SEVERAL NOTES ON SPARSE RECOVERY IN THE PRESENCE OF IMPULSIVE NOISE, AND A COMPARISON OF FEW ALGORITHMS

Marko Filipović

Rudjer Bošković Institute
Bijenička 54, 10000 Zagreb, Croatia

ABSTRACT

Several approaches have been proposed in the literature of sparse signal reconstruction to enable robust reconstruction in the presence of impulsive noise. This paper builds upon [1]. The work in [1] presented a method that can provably recover a sparse vector from a small number of measurements, under the assumption that the noise is sparse (or has a sparse representation in an orthonormal basis), and that the measurement matrix satisfies the restricted isometry property. The method is based on finding a sparse solution of an augmented linear system. It was shown in [1] that the augmented matrix satisfies the restricted isometry property with high probability when the measurement matrix is Gaussian. Although it was stated in [1] that similar results can be proved for all sub-Gaussian distributions, here we precisely derive the corresponding result for the symmetric Bernoulli distribution, based on the work in [2]. As another contribution, we complete the theorem on uniqueness of the sparse solution from [3] for the more general case. Also, we present a comparison of several sparse recovery algorithms in the presence of impulsive noise.

Index Terms— Sparse signal reconstruction, impulsive noise, sparse noise

1. INTRODUCTION

This paper discusses sparse recovery in the presence of impulsive noise, i.e. noise that has large or even infinite variance. The only assumption on the noise is that it is *sparse* or has a sparse representation in orthonormal matrix. More precisely, we consider the general model

$$y = Ax + \Omega e = [A\ \Omega]\, u \tag{1}$$

where A is $m \times n$ *measurement matrix*, where $n > m$, vector $x \in \mathbb{R}^n$ is sparse, Ω is $n \times l$ matrix with orthonormal columns (therefore, $l \leq n$), and $e \in \mathbb{R}^l$ is also a sparse vector. Here, we have denoted $u = \left[x^T\ e^T\right]^T$. This problem setting was already considered in [1] (see also references therein). The term Ωe represents possibly large errors that affect "clean" measurements Ax. Possible causes of such noise can be data corruption when sending it over a network, malfunctioning sensors or similar [1]. In information theory, similar problem setting is known as error correction [4]. However, in the context of error correction, the number of measurements ("codeword length") is *larger* than the dimension of the signal being transmitted, and the signal is not necessarily sparse.

The problem of recovering sparse vector x from measurements y 1 can be stated as

$$\min_{x \in \mathbb{R}^n} \{\|x\|_0 + \lambda \|e\|_0\}, \tag{2}$$

where $\|x\|_0$ denotes the ℓ_0 quasi-norm (not really a norm) of vector x, defined as the number of its non-zero elements. Problem 2 is hard because $\|\cdot\|_0$ is a discrete, nonconvex function. Therefore, usually a convex relaxation of 2 is considered:

$$\min_x \{\|x\|_1 + \lambda \|e\|_1\}, \tag{3}$$

where $\|\cdot\|_1$ denotes the ℓ_1 norm of a vector. The above problem 3 has been considered in several papers (we review them in Subsection 1.1). By denoting $u = [x\ \lambda e]^T$, it can be written differently, as

$$\min_{u \in \mathbb{R}^{n+m}} \|u\|_1 \quad \text{subject to} \quad \left[A\ \frac{1}{\lambda}\Omega\right] u = y \tag{4}$$

([1] and references therein). This is a convex optimization problem known as basis pursuit, commonly solved in compressed sensing. Practically more relevant formulation, which allows some error or noise, is

$$\min_{u \in \mathbb{R}^{n+m}} \|u\|_1 \quad \text{subject to} \quad \left\| \left[A\ \frac{1}{\lambda}\Omega\right] u - y \right\|_2 \leq \epsilon \tag{5}$$

where ϵ denotes an estimation of the noise level (this formulation is referred to as basis pursuit denoising). A slight difference of 5 compared to the approach in [1] is that we consider the trade-off parameter λ which can generally be different from 1. The problem 2 cannot be written in this way since $\|\cdot\|_0$ is not homogeneous (except in the case $\lambda = 1$); however, we can replace the ℓ_0 function by its continuous approximation. Often used approximations are ℓ_p pseudo-norms, $\|\cdot\|_p$,

This work was supported through grant 098-0982903-2558 funded by the Ministry of Science, Education and Sports, Republic of Croatia..

for $0 < p \leq 1$, defined as $\|x\|_p^p = \sum_{i=1}^n |x|^p$, for $x \in \mathbb{R}^n$. Therefore, a modified problem can be considered

$$\min_{u \in \mathbb{R}^{n+m}} \|u\|_p^p \quad \text{subject to} \quad \left[A\, \lambda^{-\frac{1}{p}} \Omega\right] u = y. \qquad (6)$$

Here, p-th power of ℓ_p norm is used to enable extraction of λ. Many algorithms exist in the literature on sparse recovery for solving the above problem 6 (for example, see [5] and references therein).

