SELECTING RELIABLE SENSORS VIA CONVEX OPTIMIZATION

Vassilis Kekatos and Georgios B. Giannakis

Dept. of Electrical & Computer Engrg.
University of Minnesota
Minneapolis, MN 55455, USA
Emails:{kekatos,georgios}@umn.edu

ABSTRACT

One of the key challenges in sensor networks is the extraction of trusted and relevant information by fusing data from a multitude of heterogeneous, distinct, but possibly unreliable or irrelevant sensors. Recovering the desirable view of the environment from the maximum number of dependable sensors while specifying the unreliable ones is an issue of paramount importance for active sensing and robust operation of the entire network. This problem of robust sensing is formulated here, and proved to be NP-hard. In the quest of sub-optimum but practically feasible solutions with quantifiable performance guarantees, two algorithms are developed for selecting reliable sensors via convex programming. The first relies on a convex relaxation of the original problem, while the second one is based on approximating the initial objective function by a concave one. Their performance is tested analytically, and through simulations.

1. 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.

This so-termed layered sensing apparatus entails three operational conditions: (c1) Each node's measurement vector comprising either scalar observations across time, or a snapshot of different sensor readings, is typically linearly related to the unknown variable. Such a *linear* model can arise when the sensing system is viewed as a linear filter with a known impulse response, or when a complex phenomenon is linearly represented over a fixed basis; (c2) Either because readings are costly to sense and transmit, or due to delay constraints, or simply because dimensionality reduction is invoked to cope with the "curse of dimensionality," the linear model is oftentimes *under-determined*, i.e., the dimension of the unknown variable is larger than the dimension of a single sensor observation; (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. A joint detection and parameter estimation task based on reliable sensor data, henceforth referred to as *robust (active) sensing* (RS), should be performed instead. Discerning the unreliable sensors promises higher estimation accuracy, and 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.

One possible approach to addressing this RS challenge is to formulate it as a robust estimation problem [10], [15]. Unfortunately, such a formulation treats each linear equation separately, and ignores the per-sensor structure of the problem. Alternatively, one could approach RS as a *sensor selection* [11], or, as a D/A/E-optimal experimental design problem [3, Sec. 7.5]. However, both approaches refer to the design phase of a sensor network (experiment), and are based on suitably defined criteria to choose a subset of predetermined sensors (inputs), without having acquired any observations.

The present contribution first shows that the RS task can be formulated and proved equivalent to an NP-hard problem. Then it develops two (sub-) optimum yet computationally affordable solvers. The first one relies on a convex relaxation of the original NP-hard problem, which can be efficiently solved as a second-order cone program (SOCP). It is also shown that for measurement matrices drawn from the Gaussian ensemble, and under reasonable conditions on the problem dimensions, the SOCP approach is exact almost surely. In the second method, the original objective function is surrogated by a concave function, which is (locally) minimized through a sequence of weighted SOCPs.

This research was supported by the Marie Curie Intl. Outgoing Fellowship "DIESIS"-No. 234914 within the 7th European Community Framework Programme.

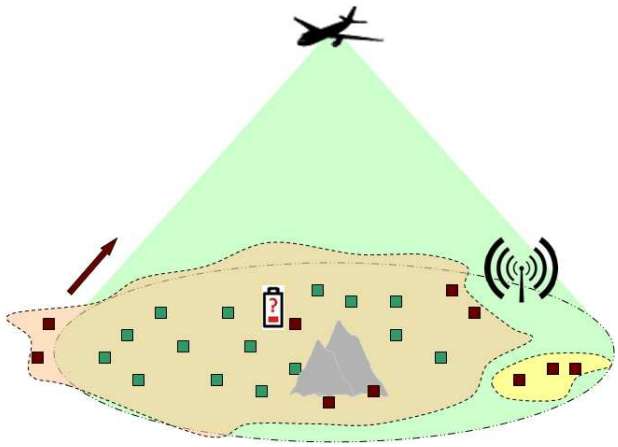

Fig. 1. A wireless sensor network linked with a fusion center (UAV). (Un)reliable sensors are color coded as (red) green.

2. PROBLEM STATEMENT AND PRELIMINARIES

Consider an agent, e.g., an unmanned aerial vehicle (UAV), collecting data vectors $\{\mathbf{b}_i = \mathbf{A}_i\mathbf{x}\}_{i=1}^k$ from k sensors, where the $m_i \times n$ matrices are known, while $\mathbf{x} \in \mathbb{R}^n$ denotes the unknown vector of interest. The task of estimating $\mathbf{x}$ is challenged by the fact that certain subsystems of equations are inconsistent in the sense that a common $\mathbf{x}$ satisfies only an *unknown subset* of sensors. The *goal* of robust sensing is to obtain an $\mathbf{x}$ satisfying the maximum number of linear subsystems of equations.

