RECONSTRUCTION OF SPARSE SIGNALS FROM HIGHLY CORRUPTED MEASUREMENTS BY NONCONVEX MINIMIZATION

Marko Filipović

Rudjer Bošković Institute
Bijenička 54, 10000 Zagreb, Croatia

ABSTRACT

We propose a method for signal recovery in compressed sensing when measurements can be highly corrupted. It is based on ℓ_p minimization for $0 < p \leq 1$. Since it was shown that ℓ_p minimization performs better than ℓ_1 minimization when there are no large errors, the proposed approach is a natural extension to compressed sensing with corruptions. We provide a theoretical justification of this idea, based on analogous reasoning as in the case when measurements are not corrupted by large errors. Better performance of the proposed approach compared to ℓ_1 minimization is illustrated in numerical experiments.

Index Terms— Compressive sensing, Sparse signal reconstruction, Nonconvex optimization, Restricted Isometry

1. INTRODUCTION

Compressed sensing (CS) has been intensively studied in recent years [1]. It is based on the fact that sparse or compressible signal $x \in \mathbb{R}^n$ can be accurately reconstructed from a small number of non-adaptive linear measurements. The measurement process in CS is usually represented as $y = Ax$, where $A \in \mathbb{R}^{m \times n}$, $m < n$, is a measurement or sensing matrix. The most natural approach to reconstruct a sparse vector x from y is to solve the optimization problem

$$\min_{\hat{x}} \|\hat{x}\|_0 \quad \text{subject to} \quad A\hat{x} = y. \tag{1}$$

Here, $\| \cdot \|_0$ denotes the ℓ_0 "norm", which counts the number of non-zero elements of a vector. Unfortunately, the above problem (1) requires combinatorial optimization and is NP-hard [2]. However, it is known that, if x is sufficiently sparse and the measurement matrix obeys certain conditions, x can be recovered by solving the convex optimization problem

$$\min_{\hat{x}} \|\hat{x}\|_1 \quad \text{subject to} \quad A\hat{x} = y. \tag{2}$$

Usual condition on A is that it satisfies the restricted isometry property (RIP) [3], which means that it is approximate isometry when operating on sparse vectors. It was shown that many random matrices satisfy the RIP with high probability [4].

In recent years, a generalization of CS was considered, which was referred to as *compressed sensing with corruptions* in [5]. Its mathematical model is

$$y = Ax + f = [A\ I]\begin{bmatrix} x \\ f \end{bmatrix},$$

where $f \in \mathbb{R}^m$ is a sparse vector and I denotes the $m \times m$ identity matrix. Here, f is modelling large errors in measurements. In other words, we assume that some elements of y are arbitrarily corrupted without knowing their locations (indexes) in y. Several papers [6, 7, 8, 9] have investigated the recovery methods for this model. They considered the following problem in the noiseless case:

$$\min_{\hat{x}, \hat{f}} \|\hat{x}\|_1 + \lambda \|\hat{f}\|_1 \quad \text{subject to} \quad A\hat{x} + \hat{f} = [A\ I]\begin{bmatrix} \hat{x} \\ \hat{f} \end{bmatrix} = y, \tag{3}$$

where $\lambda > 0$ is a parameter. In [5], the noisy case was considered:

$$\min_{\hat{x}, \hat{f}} \|\hat{x}\|_1 + \lambda \|\hat{f}\|_1 \quad \text{subject to} \quad \left\|\left(A\hat{x} + \hat{f}\right) - y\right\|_2 \leq \epsilon \tag{4}$$

(here, ϵ is related to the noise level). We discuss these papers and the novelty of our approach, explained next, in Subsection 1.1.