Before precisely stating the contributions of this work, in the following subsection we review the literature on robust sparse recovery.

1.1. Previous work

Several papers have considered sparse recovery in the presence of impulsive, i.e. *sparse* noise ([1, 3, 6, 7] is a partial list). In [7], model 3 was assumed, with application in image restoration. Since 3 can be written as a linear program, an efficient interior point algorithm was proposed. Namely, preconditioned conjugate gradient method with special type of preconditioners was used for solving a linear system in every iteration of the interior point method. The authors demonstrated good performance of the method in image restoration problems. We compare the method proposed here with the method based on the model 3 in the Experiments section. The approach from [7] was extended in [3] to sparse signal recovery problems. They also considered the $\ell_1 +$TV model (here, TV refers to *total variation*, see paper for details), wherein $\|x\|_1$ in 3 is replaced with total variation of x, $\text{TV}(x)$, for applications in image restoration. The authors in [3] also derived sufficient conditions for uniqueness of the problem 2 in terms of the number of outliers and sparsity of the solution vector.

In [6], robust *sampling* and *reconstruction* of signals for compressed sensing was considered. Namely, the authors proposed to use robust estimators of correlation (for linear sampling of a signal) and error (for signal recovery in the presence of outliers). They used the *Lorentzian norm* as the robust metric to measure the error. Lorentzian norm is a robust estimator of location in the presence of Cauchy-distributed noise (see [6] for details). Therefore, they proposed the following problem formulation to recover a sparse signal in the presence of impulsive noise:

$$\min_x \|x\|_1 \quad \text{subject to} \quad \|y - Ax\|_{LL_2,\gamma} \leq \epsilon \qquad (7)$$

where $\|\cdot\|_{LL_2,\gamma}$ denotes the Lorentzian norm with parameter γ. They proved that, if the measurement matrix A satisfies the restricted isometry property (RIP) (see Section 2) with certain constant, and if the Lorentzian norm parameter γ is chosen appropriately, the sparse solution vector can be accurately reconstructed by solving 7. However, the constraint in 7 is *nonconvex*. Therefore, since their method relies on the solution of a nonconvex problem, accurate reconstruction is generally not guaranteed. Also, the method has several parameters, and the authors in [6] gave only general recommendations on how to set them, using probabilistic analysis. The authors demonstrated that this approach performs well in some experiments.

In this paper we build upon the paper [1]. There, $\ell_1 + \ell_1$ minimization was considered (problem 5 with $\lambda = 1$). They showed that, if the elements of A are drawn as independent and identically distributed (i.i.d.) samples from Gaussian distribution, the matrix $[A\, \Omega]$ satisfies the RIP with high probability. They also mentioned that any sub-Gaussian distribution could also be used, with similar bound on probability that the RIP is satisfied. Therefore, any sparse recovery algorithm that has provable performance guarantees under the RIP assumption can be used to recover u (therefore, both x and e). They concentrated on ℓ_1 minimization, and called their algorithm *justice pursuit*. They also considered the noisy setting, wherein, apart from sparse noise (or noise with sparse representation in Ω), the measurements are also degraded by bounded (for example, Gaussian) noise. But they argued that some other reconstruction algorithm could also be used.

1.2. Contributions of this work

Our first contribution, related to the paper [1], is that we give precise constants related to probability that RIP is satisfied, for symmetric Bernoulli distribution. Related to the uniqueness of the sparse solution of 1, we complete the theorem from [3]. There, uniqueness conditions were stated only for $\lambda \geq 1$ and $\Omega = I$ (here, I denotes the $n \times n$ identity matrix). Also, we compare the robust method from [6], $\ell_1 + \ell_1$ minimization approach (that includes the approach from [1]) and the OMP algorithm applied to the problem 2 with $\lambda = 1$ on several experiments. We use OMP because of its simplicity, and since it was analyzed using the RIP [8]. To the best of the author's knowledge, these methods haven't been compared previously.