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 (e.g., in the 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 which is captured in the matrices $\{\mathbf{A}_i\}_{i=1}^k$, and is measured by a wireless sensor network deployed over the region of interest.

Suppose without loss of generality that $m_i = m$ and $\text{rank}(\mathbf{A}_i) = \min\{m, n\}$, for all $i = 1, \ldots, k$. The under- and over-determined cases are considered separately. In the over-determined case $(m \geq n)$, each one of the linear systems $\mathbf{b}_i = \mathbf{A}_i\mathbf{x}$ admits a unique solution when $\mathbf{b}_i \in \text{range}(\mathbf{A}_i)$; otherwise, it is not feasible. Thus, the infeasible subsystems can be discarded, and the vector simultaneously satisfying most of the consistent subsystems is selected readily. However, in the under-determined case $(m < n)$, every subsystem admits an infinite number of feasible solutions, and only the fusion center can in principle recover $\mathbf{x}$. The focus of this work is on the under-determined case, which apart from being more challenging, it adheres to stringent power, bandwidth, delay, or, stationarity constraints. Irrelevant or unreliable systems may arise due to obstruction, fading propagation effects, sensor failures, jamming, or, even because sensors collect data corresponding to an irrelevant $\mathbf{x}' \neq \mathbf{x}$; see Fig. 1.

Using an auxiliary vector $\mathbf{t} \in \mathbb{R}^k$ and letting $\|\mathbf{t}\|_0$ denote its ℓ_0-(pseudo)norm, the RS problem can be formulated as

$$\min_{\mathbf{x},\mathbf{t}} \{\|\mathbf{t}\|_0 : \|\mathbf{b}_i - \mathbf{A}_i\mathbf{x}\|_2 = t_i,\ i = 1, \ldots, k\}. \quad (1)$$

If the i-th subsystem is feasible, then $t_i = 0$; otherwise, t_i is strictly positive and the cost increases. The problem (1) is equivalent to

$$\min_{\mathbf{x},\mathbf{t}} \{\|\mathbf{t}\|_0 : \|\mathbf{b}_i - \mathbf{A}_i\mathbf{x}\|_2 \leq t_i,\ i = 1, \ldots, k\}, \quad (P_0)$$

which entails convex constraints, but remains non-convex.

Supposing that there exist at least s reliable sensors, a naive approach to tackle (P_0) would be to check the feasibility of all the $\binom{k}{s}$ linear systems formed by s out of k subsystems. But this approach incurs combinatorial complexity, and can only be computationally feasible for small-size networks. Unfortunately, the following result holds.

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

Proof. Consider first the related to RS problem of consistent linear equations maximization (CLEM): *"Given a system of linear equations* $\mathbf{C}\mathbf{x} = \mathbf{d}$*, where* $\mathbf{C} \in \mathbb{R}^{p\times n}$ *and* $\mathbf{d} \in \mathbb{R}^p$*, find a vector* $\mathbf{x} \in \mathbb{R}^n$ *satisfying as many equations as possible."* The CLEM problem is known to be NP-hard [1, Th. 1]. Now notice that CLEM is a special case of RS for $m = 1$ and $k = p$. Hence, if there were a deterministically polynomial-time algorithm for solving (P_0), this algorithm could also solve CLEM. This contradicts the NP-hardness of CLEM, and establishes the proposition. □

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

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

where $\mathbf{b}^T := \left[\mathbf{b}_1^T\ \ldots\ \mathbf{b}_k^T\right]$, and $\mathbf{A}^T := \left[\mathbf{A}_1^T\ \ldots\ \mathbf{A}_k^T\right]$. Alternatively, one could consider a robust estimation method based on the ℓ_1-regression

$$\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_1 \quad (3)$$

or, even rely on robust M-estimators [10], [15]. Unfortunately, all these approaches handle separately every linear equation, ignore the underlying per-sensor linear subsystem; and more importantly, they cannot reliably identify the unreliable sensors. Note also that (P_0) is reminiscent of the compressive sampling (CS) problem [7, 5]; but neither $\mathbf{x}$ here has to be sparse, nor the constraints have to be linear.