It was demonstrated in [10] that ℓ_p minimization

$$\min_{\hat{x}} \|\hat{x}\|_p^p \quad \text{subject to} \quad A\hat{x} = y \tag{5}$$

for $0 < p < 1$ recovers sparse signals from fewer linear measurements than ℓ_1 minimization (2). Here, ℓ_p "norm" $\| \cdot \|_p$ is defined as $\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$ (x_i denotes i-th element of x). Therefore, it is natural to consider a generalization of ℓ_p minimization to compressed sensing with corruptions. This is the motivation for this paper. Following [6], we consider the following general (although noiseless) formulation of the problem:

$$y = Ax + \Omega f, \tag{6}$$

where Ω is $m \times l$ matrix, with $l \leq m$, with orthonormal columns. Here, both x and f are sparse vectors. Now, we

propose solving the following nonconvex minimization problem:

$$\min_{\hat{x},\hat{f}} \|\hat{x}\|_p^p + \lambda\|\hat{f}\|_p^p \quad \text{subject to} \quad \begin{bmatrix} A & \Omega \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{f} \end{bmatrix} = y \quad (7)$$

where $\lambda > 0$ is a parameter. By change of variable $\tilde{f} = \lambda^{1/p}\hat{f}$, it can be equivalently stated as

$$\min_{\hat{x},\tilde{f}} \left\| \begin{bmatrix} \hat{x} \\ \tilde{f} \end{bmatrix} \right\|_p^p \quad \text{subject to} \quad \begin{bmatrix} A & \lambda^{-\frac{1}{p}}\Omega \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{f} \end{bmatrix} = y.$$

Therefore, any numerical algorithm for solving (5) can also be used for solving (7).

1.1. Previous work

To the best of author's knowledge, the approach to sparse reconstruction from corrupted measurements using ℓ_p minimization, proposed here, is novel. Several papers, cited above, considered solving convex formulations (3) and (4).

In [8], the authors proposed to solve (3) with $\lambda = 1$. However, they concentrated on the problem of error correction, where generally $m \geq n$. They showed that, when m is extremely large and provided x is extremely sparse, x and f can be exactly recovered in the presence of almost any error (i.e. close to 100 percent). Their analysis is based on the assumption of Gaussianity of columns of A. Also, the results in that paper require that the sparsity of x is sublinear in m.

Paper [7] discussed the model in which A is formed by selecting rows of an orthogonal matrix. The main result states that the convex program (3) correctly recovers x with $\lambda = \sqrt{\frac{1}{\gamma \log n}\frac{n}{\mu m}}$ provided $m \geq C_1\mu^2\|x\|_0 (\log n)^2$ and $\|f\|_0 \leq C_2\gamma m$, where $a_{ij}^2 \leq \mu/n$. In other words, sparse signal can be reconstructed even when close to 100 percent of measurements are corrupted. As argued in [5], the model for A used in [7] does not include some frequently used models.

In [5], the Gaussian model for A was discussed. The main result states that the model (4) with $\lambda = 1/ \left(\sqrt{\log \frac{n}{m} + 1} \right)$ recovers x up to error proportional to the noise level, provided $\|x\|_0 \leq \alpha m/ \left(\sqrt{\log \frac{n}{m} + 1} \right)$ and $\|f\|_0 \leq \alpha m$, with probability $\geq 1 - Ce^{-cm}$. This result was proved using a generalized notion of RIP, that considers sparsities of x and f separately. We use this concept in this paper too. In [5], a general model for A where rows a of A are such that $\mathbb{E}\left(aa^T\right) = I$ and $\|a\|_\infty \leq \sqrt{\mu}$, was also studied. This model includes matrices with independent and identically distributed (i.i.d.) entries.

The measurement model (6) that we use in this paper is the same as in [6]. There, problem formulations (3) and (4) were considered. It was shown that the matrix $[A\,\Omega]$, where Ω has orthonormal columns, satisfies the RIP with high probability. However, [6] did not use the generalized notion of RIP, introduced in [5], and therefore their results are sub-optimal.

In [9], a general observation model $y = Ax + Df$ was discussed, where A and D are general matrices. The authors

presented deterministic recovery guarantees using coherence of A and D. However, deterministic guarantees are more restrictive than those discussed in the above mentioned papers.

