o}_i \in \mathbb{R}^m\}_{i=1}^k$. Vector $\mathbf{o}_i$ is zero if the i-th sensor is reliable; and $\mathbf{o}_i \neq \mathbf{0}$ deterministically, otherwise. Model (1) can now be extended to incorporate the unreliable sensors as $\mathbf{b}_i = \mathbf{A}_i\mathbf{x} + \mathbf{o}_i + \mathbf{n}_i$ for every $i \in \mathcal{I}$, or collectively

$$\mathbf{b} = \mathbf{A}\mathbf{x}_o + \mathbf{o} + \mathbf{n} \tag{2}$$

where the subscript $\mathcal{I}$ has been dropped from the variables $\mathbf{b}$, $\mathbf{A}$, $\mathbf{o}$, and $\mathbf{n}$ for notational simplicity.

Based on (2), the RS problem can now be cast as a linear regression problem with respect to the vector $\left[\mathbf{x}^T\ \mathbf{o}^T\right]^T$. Note that even when $\mathbf{A}$ is tall ($km \geq n$), the involved regression matrix $[\mathbf{A}\ \mathbf{I}_{km}]$ is under-determined. To derive a well-posed regression problem, the block sparsity of $\mathbf{o}$ should be explicitly taken into account by constraining the number of nonzero $\mathbf{o}_i$'s as

$$\min_{\mathbf{x},\mathbf{o}} \ \|\mathbf{b} - \mathbf{A}\mathbf{x} - \mathbf{o}\|_2^2 \quad \text{s.to} \ \sum_{i=1}^k \mathrm{I}(\|\mathbf{o}_i\|_2 \neq 0) \leq s \tag{3}$$

where $\mathrm{I}(\cdot)$ is the indicator function. The problem in (3) is non-convex due to the constraints. Following the idea of convex relaxation that has successfully been applied in the area of compressive sampling (CS) [5], [4], the indicator function is replaced by the ℓ_2-norm of the $\mathbf{o}_i$'s which is its closest convex approximation. The optimization problem derived in its Lagrangian form is

$$\min_{\mathbf{x},\mathbf{o}} \frac{1}{2}\|\mathbf{b} - \mathbf{A}\mathbf{x} - \mathbf{o}\|_2^2 + \lambda\sum_{i=1}^k \|\mathbf{o}_i\|_2 \tag{P_1}$$

where $\lambda > 0$ can be selected in the ways described later. The second term in the penalized LS estimator of (P_1) resembles the group Lasso cost function [13], [15]. Similarly to group Lasso, we will show that the ℓ_2-norm penalty favors zero $\mathbf{o}_i$'s, and thus, the (P_1) estimator automatically reveals the (un-)reliable sensors as those having (non-)zero $\mathbf{o}_i$'s.

Interestingly, when $\lambda \to \infty$, the optimum is attained at $\mathbf{o} = \mathbf{0}$; hence, all sensors are deemed consistent and (P_1) reduces to the LS estimator. On the contrary, when $\lambda \to 0^+$, the minimizing $\mathbf{o}_i$'s are $\{\mathbf{o}_i = \mathbf{b}_i - \mathbf{A}_i\mathbf{x}\}_{i=1}^k$ where $\mathbf{x}$ is the minimizer of (P_1) and all sensors are considered outliers (cf. the subsequent analysis). In the latter case, (P_1) becomes equivalent to the problem

$$\min_{\mathbf{x}} \sum_{i=1}^k \|\mathbf{b}_i - \mathbf{A}_i\mathbf{x}\|_2. \tag{4}$$

In [9], [10], the optimization problem in (4) has been developed for the RS task when the consistent measurements do not contain noise, or practically, in the high SNR regime. We also provide necessary and sufficient conditions on $\mathbf{A}$ for the identifiability of the problem, and show that when $\mathbf{A}$ is drawn from the Gaussian ensemble the probability of exact $\mathbf{x}_o$ recovery is overwhelmingly high.