3. A CONVEX RELAXATION

It is known that if $\|\mathbf{t}\|_\infty \leq 1$, then $\|\mathbf{t}\|_1$ is the convex envelope (a.k.a. the largest convex under-estimator) of $\|\mathbf{t}\|_0$; see e.g., [3, p. 119]. This has been used for compressed sensing in [17], and prompted us to relax the NP-hard problem (P_0) to

$$\min_{\mathbf{x},\mathbf{t}} \{\|\mathbf{t}\|_1 :\ \|\mathbf{b}_i - \mathbf{A}_i\mathbf{x}\|_2 \leq t_i,\ i = 1,\ldots,k\}. \quad (4)$$

The latter is a second-order cone program (SOCP) in the standard form, and can be efficiently implemented by several existing solvers [3]. By exploiting the implicit constraint $\mathbf{t} \geq \mathbf{0}$, the problem (4) is equivalent to

$$\min_{\mathbf{x}} \sum_{i=1}^{k} \|\mathbf{b}_i - \mathbf{A}_i\mathbf{x}\|_2 \quad (P_1)$$

which is still an SOCP, albeit unconstrained.

The cost is the sum of the ℓ_2-norms of the residual vectors associated with the linear subsystems, it is continuous, but non-smooth (almost everywhere differentiable). This problem is known in the optimization parlance as the *minimization of the sum of (Euclidean) norms* (MSN) problem [3, Sec. 6.4]. It has served as a heuristic for optimizing Steiner trees, locations, as well as in total-variation image restoration problems; see e.g., [14, Sec. 2.2], and the references therein. Algorithmically, it can be solved either by generic SOCP solvers [14], or, by customized interior-point algorithms [2], [6].

The following remarks are now in order.

Remark 1. Interestingly, the LS problem in (2) can be rewritten as

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

which is again a convex approximation of (P_0), though not the closest one.

Remark 2. A problem equivalent to (P_1) can be obtained by simply squaring its objective function:

$$\min_{\mathbf{x}} \|\mathbf{b}-\mathbf{A}\mathbf{x}\|_2^2 + 2\sum_{i=1}^{k}\sum_{j=i+1}^{k} \|\mathbf{b}_i-\mathbf{A}_i\mathbf{x}\|_2\|\mathbf{b}_j-\mathbf{A}_j\mathbf{x}\|_2. \quad (5)$$

This last form reveals that (P_1) actually minimizes the conventional LS error (2) *regularized* by the sum of the products of Euclidean error norms between subsystem pairs.

Remark 3. In the degenerate case $m = 1$, where every subsystem reduces to a single equation, (P_1) becomes identical to a regression problem minimizing the ℓ_1-norm of the error, which is known to be robust against outliers [3], [10].

Remark 4. Assuming again $m = 1$ and additionally $k > n$, problem (P_1) can be equivalently expressed as

$$\min_{\mathbf{t}} \{\|\mathbf{t}\|_1 :\ \mathbf{C}\mathbf{t} = \mathbf{C}\mathbf{b}\} \quad (6)$$

where $\mathbf{C} \in \mathbb{R}^{(k-n)\times k}$ such that $\mathbf{C}\mathbf{A} = \mathbf{0}$. Problem (6) is in the form of a basis pursuit (BP) scheme [5].

Remark 5. For the generic $m \geq 1$ case, we may define the residual error vectors $\mathbf{r}_i := \mathbf{b}_i - \mathbf{A}_i\mathbf{x}$ for $i = 1,\ldots,k$ and the collective vector $\mathbf{r}^T := [\mathbf{r}_1^T\ \ldots\ \mathbf{r}_k^T]$, and rewrite (P_1) as

$$\min_{\mathbf{r}} \left\{ \sum_{i=1}^{k} \|\mathbf{r}_i\|_2 :\ \mathbf{C}\mathbf{r} = \mathbf{C}\mathbf{b} \right\} \quad (7)$$

where $\mathbf{C} \in \mathbb{R}^{(km-n)\times km}$ such that again $\mathbf{C}\mathbf{A} = \mathbf{0}$. The problem in (7) has been proposed for the reconstruction of a *block-sparse* signal $\mathbf{r}$, i.e., a signal which is partitioned in predetermined blocks that are likely to be zero, through incomplete linear measurements; see [16] and the references therein. For the RS task, each block corresponds to the residual error vector of a sensor.