1.2. Organization of the paper

Conditions under which the global solution of (7) is exactly x, expressed in terms of the generalized notion of RIP introduced in [5], are discussed in Section 2. A short discussion on generalized RIP of random matrices is also included there. Numerical experiments illustrating good performance of the proposed method, compared to the convex formulation (3), are described in Section 3. Conclusions are given in Section 4.

2. RESTRICTED ISOMETRY PROPERTY

We repeat the following definition of generalized RIP from [5].

Definition 1. For a matrix $\Phi \in \mathbb{R}^{l \times (n+m)}$, define the restricted isometry constant δ_{s_1,s_2} as the smallest number δ such that

$$(1-\delta)\left\| \begin{bmatrix} x \\ f \end{bmatrix} \right\|_2^2 \leq \left\| \Phi \begin{bmatrix} x \\ f \end{bmatrix} \right\|_2^2 \leq (1+\delta)\left\| \begin{bmatrix} x \\ f \end{bmatrix} \right\|_2^2 \tag{8}$$

holds for all $x \in \mathbb{R}^n$ with $|\mathrm{supp}(x)| \leq s_1$ and all $f \in \mathbb{R}^m$ with $|\mathrm{supp}(f)| \leq s_2$ (here, $\mathrm{supp}(x)$ denotes the support set of x).

The following theorem then holds.

Theorem 1. Let Φ be an $l \times (n+m)$ matrix, with $l < (n+m)$. Denote $K_1 = |\mathrm{supp}(x)|$ and $K_2 = |\mathrm{supp}(f)|$. Let $a_1 \geq 1$ and $a_2 \geq 1$ such that $a_1 K_1$ and $a_2 K_2$ are integers and $a = \min(a_1, a_2)$. Let $c \geq 1$ and b be such that

$$b = \frac{a^{\frac{1}{p}-\frac{1}{2}}}{2^{\frac{1}{p}}c^2} > 1. \tag{9}$$

If

$$\lambda \in \left[\frac{1}{c^p}\left(\frac{a_1 K_1}{a_2 K_2} \right)^{1-\frac{p}{2}}, c^p \left(\frac{a_1 K_1}{a_2 K_2} \right)^{1-\frac{p}{2}} \right]$$

and Φ satisfies

$$\delta_{a_1 K_1, a_2 K_2} + b^2 \delta_{(a_1+1)K_1, (a_2+1)K_2} < b^2 - 1, \tag{10}$$

then the unique minimizer of (7) (with $\Phi = [A\,\Omega]$) is exactly the pair (x, f).

Proof. Let us denote by $(\hat{x}, \hat{f})$ the solution pair of (7), and write

$$h = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} x \\ f \end{bmatrix} - \begin{bmatrix} \hat{x} \\ \hat{f} \end{bmatrix}.$$

Let us denote the support set of x by T_0, the support set of f (both in (in $\{1, \ldots, n+m\}$)) by V_0, and their union $T_0 \cup V_0$ by W_0. Since $(\hat{x}, \hat{f})$ is the solution of (7), we have

$$\|\hat{x}\|_p^p + \lambda\|\hat{f}\|_p^p \leq \|x\|_p^p + \lambda\|f\|_p^p. \qquad (11)$$

Also, from the triangle inequality for $\|\cdot\|_p^p$ and the fact that $\text{supp}(x) = T_0$,

$$
\begin{aligned}
\|\hat{x}\|_p^p &= \|x + h_1\|_p^p \\
&\geq \|x\|_p^p - \|h_{1_{T_0}}\|_p^p + \|h_{1_{T_0^C}}\|_p^p.
\end{aligned}
$$

Here and further, notation x_T refers to the sub-vector of x consisting of elements at indexes in the set T. Using the above, the analogous inequality for $\|\hat{f}\|_p^p$, and (11), we have

$$\|h_{1_{T_0^C}}\|_p^p + \lambda\|h_{2_{V_0^C}}\|_p^p \leq \|h_{1_{T_0}}\|_p^p + \lambda\|h_{2_{V_0}}\|_p^p.$$