To better understand the optimization problem (P_1) and to later develop an efficient solver, we study the form of its minimizer(s). Let $[(\mathbf{x}^\star)^T\ (\mathbf{o}^\star)^T]^T$ denote a minimizer of (P_1), and define the associated residual vector $\mathbf{r}^\star := \mathbf{b} - \mathbf{A}\mathbf{x}^\star$. Given $\mathbf{x}^\star$, the vectors $\{\mathbf{o}_i^\star\}_{i=1}^k$ in (P_1) can be found separately as the minimizers of

$$\min_{\mathbf{o}_i} \ \phi(\mathbf{o}_i) \quad \text{s.t.} \ \phi(\mathbf{o}_i) := \frac{1}{2}\|\mathbf{r}_i^\star - \mathbf{o}_i\|_2^2 + \lambda\|\mathbf{o}_i\|_2,\ i = 1,\ldots,k. \tag{5}$$

Although $\phi(\mathbf{o}_i)$ is not everywhere differentiable, its subdifferential $\partial\phi(\mathbf{o}_i)$ can be defined as [2]

$$\partial\phi(\mathbf{o}_i) := \begin{cases} \mathbf{o}_i\left(1+\frac{\lambda}{\|\mathbf{o}_i\|_2}\right) - \mathbf{r}_i^\star & , \ \mathbf{o}_i \neq \mathbf{0} \\ \{\lambda\mathbf{g}_i - \mathbf{r}_i^\star : \ \|\mathbf{g}_i\|_2 \leq 1\} & , \ \mathbf{o}_i = \mathbf{0}. \end{cases} \quad (6)$$

Vector $\mathbf{o}_i^\star$ is a minimizer of (5) if and only if $\mathbf{0} \in \partial\phi(\mathbf{o}_i^\star)$. If $\mathbf{o}_i^\star \neq \mathbf{0}$, the first order optimality condition yields $\mathbf{o}_i^\star = \mathbf{r}_i^\star\left(1-\frac{\lambda}{\|\mathbf{r}_i^\star\|_2}\right)$ which holds only if $\|\mathbf{r}_i^\star\|_2 > \lambda$ [9]. On the other hand, the minimizer is $\mathbf{o}_i^\star = \mathbf{0}$ when $\|\mathbf{r}_i^\star\|_2 \leq \lambda$. The latter proves that (P_1) indeed admits a block-sparse minimizer $\mathbf{o}^\star$.

Substituting $\mathbf{o}_i^\star$ into (5), yields $\phi(\mathbf{o}_i^\star) = \|\mathbf{r}_i^\star\|_2^2/2$, when $\|\mathbf{r}_i^\star\|_2 \leq \lambda$; and $\phi(\mathbf{o}_i^\star) = \lambda\|\mathbf{r}_i^\star\|_2 - \lambda^2/2$, otherwise. Having minimized (P_1) over the $\mathbf{o}_i$'s, the minimizer $\mathbf{x}^\star$ can then be found as

$$\min_{\mathbf{x}} \quad \sum_{i=1}^{k} \rho_v(\mathbf{b}_i - \mathbf{A}_i\mathbf{x}) \quad (7a)$$

$$\text{s.t.} \quad \rho_v(\mathbf{r}_i) := \begin{cases} \frac{1}{2}\|\mathbf{r}_i\|_2^2 & , \ \|\mathbf{r}_i\|_2 \leq \lambda \\ \lambda\|\mathbf{r}_i\|_2 - \frac{\lambda^2}{2} & , \ \|\mathbf{r}_i\|_2 > \lambda \end{cases} \quad (7b)$$

It is now evident that (P_1) is equivalent to (7), which rather surprisingly turns out to be a generalization of Huber's M-estimator to the vector case [8], [12]; see also [7] for the scalar $(m = 1)$ case. The sensors capable of achieving a lower $\|\mathbf{r}_i\|_2$ value, and are more likely to be reliable, are treated by (7) through the conventional LS criterion. But the sensors having $\|\mathbf{r}_i\|_2 > \lambda$, contribute $(\lambda\|\mathbf{r}_i\|_2 - \lambda^2/2) < \|\mathbf{r}_i\|_2^2/2$ to the cost, and are deemed "less important" in specifying $\mathbf{x}$. For the latter set of sensors, $\mathbf{o}_i^\star \neq \mathbf{0}$ holds too. Thus, (P_1) not only estimates the unknown vector $\mathbf{x}$, but also reveals the sensors most likely to be unreliable in the presence of noise.

A heuristic rule of thumb for practically selecting λ is setting it to $\tau\sqrt{m}$, where τ is the equivalent parameter for the scalar Huber case. Regarding τ, when the outliers' distribution is known a-priori, its value is available in closed form so that Huber's M-estimator is optimal in a well-defined sense; see [8], [12]. Alternatively, assuming that the noise is standard Gaussian, τ is usually set to $\tau = 1.34$ such that the Huber M-estimator derived is 95% asymptotically efficient at the normal distribution [8], [12]. To render Huber's M-estimator invariant to any noise variance σ^2, one has to multiply τ by σ. If σ is unknown, a robust estimate of it is commonly used instead [8], [12].