Having replaced the NP-hard (P_0) with the convex (P_1), a legitimate question is whether the latter can provide a solution identical to the former. Apparently, due to the NP-hardness of (P_0), the two solvers cannot be equivalent for every instance $(\mathbf{b}, \mathbf{A})$ of the problem. Surprisingly though, by using concentration of measure tools, we have proved the following theorem which quantifies the probability of exactness for Gaussian measurement matrices [12].

Theorem 1. *Consider the matrix $\mathbf{A} \in \mathbb{R}^{km\times n}$ with entries drawn independently from $\mathcal{N}(0,1)$, and suppose there exist at least s consistent subsystems. If*

$$r := \frac{s}{k} > \frac{\sqrt{\gamma}+1}{2} \quad (8)$$

where $\gamma := n/(km)$, there exist positive $c_1(r,\gamma)$ and $c_2(r,\gamma)$ such that whenever $m \geq c_2(r,\gamma) r \log\left(\frac{e}{r}\right)$, the solution of (P_1) is identical to the solution of (P_0) with probability exceeding $1 - e^{-c_1(r,\gamma)km+o(km)}$.

By using the probability bound of this theorem and the Borel-Cantelli lemma [9], the next result follows readily.

Corollary 1. *Whenever the quadruplet (n,m,k,s) satisfies the conditions of Theorem 1, the (P_1) minimization problem recovers the (P_0) solution almost surely as $n \to \infty$.*

Two comments on the condition (8) are worth mentioning. Firstly, the condition in (8) implies that $s > k/2$, meaning that the reliable sensors should be strictly more than the inconsistent ones for the relaxation to yield the exact solution with high probability. Secondly, the inequalities $\sqrt{\gamma} < 2r - 1 < \sqrt{r}$ for $r \in (0.5, 1)$ lead to the requirement $km > sm > n$, which implies that not only the initial linear system $\mathbf{b} = \mathbf{A}\mathbf{x}$, but also the linear system eventually obtained after discarding the irrelevant subsystems should be over-determined.

4. A CONCAVE APPROXIMATION

In the previous section, the nonconvex cost $\|\mathbf{t}\|_0$ was replaced by its closest convex approximation, namely $\|\mathbf{t}\|_1$. However,

by letting the surrogate function to be nonconvex, tighter approximations are possible. For example, the ℓ_0-norm of a vector $\mathbf{x} \in \mathbb{R}^n$ is surrogated by the logarithm of its geometric mean, that is $\sum_{i=1}^n \log(|x_i|)$ [4]. Likewise, in the context of the matrix rank minimization problem (RMP), $\mathrm{rank}(\mathbf{X})$ for $\mathbf{X} \in \mathcal{S}_+^n$ is replaced by $\log\det(\mathbf{X} + \delta\mathbf{I})$ for a small $\delta > 0$ [8, Sec. 5.2]. Building on this idea, the problem (P_0) can be surrogated by

$$\min_{\mathbf{x},\mathbf{t}} \ \sum_{i=1}^{k} \log(t_i + \delta) \qquad (P_2)$$
$$\text{s.t. } \|\mathbf{b}_i - \mathbf{A}_i\mathbf{x}\|_2 \leq t_i, \ i = 1, \ldots, k$$

where δ is a small positive constant introduced to avoid numerical instability.

The new problem (P_2) is concave, and its minimization is thus nontrivial. However, due to the smoothness of the logarithmic function in $\mathbb{R}_+^n$, local optimization methods can be employed. Specifically, given an initial point $\left(\mathbf{x}^{(0)}, \mathbf{t}^{(0)}\right)$, iterative linearization will lead to a local minimum. The concavity of the logarithm implies that the first-order approximation of $\log(t_i + \delta)$ around $t_i^{(0)}$ is upper bounded by

$$\log(t_i + \delta) \leq \log\left(t_i^{(0)} + \delta\right) + \frac{1}{t_i^{(0)} + \delta}\left(t_i - t_i^{(0)}\right). \quad (9)$$