The above inequality can be written as

$$\left\| \begin{bmatrix} h_{1_{T_0^C}} \\ \lambda^{\frac{1}{p}} h_{2_{V_0^C}} \end{bmatrix} \right\|_p^p \leq \left\| \begin{bmatrix} h_{1_{T_0}} \\ \lambda^{\frac{1}{p}} h_{2_{V_0}} \end{bmatrix} \right\|_p^p.$$

Now, using the reverse triangle inequality for $\|\cdot\|_p$ and the inequality $(c_1^p + c_2^p)^{\frac{1}{p}} \leq (c_1 + c_2)\, 2^{\frac{1}{p} - \frac{1}{2}}$ (which is valid for $c_1 \geq 0, c_2 \geq 0$), it follows

$$\|h_{1_{T_0^C}}\|_p + \lambda^{\frac{1}{p}}\|h_{2_{V_0^C}}\|_p \leq 2^{\frac{1}{p} - \frac{1}{2}}\left(\|h_{1_{T_0}}\|_p + \lambda^{\frac{1}{p}}\|h_{2_{V_0}}\|_p\right). \qquad (12)$$

We denote $L_1 = a_1 K_1$ and $L_2 = a_2 K_2$. Let us partition T_0^C as $T_0^C = T_1 \cup \cdots \cup T_{J_1}$, where T_1 contains L_1 largest (in absolute value) elements of $h_{1_{T_0^C}}$, T_2 contains next L_1 largest elements, and so on. Here, $|T_1| = \cdots = |T_{J_1-1}| = L_1$ and $|T_{J_1}| \leq L_1$. In the same way, we partition V_0^C as $V_0^C = V_1 \cup \cdots \cup V_{J_2}$ so that V_1 contains L_2 largest (in absolute value) elements of $h_{2_{V_0^C}}$, V_2 contains next L_2 largest elements, and so on, where $|V_1| = \cdots = |V_{J_2-1}| = L_2$ and $|V_{J_2}| \leq L_2$. We also define $W_j = T_j \cup V_j$ for $j = 1, \ldots, J = \min(J_1, J_2)$, while for $j > J$ we define $T_j = \emptyset$, i.e. $V_j = \emptyset$, depending on whether $J_2 > J_1$ or $J_1 > J_2$, respectively. The following inequalities then hold:

$$
\begin{aligned}
0 &= \left\| \Phi \begin{bmatrix} x \\ f \end{bmatrix} - \Phi \begin{bmatrix} \hat{x} \\ \hat{f} \end{bmatrix} \right\|_2 = \|\Phi h\|_2 \\
&\geq \|\Phi h_{W_0 \cup W_1}\|_2 - \sum_{j \geq 2} \|\Phi h_{W_j}\|_2 \\
&\geq \sqrt{1 - \delta_{K_1+L_1,\, K_2+L_2}}\, \|h_{W_0 \cup W_1}\|_2 - \\
&\quad \sqrt{1 + \delta_{L_1, L_2}} \sum_{j \geq 2} \|h_{W_j}\|_2. \qquad (13)
\end{aligned}
$$

$\|h_{W_j}\|_2$ needs to be controlled, for all $j \geq 2$. Using inequalities $\|h_{1_{T_j}}\|_2 \leq \|h_{1_{T_{j-1}}}\|_p / L_1^{\frac{1}{p} - \frac{1}{2}}$ and $\|h_{2_{V_j}}\|_2 \leq$

$\|h_{2_{V_{j-1}}}\|_p / L_2^{\frac{1}{p} - \frac{1}{2}}$, we have

$$
\begin{aligned}
\|h_{W_j}\|_2 &\leq \|h_{1_{T_j}}\|_2 + \|h_{2_{V_j}}\|_2 \\
&\leq \frac{c}{L_1^{\frac{1}{p} - \frac{1}{2}}}\left(\|h_{1_{T_{j-1}}}\|_p + \frac{1}{c}\left(\frac{L_1}{L_2}\right)^{\frac{1}{p} - \frac{1}{2}} \|h_{2_{V_{j-1}}}\|_p\right) \\
&\leq \frac{c}{L_1^{\frac{1}{p} - \frac{1}{2}}}\left(\|h_{1_{T_{j-1}}}\|_p + \lambda^{\frac{1}{p}} \|h_{2_{V_{j-1}}}\|_p\right)
\end{aligned}
$$