Alternatively, If the number of reliable sensors is roughly known (e.g., based on prior operation of the network), another approach is solving (P_1) for a grid of λ values and selecting the one identifying the prescribed number of outliers. Note that solving (P_1) for several values of λ can be efficiently performed either through the group-LARS algorithm [15], or, by using the block coordinate descent algorithm presented next with what is called "warm startup" [6]. The latter initializes the tentative solutions of (P_1) for a grid value of λ with the solution derived for the previous grid value of λ. The computational efficiency of such an approach has been numerically verified for the Lasso problem [6], [13].

Remark 1. The models in (1) (2) assumed the noise term to be independent. Specifications such as the geographical distribution of sensors may impose correlation across different sensor readings. In this case, if the covariance matrix $\mathbf{\Sigma}$ of the aggregate noise vector $\mathbf{n}$ is known, a standard preprocessing step is to prewhiten the data as $\mathbf{b}' := \mathbf{\Sigma}^{-1/2}\mathbf{b}$ and $\mathbf{A}' := \mathbf{\Sigma}^{-1/2}\mathbf{A}$. Prewhitening "spreads" the influence of unreliable sensors across the entries of $\mathbf{b}'$. As a result, the LS and ℓ_1-error regression estimators and even the robust Huber M-estimator are not applicable; see also [7] for similar observations in the scalar case $(m = 1)$. On the contrary, given that $\mathbf{o}$ remains block sparse, the (P_1) estimator can successfully handle a colored noise setup by simply modifying its cost as

$$\min_{\mathbf{x},\mathbf{o}} \frac{1}{2}\|\mathbf{b}' - \mathbf{A}'\mathbf{x} - \mathbf{\Sigma}^{-1/2}\mathbf{o}\|_2^2 + \lambda\sum_{i=1}^{k}\|\mathbf{o}_i\|_2. \quad (8)$$

IV. A Non-Convex Approximation

To derive (P_1), the non-smooth indicator function $\mathrm{I}(\|\mathbf{o}_i\|_2 \neq \mathbf{0})$ was approximated by its closest convex approximation, i.e., $\|\mathbf{o}_i\|_2$. However, by letting the surrogate function to be non-convex, tighter approximations are possible. In CS for example, the indicator function $\mathrm{I}(x \neq 0)$ for $x \in \mathbb{R}$ has been surrogated by $\log(|x| + \delta)$ for a small $\delta > 0$ [3]. Building on this idea, the optimization in (3) can be replaced by

$$\min_{\mathbf{x},\mathbf{o}} \frac{1}{2}\|\mathbf{b} - \mathbf{A}\mathbf{x} - \mathbf{o}\|_2^2 + \lambda\sum_{i=1}^{k}\log(\|\mathbf{o}_i\|_2 + \delta) \quad (P_2)$$

which is a non-convex problem and its minimization is non-trivial. In the quest of local optimization methods, the majorization-minimization (MM) approach is followed. The concavity of the logarithm implies that a linearization of $\log(x + \delta)$ around any $x^{(0)} > 0$ serves as a majorizer, i.e.,

$$\log(x + \delta) \leq \log\left(x^{(0)} + \delta\right) + \frac{x - x^{(0)}}{x^{(0)} + \delta}.$$

Thus, given any $\mathbf{o}^{(t)}$, the cost in (P_2) is majorized by

$$\frac{1}{2}\|\mathbf{b} - \mathbf{A}\mathbf{x} - \mathbf{o}\|_2^2 + \lambda\sum_{i=1}^{k}\frac{\|\mathbf{o}_i\|_2}{\|\mathbf{o}_i^{(t)}\|_2 + \delta}.$$

Following the MM rationale and letting $(\mathbf{x}^{(t)}, \mathbf{o}^{(t)})$ be the tentative estimates at the t-th iteration, (P_2) can be driven to a stationary point [11] by updating $(\mathbf{x}^{(t+1)}, \mathbf{o}^{(t+1)})$ as the minimizers of

$$\min_{\mathbf{x},\mathbf{o}} \quad \frac{1}{2}\|\mathbf{b} - \mathbf{A}\mathbf{x} - \mathbf{o}\|_2^2 + \lambda\sum_{i=1}^{k} w_i^{(t)}\|\mathbf{o}_i\|_2, \quad (9)$$

$$w_i^{(t)} := \left(\|\mathbf{o}_i^{(t)}\|_2 + \delta\right)^{-1}, \ i = 1, \ldots, k.$$

Interestingly, a single iteration of (9) is a weighted version of (P_1). When the residual error of a sensor is small, the sensor becomes more influential at the minimization of the next iteration. Iterations can be initialized by the solution of (P_1), while simulation results indicate that even one additional iteration can significantly improve the mean-square error (MSE) performance of (P_2) over (P_1).