1.3. Organization of the paper

In Section 2 we discuss conditions for provable recovery of the solution of the problem 2 with $\lambda = 1$. In Subsection 2.1 we show that, if the elements of A are generated from symmetric Bernoulli distribution, the matrix $[A\, \Omega]$ satisfies the restricted isometry property with high probability. In Subsection 2.2 we complete the theorem on the uniqueness of the solution of the problem 2 from [3]. The results of numerical experiments are presented in Section 3.

2. MAIN

We consider the problem formulation 2 with $\lambda = 1$. It can be written differently as

$$\min_u \|u\|_0 \quad \text{subject to} \quad [A\, \Omega]\, u = y, \qquad (8)$$

where $u = [x\, e]^T$ as before, since $\|u\|_0 = \|x\|_0 + \|e\|_0$. Corresponding formulation that allows some noise is

$$\min_u \|u\|_0 \quad \text{subject to} \quad \|[A\ \Omega]\, u - y\|_2 \leq \epsilon. \tag{9}$$

Corresponding ℓ_p norm minimization problem 6 can handle the case $\lambda \neq 1$; however, it is unclear how it affects the performance of the algorithm. Any greedy algorithm for sparse recovery can be used to solve 8 or 9. Sufficient conditions for greedy algorithms to provably recover u are usually stated in terms of the restricted isometry property (RIP) of the measurement matrix [9, 8, 10]. Matrix $A \in \mathbb{R}^{n\times N}$ is said to satisfy the (symmetric) restricted isometry property of order k with constant δ_k if

$$(1-\delta_k)\,\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta_k)\,\|x\|_2^2$$

for all $x \in \mathbb{R}^N$ such that $\|x\|_0 \leq k$. It was shown in [1] (based on [11]), Lemma 1 and Theorem 1, that the matrix $[A\ \Omega]$, with Ω orthonormal, satisfies the restricted isometry property if the elements of A are generated as i.i.d. samples from normal distribution. There, it was stated that similar results can be shown for the more general class of sub-Gaussian matrices. In the following subsection, we derive precise constants for the case when the elements of A are generated i.i.d. from symmetric Bernoulli distribution. In Subsection 2.2 we discuss the uniqueness of the solution of 2.

2.1. RIP of union of Bernoulli and orthonormal matrix

Using analogous reasoning as the one presented in [2], we show that the matrix $[A\ \Omega]$, where A has elements generated as i.i.d. samples from symmetric Bernoulli distribution, also satisfies the RIP with high probability. More precisely, two types of distributions are considered:

$$a_{ij} = \frac{1}{\sqrt{n}} \begin{cases} +1\,, \text{ with probability } \frac{1}{2} \\ -1\,, \text{ with probability } \frac{1}{2} \end{cases} \tag{10}$$

and

$$a_{ij} = \sqrt{\frac{3}{n}} \times \begin{cases} +1\,, \text{ with probability } \frac{1}{6} \\ 0\quad, \text{ with probability } \frac{2}{3} \\ -1\,, \text{ with probability } \frac{1}{6} \end{cases} \tag{11}$$

Here, the $1/\sqrt{n}$ normalization factor is included to ensure that $\mathbb{E}\left(\|[A\ \Omega]\, u\|^2\right) = \|u\|^2$. These two distributions were considered in [2]. We are interested in the bound of the form

$$\mathbb{P}\left(\left|\|[A\ \Omega]\, u\|_2^2 - \|u\|_2^2\right| \geq \epsilon\|u\|_2^2\right) \leq ce^{-nc_0(\epsilon)} \tag{12}$$

where $\epsilon \in (0,\, 1)$ and $c_0(\epsilon)$ depends on ϵ only. Note that the term $\|[A\ \Omega]\, u\|_2^2$ can be expanded as

$$\|[A\ \Omega]\, u\|_2^2 = \|Ax\|_2^2 + \|e\|_2^2 + 2e^T\Omega^T Ax \tag{13}$$

since Ω is orthonormal. The term $2e^T\Omega^T Ax$ can be written as the sum of scaled symmetric Bernoulli variables. Namely, if we denote $z = 2\Omega e$, we have

$$2e^T\Omega^T Ax = \sum_{i,\,j} (x_i z_j)\, a_{ij}. \tag{14}$$

We want to show a bound of the form

$$\mathbb{P}\left(\left|2e^T\Omega^T Ax\right| \geq \epsilon\|x\|_2\|e\|_2\right) \leq e^{-n\tilde{c}(\epsilon)}. \tag{15}$$