Therefore, using (12),

$$
\begin{aligned}
\sum_{j \geq 2} \|h_{W_j}\|_2 &\leq \frac{c}{L_1^{\frac{1}{p} - \frac{1}{2}}}\left(\|h_{1_{T_0^C}}\|_p + \lambda^{\frac{1}{p}} \|h_{2_{V_0^C}}\|_p\right) \\
&\leq \frac{c 2^{\frac{1}{p} - \frac{1}{2}}}{L_1^{\frac{1}{p} - \frac{1}{2}}}\left(\|h_{1_{T_0}}\|_p + \lambda^{\frac{1}{p}} \|h_{2_{V_0}}\|_p\right) \\
&\leq c\, 2^{\frac{1}{p} - \frac{1}{2}}\left(\frac{K_1}{L_1}\right)^{\frac{1}{p} - \frac{1}{2}}\left(\|h_{1_{T_0}}\|_2 + \right. \\
&\quad \left. \lambda^{\frac{1}{p}} \frac{1}{c}\left(\frac{K_2}{K_1}\right)^{\frac{1}{p} - \frac{1}{2}} \|h_{2_{V_0}}\|_2\right) \qquad (14) \\
&\leq \|h_{W_0}\|_2 / b. \qquad (15)
\end{aligned}
$$

(14) follows from generalized Hölder inequality, while (15) follows from the inequality $c_1 + c_2 \leq \sqrt{2} \cdot \sqrt{c_1^2 + c_2^2}$ (which is valid for $c_1 \geq 0, c_2 \geq 0$).

Going back to (13), we have

$$
\begin{aligned}
0 &\geq \sqrt{1 - \delta_{K_1+L_1,\, K_2+L_2}}\, \|h_{W_0 \cup W_1}\|_2 - \\
&\quad \sqrt{1 + \delta_{L_1, L_2}} \sum_{j \geq 2} \|h_{W_j}\|_2 \\
&\geq \left(\sqrt{1 - \delta_{K_1+L_1,\, K_2+L_2}} - \sqrt{1 + \delta_{L_1, L_2}}/b\right) \|h_{W_0}\|_2. \qquad (16)
\end{aligned}
$$

Because of the condition (10) of the theorem, the scalar factor is strictly positive, so that $h_{W_0} = 0$, and therefore (from (12)) $h = 0$. $\qquad \square$

The condition (10) is somewhat restrictive since (9) implies $a > 2$, which is worse than the corresponding condition in [5]. This should be the artifact of the proof since numerical experiments in Section 3 illustrate better empirical performance of ℓ_p minimization. The theorem gives an optimal value of trade-off parameter λ, however it depends on a_1 and a_2, which are related to restricted isometry constants of Φ and are generally unknown and hard to determine. Intuitively, smaller values of λ should enable recovery of very sparse signals when a large number of measurements are corrupted.

Many random matrices satisfy the condition (10) of the theorem with high probability. In the following, we suppose that, for fixed vectors $x \in \mathbb{R}^n$ and $f \in \mathbb{R}^m$, A satisfies

$$\mathbb{P}\left(\left| \|Ax\|_2^2 - \|x\|_2^2 \right| \geq \delta\|x\|_2^2\right) \leq 2e^{-c_1(\delta)m} \qquad (17)$$

$$\mathbb{P}\left(|2f^T \Omega^T Ax| \geq \delta\|x\|_2\|f\|_2\right) \leq e^{-c_2(\delta)m}, \qquad (18)$$

where $c_1(\delta)$ and $c_2(\delta)$ are constants that depend only on δ and such that $c_1(\delta) > 0$ and $c_2(\delta) > 0$ for all $\delta \in (0, 1)$. Such inequalities hold for normal and Bernoulli distribution [4, 6], but also for sub-gaussian distributions. The following lemma then holds by the same reasoning as in Lemma 5.1 in [4].