V. A Block Coordinate Descent Algorithm

The optimization problem in (P_1) can be cast as a SOCP, and solved by generic interior point-based solvers. To leverage the special structure of the cost and offer computational savings, a block coordinate descent algorithm is developed. Specifically, (P_1) can be solved by iteratively minimizing with respect to $\mathbf{x}$ while keeping $\mathbf{o}$ fixed, and vice-versa. At the l-th iteration and for a fixed $\mathbf{o} = \mathbf{o}^{(l-1)}$, the solution becomes available in closed form as

$$\mathbf{x}^{(l)} := (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T(\mathbf{b} - \mathbf{o}^{(l-1)}) \tag{10}$$

which is a LS estimator on the outlier-compensated data $(\mathbf{b} - \mathbf{o}^{(l-1)})$. Then, by setting $\mathbf{x} = \mathbf{x}^{(l)}$ and

$$\mathbf{r}_i^{(l)} := \mathbf{b}_i - \mathbf{A}_i\mathbf{x}^{(l)} \tag{11}$$

the cost is separable over the $\mathbf{o}_i$'s and the minimizers are provided again in closed form as

$$\mathbf{o}_i^{(l)} = \mathbf{r}_i^{(l)} \left[1 - \frac{\lambda}{\|\mathbf{r}_i^{(l)}\|_2}\right]_+ \tag{12}$$

for $i = 1, \ldots, k$ and where $[x]_+ = \max\{0, x\}$. The block-coordinate descent algorithm consists of the updates (10), (11), and (12), and it is initialized at $\mathbf{o}^{(0)} = \mathbf{0}$. The convergence of this iterative algorithm to the global optimum of (P_1) follows readily from [14]. Its complexity can be as low as $\mathcal{O}(kmn)$ per iteration, while the simulated tests demonstrate that it is faster than interior point-based algorithms. The algorithm can be readily applied to the iterations in (9) by simply replacing λ by $\lambda_i^{(l)} := \lambda w_i^{(l)}$.

VI. Simulated Results

To evaluate the performance of the RS solvers developed, a setup involving a network of $k = 16$ sensors, $m = 4$ per sensor measurements, and an unknown vector of dimension $n = 20$, was simulated. The entries of $\mathbf{A}$ were independently drawn from $\mathcal{N}(0, 1)$. The reliable sensors followed the model (1) for $\mathbf{x}_o = n^{-1/2}\mathbf{1}_n$ and $\mathbf{n}_i \sim \mathcal{N}(\mathbf{0}, 0.1 \cdot \mathbf{I}_m)$, whereas the unreliable ones were drawn as $\{\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, 1.1 \cdot \mathbf{I}_m)\}_{i \notin \mathcal{S}_o}$. The MSE, $\mathbb{E}\left[\|\mathbf{x}_o - \hat{\mathbf{x}}\|_2^2\right]$, of each method was empirically estimated by averaging over 1,000 Monte Carlo experiments.

The comparison included: (i) the LS estimator; (ii) a genie-aided (GA) LS estimator which knows a-priori the subset of reliable sensors $\mathcal{S}_o$; (iii) the ℓ_1-error regression estimator; (iv) the Huber M-estimator [8]; (v)-(vi) the two solvers developed in [9] called here (W-)MSN; (vii) the estimator defined by (P_1); and (viii) the (P_2) estimator run for a single iteration. The parameter λ for both (P_1)-(P_2) was set using the first rule of thumb described in Section III, and δ was set to 10^{-4} even though the solution of (P_2) was insensitive in the range of values 10^{-2} down to 10^{-8} tested.

In Fig. 2 the MSE is plotted versus the number of consistent sensors s. The curves show that the block-sparsity-ignorant LS, ℓ_1, and Huber's estimators are generally outperformed by the novel schemes. The (P_2) estimator initialized at the solution of (P_1) clearly combines robustness when outliers are present with efficiency when they are absent.