For "normal" A (i.e., the elements of A i.i.d. from normal distribution), it was shown in [2] that

$$\mathbb{P}\left(\left|2e^T\Omega^T Ax\right| \geq \epsilon\|x\|_2\|e\|_2\right) \leq e^{-n\epsilon^2/8}. \tag{16}$$

On the other hand, it was shown in [2] that for both symmetric Bernoulli and Gaussian A we have

$$\mathbb{P}\left(\left|\|Ax\|_2^2 - \|x\|_2^2\right| \geq \epsilon\|x\|_2^2\right) \leq 2e^{-n\left(\epsilon^2/2-\epsilon^3/3\right)}. \tag{17}$$

In [1], different bound on the above probability was stated (for Gaussian case), namely $2e^{-n\epsilon^2/8}$. It is valid for $\epsilon \leq 3/4$ since then $\epsilon^2/4 - \epsilon^3/6 \geq \epsilon^2/8$. Bound 17 was derived both in [2] and [12] (for Gaussian case).

Using the ideas from [2], we can use the bound 16 for normal A to derive the bound for the symmetric Bernoulli case. Let use denote $c = \|x\|_2\|e\|_2$. From Markov inequality, we have

$$\mathbb{P}\left(z^T Ax \geq c\epsilon\right) \leq \mathbb{E}\left(e^{hz^T Ax}\right) e^{-hc\epsilon} \tag{18}$$

for all $h > 0$. Let us denote by T a random matrix with i.i.d elements from $\mathcal{N}(0,\, 1/n)$. We have that $\mathbb{E}\left(\exp\left(hz^T Tx\right)\right)$ is finite for small h (since $z^T Tx$ is normally distributed, with expectation 0 and variance $\|x\|_2^2\|y\|_2^2/n$). Therefore, we have

$$\begin{aligned} \mathbb{E}\left(\exp\left(hz^T Tx\right)\right) &= \prod_{i,\,j} \mathbb{E}\left(\exp\left(hz_i x_j t_{ij}\right)\right) \\ &= \prod_{i,\,j}\sum_{l\geq 0} \frac{\left(hz_i x_j\right)^l}{l!}\mathbb{E}\left(t_{ij}\right) \\ &\geq \prod_{i,\,j}\sum_{l\geq 0} \frac{\left(hz_i x_j\right)^l}{l!}\mathbb{E}\left(a_{ij}\right) \qquad (19) \\ &= \mathbb{E}\left(\exp\left(hz^T Ax\right)\right) \qquad (20) \end{aligned}$$

where 19 follows since all moments of symmetric Bernoulli variable are dominated by the moments of Gaussian variable with mean zero and the same variance [2] (this is valid for both distributions 10 and 11). In the above calculations, the expectation and sum can be swaped because of the monotone convergence theorem [2]. We have that $\mathbb{E}\left(\exp\left(hy^T Tx\right)\right) = e^{h^2 v/2}$, where $v = \|x\|_2^2\|y\|_2^2/n$. Therefore, we have

$$\mathbb{P}\left(y^T Ax \geq c\epsilon\right) \leq e^{h^2 v/2 - hc\epsilon}. \tag{21}$$

By optimizing this bound over h, we get

$$\mathbb{P}\left(y^T A x \geq c\epsilon\right) \leq e^{-n\epsilon^2/8}. \quad (22)$$

Similarly, we have

$$\mathbb{P}\left(y^T A x \leq -c\epsilon\right) \leq e^{-n\epsilon^2/8}. \quad (23)$$

This follows because of the symmetry of the distribution of a_{ij}.

Going back to 12, we have that, with probability greater than $1 - 4e^{-n\epsilon^2/8}$ (for $\epsilon \leq 3/4$)

$$\begin{aligned} \|[A\ \Omega]\, u\|_2^2 &= \|Ax\|_2^2 + \|e\|_2^2 + 2e^T\Omega^T Ax && (24)\\ &\leq (1+\epsilon)\|x\|_2^2 + \epsilon\|x\|_2\|e\|_2 + \|e\|_2^2 && (25)\\ &\leq (1+2\epsilon)\|u\|_2^2 && (26)\end{aligned}$$

and

$$\|[A\ \Omega]\, u\|_2^2 \geq (1-2\epsilon)\|u\|_2^2. \quad (27)$$

The above inequalities are analogous to those in [1]. It should be noted that the above inequalities are valid with even higher probability than stated here, by combining bounds for both inequalities in 12. Of course, as noted in [1], similar bounds are valid for all sub-gaussian distributions, i.e. those that are dominated by the centered Gaussian distribution ([13], Chapter 5).