Lemma 1. *If A is a random matrix that satisfies the inequality* (18)*, then for any sets V with $|V| = s_1$ and W with $|W| = s_2$ we have $\left|2f^T\Omega^T Ax\right| \leq \delta\|x\|_2\|f\|_2$ for all $x \in \mathbb{R}^n$ with $\mathrm{supp}(x) \subseteq V$ and $f \in \mathbb{R}^m$ with $\mathrm{supp}(f) \subseteq W$, with probability $\geq 1 - \left(\frac{12}{\delta}\right)^{s_1+s_2} e^{-c_2\left(\frac{\delta}{4}\right)m}$.*

We omit the proof for brevity. Now, an analogue of the Theorem 5.2 in [4] holds. Again, the proof is omitted for the lack of space (it follows using the same approach as in [4]).

Theorem 2. *Let $\delta \in (0, 1)$. If the probability distribution generating A satisfies* (17) *and* (18)*, then there exist constants $c_3, \tilde{c}_3 > 0$ and $c_4, \tilde{c}_4 > 0$, depending only on δ, such that the matrix $\Phi = [A\ \Omega]$ satisfies the RIP (as defined in* (8)*) with constants $s_1 \leq c_3 m/\log\frac{n}{s_1}$ and $s_2 \leq c_4 m/\log\frac{m}{s_2}$ and with prescribed δ with probability $\geq 1 - \left(2e^{-\tilde{c}_3 m} + e^{-\tilde{c}_4 m}\right)$.*

The above theorem follows easily from the corresponding results from [4]; however, it gives sharper bounds because it considers sparsities of x and f separately. This was not emphasized in [6]. On the other hand, an analogous result was shown in [5], but there a different bound was used for the probability $\mathbb{P}\left(\left|2f^T\Omega^T Ax\right| > \delta\|x\|_2\|f\|_2\right)$. Bound on the sparsity of f that follows from the above theorem is more restrictive than that in [8, 7], but they considered different models. We note that the exact probability for satisfying the condition (10) for given probability distribution and parameters p, a_1 and a_2 in Theorem 1 follows from the above theorem (exact expressions are omitted for simplicity and lack of space).

We also note that, unfortunately, it seems that the approach from [10] (using the ℓ_p variant of the restricted isometry property), which would yield better bounds for smaller p, cannot be extended to compressed sensing with corruptions.

3. EXPERIMENTS

In this section we perform some empirical tests to check how many corruptions (7) can tolerate. In all numerical experiments performed here, p was set to 0.5 and (7) was solved using an iteratively reweighted least squares (IRLS) method from [11].

We set $m = 100$, $n = 500$, and we vary $s_1 = \|x\|_0$ in the range $[1, 10]$, and the number s_2 of corrupted measurements in the range $[30, 90]$. The elements of $m \times n$ matrix A are generated i.i.d. from normal distribution with mean zero and variance $1/m$. Then, the columns of A are normalized. We set $\Omega = I$. Locations of nonzero indexes of x and f are generated randomly, while the values of nonzero elements of x and f are generated i.i.d. from standard normal distribution. Noise was not added to the measurements (adding noise

would require using a noise-aware algorithm for ℓ_p minimization, which we avoid here for simplicity).

We compare IRLS with the convex approach (3). To solve (3), CVX package [12, 13] for MATLAB was used. Parameter λ in (3) and (7) was set to 0.7. Other values were also tested, but this choice gave representative result. Of course, it should be noted that better results (both for IRLS and (3)) could possibly be obtained by tuning the value of λ for every value of signal sparsity and the number of corrupted measurements, but the optimal value is hard to determine (and is generally unavailable in practice).