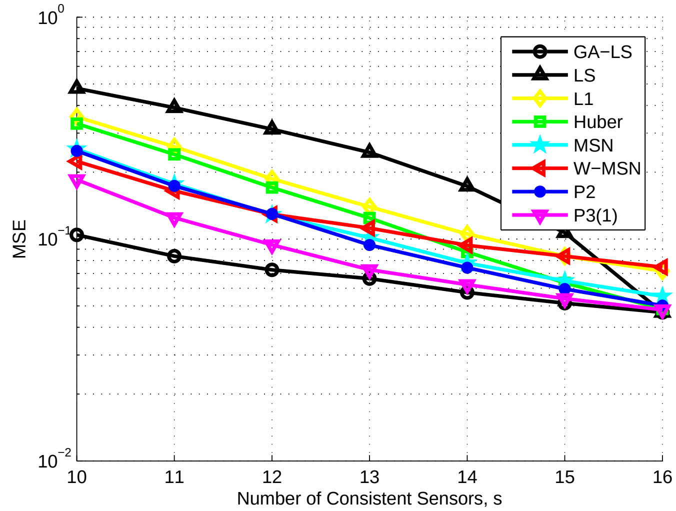

Fig. 2. MSE performance of RS solvers for white noise with $(n, m, k) = (20, 4, 16)$

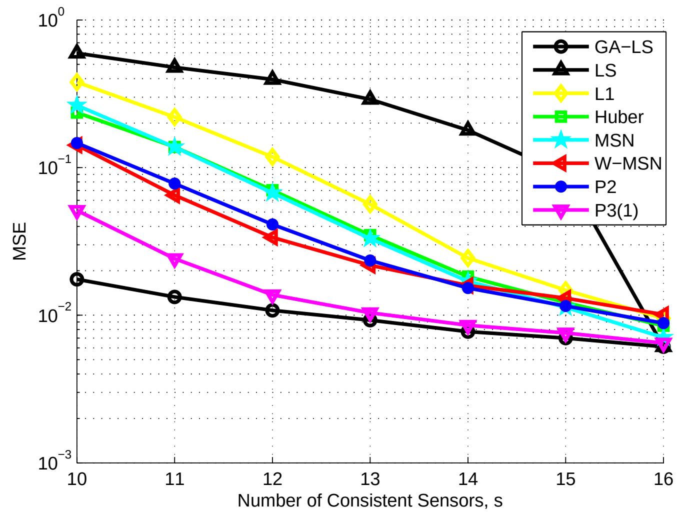

Fig. 3. MSE performance of RS solvers for colored noise with $(n, m, k) = (20, 4, 16)$.

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

VII. Conclusions

The fresh viewpoint offered here broadens the scope of sparsity-exploiting algorithms to settings where model mismatch induced by unreliable sensors or outliers gives rise to sparsity in an appropriately defined residual domain and not necessarily in the signal of interest domain. This perspective bridges compressive sampling and sparse linear regression with robust multivariate linear regression. Leveraging this connection, robust sensing algorithms were developed to reveal unreliable sensors and recover the signal of interest based on reliable sensors. The RS task was reformulated to a combinatorially complex 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 the success of the methods proposed.

References

[1] 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.

[2] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY: Cambridge University Press, 2004.

[3] E. J. Candès, M. B. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, Dec. 2008.

[4] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[5] S. S. Chen, D. L. Donoho, Michael, and A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1998.

[6] J. Friedman, T. Hastie, and R. Tibshirani, "Regularized paths for generalized linear models via coordinate descent," *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010.

[7] J.-J. Fuchs, "An inverse problem approach to robust regression," in *Proc. ICASSP*, Phoenix, AZ, 1999, pp. 1809–1812.

[8] P. J. Huber and E. M. Ronchetti, *Robust Statistics*. New York: Wiley, 2009.

[9] V. Kekatos and G. B. Giannakis, "From sparse signals to sparse residuals for robust sensing," Submitted May 2010; Revised in Nov 2010; Available online as `arXiv:1011.0450`.

[10] ——, "Selecting reliable sensors via convex optimization," in *Proc. SPAWC*, Marrakech, Morocco, Jun. 2010.

[11] K. Lange, D. Hunter, and I. Yang, "Optimization transfer using surrogate objective functions (with discussion)," *Journal of Computational and Graphical Statistics*, vol. 9, pp. 1–59, 2000.

[12] R. A. Maronna, R. D. Martin, and V. J. Yohai, *Robust Statistics: Theory and Methods*. Wiley, 2006.

[13] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society, Series B.*, vol. 58, no. 1, pp. 267–288, 1996.

[14] P. Tseng, "Convergence of block coordinate descent method for nondifferentiable minimization," *Journal on Optimization Theory and Applications*, vol. 109, pp. 475–494, Jun. 2001.

[15] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society, Series B.*, vol. 68, no. 1, pp. 49–67, 2006.

[16] Y. Zhang, N. Meratnia, and P. Havinga, "Outlier detection techniques for wireless sensor networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 12, no. 2, pp. 159–170, 2010.