Based on the above concentration inequalities, it follows that the matrix $[A\ \Omega]$, where the elements of A are generated as i.i.d. elements from symmetric Bernoulli distribution with variance $1/n$, satisfies the RIP with high probability (Theorem 1 from [1], which follows from Theorem 5.2 in [11]).

2.2. Uniqueness of the sparse solution

Sufficient conditions for uniqueness of the solution of the problem 2 were derived in [3], under the assumption $\lambda \geq 1$. However, analogous reasoning can be used to derive sufficient conditions for uniqueness for $\lambda < 1$ too. Also, analogous result can be proved for more general setting 1 with Ω with orthonormal columns. Here we state the Proposition 2.2 from [3] with these extensions.

Proposition 1 *Let $A \in \mathbb{C}^{n\times N}$, where $n \leq N$, be such that it has only full-rank submatrices. Let $x_0 \in \mathbb{C}^N$ with $\|x_0\|_0 = m$ and $e_0 \in \mathbb{C}^n$ with $\|e_0\|_0 = K$ be given. Let $b = Ax_0 + \Omega e_0$. Then, for $\lambda \geq 1$ ($0 < \lambda < 1$), if $n \geq 2K + (\lambda+1)m$ ($n \geq 2m + \left(1+\frac{1}{\lambda}\right)K$), problem 2 has the unique solution $\hat{x} = x_0$.*

Proof. The proposition was proved for $\lambda \geq 1$ and $\Omega = I$ (here, I denotes $n \times n$ identity matrix) in [3]. Here we prove the other part ($0 < \lambda < 1$) in the more general setting (we do not suppose $\Omega = I$) analogously. The proof of the case $\lambda \geq 1$ and for general orthonormal Ω is analogous. Note that the problem 2 is equivalent to

$$\min_x\{\frac{1}{\lambda}\|\Omega^T(y - Ax)\|_0 + \|x\|_0\} \quad (28)$$

i.e. x_0 is the solution of 2 if and only if it is the solution of 28. Let $\hat{x} \neq x_0$ be another solution of 28, and $\hat{e} = \Omega^T(y - A\hat{x})$. We have $\|\hat{x}\|_0 + \frac{1}{\lambda}\|\hat{e}\|_0 \leq m + \frac{1}{\lambda}K$. Let us denote by S the support (set of indexes of non-zero elements) of $\hat{x} - x_0$, and by T the support of $\hat{e} - e_0$. If we denote $\|\hat{e}\|_0 = s$, we have $|S| \leq 2m + \frac{1}{\lambda}K - \frac{1}{\lambda}s$ and $|T| \leq s + K$. From $A(\hat{x} - x_0) = \Omega(e_0 - \hat{e})$ we have that $\left(\Omega^T A\right)_{T^C,\Omega}(\hat{x} - x_0)_S = 0$. Matrix $\left(\Omega^T A\right)_{T^C,S}$ is of full rank by assumption, since it can be written as

$$\left(\Omega^T A\right)_{T^C,S} = \left(\Omega_{T^C}\right)^T A_S \quad (29)$$

and Ω_{T^C} is of full rank (since it has orthonormal columns). The condition $\left|T^C\right| \geq |S|$, which would yield $\hat{x} = x_0$, is equivalent to $n \geq 2m + (1+\frac{1}{\lambda})K + (1-\frac{1}{\lambda})s$. Since $\lambda < 1$, the proposition is proved. □

The assumption used in the proposition, that every submatrix of A is of full rank, is satisfied with high probability for random matrices. In the case $\lambda = 1$, the statement of the proposition reduces to the usual requirement for the number of measurements, namely that it is at least two times larger than the ℓ_0 quasi-norm of the solution.

3. EXPERIMENTS

We have compared the proposed method with the methods from [6] and [3] on some synthetic experiments. We have used problem settings from that papers. All reported numerical experiments were done in MATLAB R2011b on a 3 GHz Quad-Core Windows 7 PC with 12GB memory. Code for reproducing the results is available on author's webpage[1]. We have implemented the method from [6] using the fmincon function in MATLAB. The absolute value in the definition of the Lorentzian norm in 7 was approximated as $|x| \approx \sqrt{x^2 + \epsilon}$ to make it differentiable. In all experiments, ϵ was set to 0.001.