For every fixed s_1 and s_2, 100 repetitions were performed, every time randomly generating A, x and f. Recovery is considered successful if the reconstruction signal-to-noise ratio (SNR), defined as $-20\log_{10}\frac{\|x-\hat{x}\|_2}{\|x\|_2}$ (where $\hat{x}$ is the output of an algorithm), is above 20 dB (i.e., relative error is below 0.1). Figure 1 shows the results.
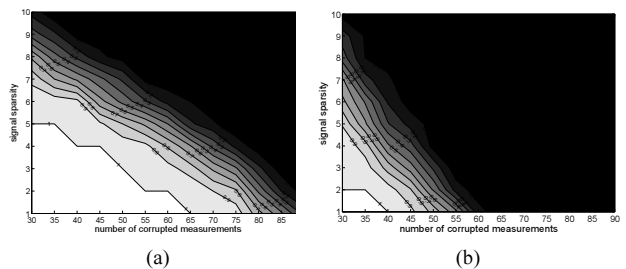

Fig. 1: The plot shows frequency of exact reconstruction (over 100 runs) by solving: (a) (7) using IRLS algorithm; (b) (3), for a range of signal sparsities and numbers of highly corrupted measurements. See text for details.

4. CONCLUSIONS

In this paper, an approach to compressed sensing with corruptions based on nonconvex optimization was proposed. Its theoretical analysis is based on the analysis of ℓ_p minimization in the case when measurements are not highly corrupted [10]. Sufficient conditions for the success of ℓ_p minimization were expressed in terms of the generalized notion of restricted isometry property, as introduced in [5]. Although the algorithm can only be expected to produce a local minimum of the problem, numerical experiments confirm better performance of ℓ_p minimization compared to ℓ_1 minimization empirically. We also emphasize that the approach discussed here can be straightforwardly extended to the case of noisy measurements, in similar way as in [14], which was not done here for the lack of space.

5. REFERENCES

[1] Y. Eldar and G. Kutyniok, *Compressed sensing: Theory and applications*, Cambridge University Press, 2012.

[2] B. Natarajan, "Sparse approximate solutions to linear systems," *SIAM Journal on Computing*, vol. 24, no. 2, pp. 227–234, 1995.

[3] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[4] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, pp. 253–263, 2008.

[5] X. Li, "Compressed sensing and matrix completion with constant proportion of corruptions," *Constructive Approximation*, vol. 37, no. 1, pp. 73–99, 2013.

[6] J. N. Laska, M. A. Davenport, and R. G. Baraniuk, "Exact signal recovery from sparsely corrupted measurements through the pursuit of justice," in *Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*, nov. 2009, pp. 1556 – 1560.

[7] N. H. Nguyen and T. D. Tran, "Exact recoverability from dense corrupted observations via ℓ_1-minimization," *Information Theory, IEEE Transactions on*, vol. 59, no. 4, pp. 2017–2035, 2013.

[8] J. Wright and Y. Ma, "Dense error correction via ℓ_1-minimization," *Information Theory, IEEE Transactions on*, vol. 56, no. 7, pp. 3540–3560, 2010.

[9] C. Studer, P. Kuppinger, G. Pope, and H. Bolcskei, "Recovery of sparsely corrupted signals," *Information Theory, IEEE Transactions on*, vol. 58, no. 5, pp. 3115–3130, 2012.

[10] R. Chartrand and V. Staneva, "Restricted isometry properties and nonconvex compressive sensing," *Inverse Problems*, vol. 24, no. 3, pp. 035020, 2008.

[11] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2008, pp. 3869–3872.

[12] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.0 beta," http://cvxr.com/cvx, Sept. 2012.

[13] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, V. Blondel, S. Boyd, and H. Kimura, Eds., Lecture Notes in Control and Information Sciences, pp. 95–110. Springer-Verlag Limited, 2008.

[14] R. Saab, R. Chartrand, and O. Yilmaz, "Stable sparse approximations via nonconvex optimization," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2008, pp. 3885–3888.