By viewing (9) under the majorization - minimization (MM) framework [13], instead of minimizing $\log(t_i + \delta)$, it is possible to minimize its *majorizing* cost function on the right-hand side of (9), and then iterate. Thus, the problem (P_2) can be iteratively driven to a (local) minimum as

$$\left(\mathbf{x}^{(l)}, \mathbf{t}^{(l)}\right) := \arg\min_{\mathbf{x},\mathbf{t}} \ \sum_{i=1}^{k} \frac{t_i}{t_i^{(l-1)} + \delta}$$
$$\text{s.t. } \|\mathbf{b}_i - \mathbf{A}_i\mathbf{x}\|_2 \leq t_i, \ i = 1, \ldots, k,$$

or equivalently,

$$\mathbf{x}^{(l)} := \arg\min_{\mathbf{x}} \sum_{i=1}^{k} w_i^{(l)} \|\mathbf{b}_i - \mathbf{A}_i\mathbf{x}\|_2, \text{ where} \qquad (10)$$

$$w_i^{(l)} := \left(\|\mathbf{b}_i - \mathbf{A}_i\mathbf{x}^{(l-1)}\|_2 + \delta\right)^{-1}, \ i = 1, \ldots, k. \quad (11)$$

Remarkably, this iterative method involves convex optimization of the weighted cost in (10) at every iteration. 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 the (P_1) problem, which corresponds to one iteration of (10) with equal weights.

5. SIMULATED TESTS

In this section, the performance of the developed methods is evaluated through computer simulations. The simulation setup involves a network of $k = 16$ sensors with size of observation vectors $m = 4$, and an unknown vector of size $n = 20$. The unknown vector is modeled as $\mathbf{x}_o \sim \mathcal{N}(\mathbf{0}, n^{-1/2}\mathbf{I}_n)$, and the entries of matrix $\mathbf{A}$ are drawn independently from $\mathcal{N}(0, 1)$. The comparison includes: (i) the LS estimator of (2); (ii) the ℓ_1 estimator of (3); (iii) the Huber estimator; (iv) the bisquare M-estimator [15], [10]; (v) the (P_1) estimator; and (vi) the (P_2) with a single iteration. In addition, a *genie-aided* LS estimator knowing a-priori the reliable sensors serves as a benchmark. Note that for each of the two robust M-estimators, a cutoff parameter should be specified. For our experiments, the parameters were set so that the estimators are 95% efficient to the normal distribution [15]. Similar simulation tests were performed for other values.

Table 1. Empirical probability of successful recovery of the consistent sensors (%) for $n = 20$, $m = 4$, and $k = 16$.

	Number of consistent sensors s				
Method	**6**	**8**	**10**	**12**	**14**
LS-GA	100.0	100.0	100.0	100.0	100.0
LS	0.0	0.0	0.0	0.0	0.0
l_1	0.0	0.0	11.3	90.6	100.0
Huber	0.0	0.0	0.0	0.0	0.0
Bisquare	0.0	0.0	0.0	0.0	0.0
P_1	0.0	0.4	45.2	99.2	100.0
$P_2(1)$	1.0	50.3	99.0	100.0	100.0

In the first scenario, a subset of $s \in [6, 14]$ sensors have consistent observations $\mathbf{b}_i = \mathbf{A}_i\mathbf{x}_o$, while the remaining $(k - s)$ unreliable sensor data are modeled as $\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_m)$, so that $\mathbb{E}[\|\mathbf{b}_i\|_2^2] = m$ for all sensors. The reliable sensor detection probability is empirically estimated through 1,000 Monte Carlo experiments. An estimate $\hat{\mathbf{x}}$ is considered to have successfully recovered the dependable subset, whenever $\|\mathbf{b}_i - \mathbf{A}_i\hat{\mathbf{x}}\|_\infty \leq 10^{-4}$ for all consistent sensors. As evidenced by Table 1, the LS, the Huber, and the bisquare estimators fail to recover the reliable subset. The novel scheme corresponding to (P_1) has a clear advantage over the ℓ_1 estimator, while the empirical detection probability further improves for the (P_2) method, even after a single iteration.

In a second more practical scenario, there are $s = 14$ reliable sensors but their measurements do not exactly satisfy the linear equations; instead, they obey the noisy model $\mathbf{b}_i = \mathbf{A}_i\mathbf{x}_o + \mathbf{v}_i$, where the noise is distributed as $\mathbf{v}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_m)$. The vector $\mathbf{x}_o$ and the matrix $\mathbf{A}$ follow the same model as in the previous experiment. For the unreliable sensors, $\mathbf{b}_i \sim \mathcal{N}(\mathbf{0}, (1 + \sigma^2)\mathbf{I}_m)$, so that $\mathbb{E}[\|\mathbf{b}_i\|_2^2]$ remains the same for all sensors. The average signal-to-noise ratio (SNR) is defined over the reliable sensors as $10\log_{10}\sigma^{-2}$. The figure of merit in this setup is the mean-square error (MSE), $\mathbb{E}[\|\mathbf{x}_o - \hat{\mathbf{x}}\|_2^2]$, which is empirically estimated over 10,000 Monte Carlo experiments.