3.1. First experiment

In the first experiment, the problem setting was the following (it is taken from [6]. The size of matrix A was set to 128×1024. Elements of A were generated i.i.d. from standard normal distribution. Columns of A were normalized to unit norms. Elements of the solution vector were generated randomly from $-1, 1$. Afterwards, S was normalized so that its average power is 0.78 (as in [6]). Sparsity of the solution vector was varied in the range from 8 (as fixed in [6]) to 20.

OMP algorithm was used for sparse recovery (problem 8). Stopping criterion was set to maximal number of non-zero elements of the solution, and it was set to $2(k + n_{\text{imp}})$, where k denotes the sparsity (ℓ_0 norm) of the solution vector, and n_{imp} denotes the number of impulses. In this experiment, the number of impulses was 13. The amplitude of all impulses was 50, as in [6]. Parameters of the proposed method were as

[1] MATLAB code

follows. γ was set to 0.1, ϵ in 7 was set to $2\|e\|_{LL_2,\gamma}$, where e is the true noise vector. Debiasing was used after reconstruction, as proposed in [6]. For solving the $\ell_1 + \ell_1$ problem 3, we have used the cvx package. λ was set to 1. Experiment was repeated 10 times. Figure 1 shows the results. OMP algorithm clearly performs better than both robust method [6] and $\ell_1 + \ell_1$ approach.

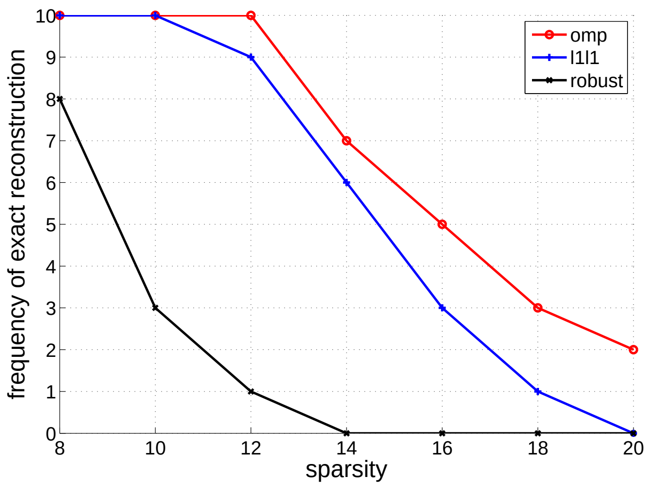


Fig. 1: Frequency of exact reconstruction (over 10 runs) when 10 impulses of amplitude 50 were added to the measurements, for several values of the sparsity of the solution

3.2. Experiment 2

In the second experiment, A was generated as in the first. Sparsity of the solution vector was fixed to 8. This time, p-contaminated noise was added. Namely, the number of impulses was $\lfloor pm + 0.5 \rfloor$ (here, $\lfloor \cdot \rfloor$ denotes the floor function), where contamination factor p was varied in the range from 0.001 to 0.5. Gaussian noise with standard deviation 0.1 was added to the measurements too. This setting was also adopted from [6]. The experiment was repeated 100 times, every time randomly generating the system. Method [6] wasn't used in the simulations because of the slow implementation; however, all paramaters of the problem are the same as in [6], and the results reported here should be compared with Figure 15(a) from that paper. Results of our experiments are shown in Figure 2. It can be seen that both the proposed and the $\ell_1 + \ell_1$ approach perform well in this experiment, better than the method from [6]. $\ell_1 + \ell_1$ approach is sensitive to the choice of the parameter λ, and possibly even better average reconstruction SNR could be obtained with some other choice of λ.

3.3. Experiment 3

In this experiment, problem setting was as in the previous, but the noise was generated from impulsive alpha-stable distribution. The parameter α of the distribution was varied in the range from 0.2 (very impulsive) to 2 (Gaussian). In this case, stopping criterion for OMP algorithm was the norm of the residual. It was set to $\sqrt{m}\,\mathrm{mad}\, x/10$, where $\mathrm{mad}\, x$ denotes the median absolute deviation of vector x (it is a robust estimator of scale/standard deviation of x). This choice doesn't use prior information about noise, and therefore obtained results can be considered realistic. For $\ell_1 + \ell_1$ approach, λ was again set to 1. Again, the method from [6] wasn't used in our simulations because the implementation is slow. The experiment was repeated 100 times. Results are shown in Figure 3, and should be compared with Figure 15(b) in [6]. It can be seen that, in this case, the method [6] performs better, but it is tailored for this problem setting. Namely, in this case the noise vector is not sparse, it can only be viewed as approximately sparse. This is the reason for better performance of robust method that uses the Lorentzian norm. However, even here, OMP gave comparable results.

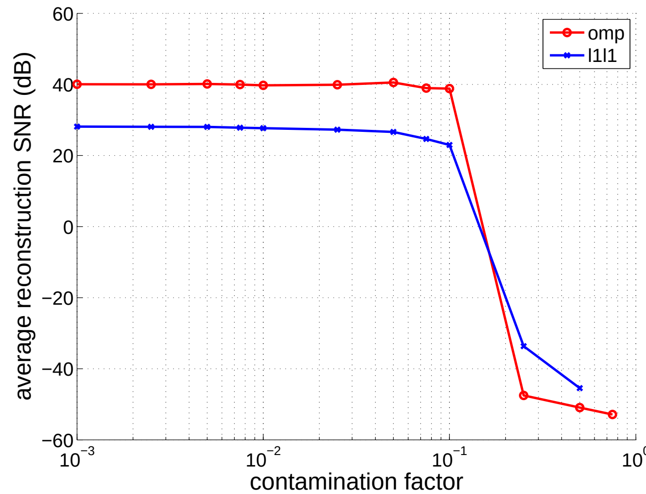


Fig. 2: Average reconstruction signal-to-noise ratio (SNR) (over 100 runs) as a function of the contamination factor (see text for details). These graphs should be compared with the Figure 15(a) in [6].

3.4. Experiment 4

In this experiment, we have used the setting from the first experiment 3.1, but the elements of the solution vector were generated from standard normal distribution. Parameters of all the methods were the same as in the first experiment. Results are shown in Figure 4. Robust method from [6] couldn't recover the true solution in any of the cases.

3.5. Experiment 5

In the next experiment, we have compared the proposed method with the $\ell_1 + \ell_1$ aproach from [3]. The problem setting was as in [3]. Namely, the size of the measurement

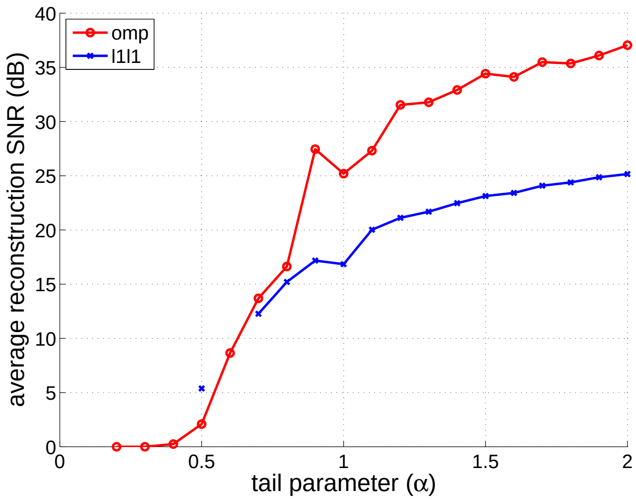


Fig. 3: Average reconstruction signal-to-noise ratio (SNR) (over 100 runs) as a function of the tail parameter of the noise (see text for details). These graphs should be compared with the Figure 15(b) in [6].

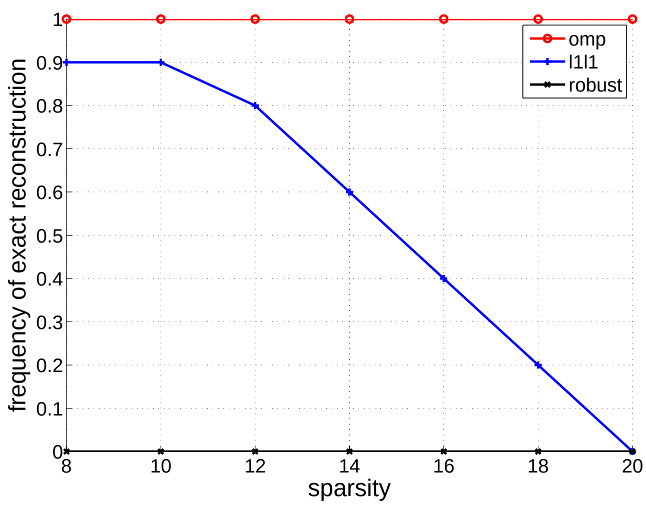
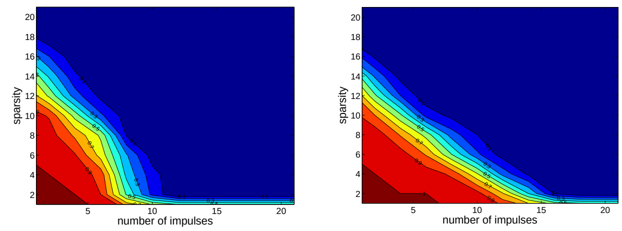