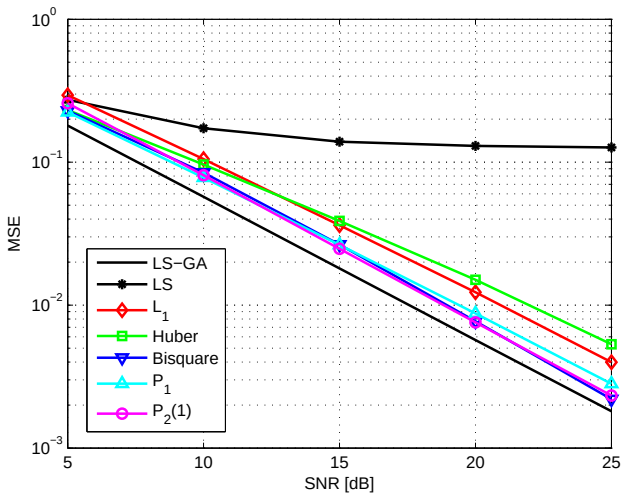

Fig. 2. MSE performance for $(n, m, k, s) = (20, 4, 16, 14)$.

The MSE curves plotted in Fig. 2 confirm that the LS estimator is not appropriate for the RS problem. The ℓ_1 and the Huber estimators belong to a second tier of MSE performance, while the (P_1), the bisquare, and the (P_2) estimators perform comparably. In the high SNR regime, the performance of (P_1) is improved by just a single re-weighting performed by (P_2). The results of Fig. 2 should be examined jointly with the probabilities of Table 1 for $s = 14$, which correspond to the asymptotic SNR regime (SNR$\to \infty$). It is evident that the proposed robust sensing methods can simultaneously attain high estimation accuracy, and high detection probability without having to tune a cutoff parameter.

6. CONCLUSIONS

This paper dealt with selecting reliable sensors, a problem inherently challenging the performance of robust active sensing applications. The quest for a joint sensor detector and parameter estimator led to an NP-hard problem. Two suboptimum alternatives were developed based on a convex relaxation and a concave approximation of the original objective function. Both methods can be efficiently implemented by SOCP solvers. Analysis and corroborating simulations establish that the novel schemes can provide high estimation accuracy and high detection probability.

7. REFERENCES

[1] E. Amaldi and V. Kann, "The complexity and approximability of finding maximum feasible subsystems of linear relations," *Theoretical Computer Science*, vol. 147, pp. 181–210, 1995.

[2] K. D. Andersen, E. Christiansen, A. R. Conn, and M. L. Overton, "An efficient primal-dual interior-point method for minimizing a sum of Euclidean norms," *SIAM J. Scient. Comp.*, vol. 22, pp. 243–262, 2000.

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

[4] 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.

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

[6] T. F. Chan, G. H. Golub, and P. Mulet, "A nonlinear primal-dual method for total variation-based image restoration," *SIAM J. Scient. Comput.*, vol. 20, no. 6, pp. 1964–1977, 1999.

[7] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Scient. Comp.*, vol. 20, pp. 33–61, 1999.

[8] M. Fazel, "Matrix rank minimization with applications," Ph.D. dissertation, Stanford University, 2002.

[9] M. Gray, *Probability, Random Processes, and Ergodic Properties*. New York, NY: Springer-Verlag, 2009.

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

[11] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Trans. Signal Processing*, vol. 57, no. 2, pp. 451–462, Feb. 2009.

[12] V. Kekatos and G. B. Giannakis, "Robust sensing via convex optimization," submitted for publication, Apr. 2010.

[13] K. Lange, D. Hunter, and I. Yang, "Optimization transfer using surrogate objective functions (with discussion)," *J. Comput. Graph. Statist.*, vol. 9, pp. 1–59, 2000.

[14] M. S. Lobo, L. Vandenberghe, H. Lebret, and S. Boyd, "Applications of second-order cone programming," *Linear Algebra and its Applications*, vol. 284, pp. 193–228, 1998.

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

[16] M. Stojnic, F. Parvaresh, and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements," *IEEE Trans. Signal Processing*, vol. 57, no. 8, pp. 3075–3085, Aug. 2009.

[17] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals," *IEEE Trans. Inform. Theory*, vol. 51, no. 3, pp. 1030–1051, Mar. 2006.