Fig. 4: Frequency of exact reconstruction (over 10 runs) when 10 impulses of amplitude 50 were added to the measurements, for several values of the sparsity of the solution. Nonzero elements of the true solution were generated from standard normal distribution.

matrix A was 40×64. A was generated by taking a subset of rows of two-dimensional discrete cosine transform (DCT) matrix with randomly generated indexes. Indexes of nonzero elements of the solution vector were generated randomly. Nonzero values of the solution were generated uniformly from $[-1, -0.1] \cup [0.1, 1]$. Nonzero values of the noise vector were taken from the set $\{\min Ax, \max Ax\}$. To create phase transition plots for both methods, sparsities of the solution vector and the noise vector were varied in the range from 0 to 20. For every fixed pair of sparsity of the solution and noise, 100 runs were performed, every time randomly generating indexes of rows of DCT matrix, indexes of nonzero elements of the solution and of the noise vector, and the values of nonzero elements of the solution and noise. Recovery was considered successful if the relative error of the reconstruction was below 10^{-4}. For both methods, fraction of successful recoveries in 100 runs was calculated. Results are shown in Figures 5a and Figure 5b. It can be seen that the



(a) (b)

Fig. 5: (a) Phase transition plot for the OMP algorithm. Measurement matrix was generated by randomly selecting 40 rows of the 2-D DCT matrix. See text for details. (b) Phase transition plot for the $\ell_1 + \ell_1$ approach. Problem setting was the same as in (a).

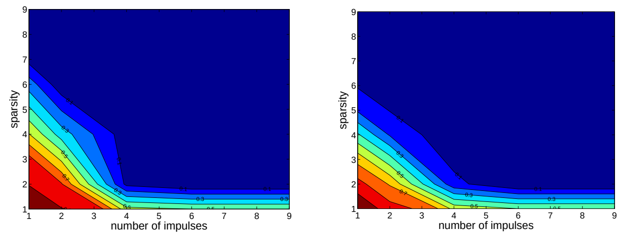
$\ell_1 + \ell_1$ approach performed little better than the OMP, but this example can be considered unrealistic since the number of measurements is relatively large with respect to the dimension of the solution vector. We included this experiment since this setting was used in [3].

3.6. Experiment 6

The problem setting in this experiment was similar to that in the previous 3.5, except that the measurement matrix was generated by randomly selecting 16 rows of 64×64 2-D DCT matrix, and the non-zero elements of the solution vector were generated from standard normal distribution. The results are shown in Figures 6a and 6b. This time, OMP performed better than $\ell_1 + \ell_1$ approach. This is expected since it is known that OMP performs better when the non-zero elements of the solution have some decay. On the other hand, ℓ_1 minimization is insensitive to the distribution of non-zero elements [14].

4. CONCLUSIONS

We have discussed sparse recovery in the presence of impulsive (sparse) noise. It can be argued that the results presented in Subsections 2.1 and 2.2 follow straightforwardly from papers [1, 3]. Still, they are relevant for the problem of sparse recovery with impulsive noise. However, the main focus in this paper has been on numerical experiments that show generally better performance of the proposed method compared to robust sparse recovery method from [6]. The method from



(a) (b)

Fig. 6: (a) Phase transition plot for the OMP algorithm. Measurement matrix was generated by randomly selecting 16 rows of the 2-D DCT matrix. See text for details. (b) Phase transition plot for the $\ell_1 + \ell_1$ approach. Problem setting was the same as in (a).

[1], which is based on convex relaxation of 8 by replacing ℓ_0 quasi-norm with ℓ_1 norm, has also been used in comparative performance analysis, showing results similar to OMP, as expected. We have used the OMP algorithm, however any other method for solving the sparse recovery problem 8 can be used (as already suggested in [1]).

5. REFERENCES

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

[2] Dimitris Achlioptas, "Database-friendly random projections: Johnson-lindenstrauss with binary coins," *Journal of Computer and System Sciences*, vol. 66, no. 4, pp. 671 – 687, 2003, Special Issue on PODS 2001.

[3] B. Popilka, S. Setzer, and G. Steidl, "Signal recovery from incomplete measurements in the presence of outliers," *Inverse problems and imaging*, vol. 1, no. 4, pp. 661–672, 2007.

[4] E.J. Candes and P.A. Randall, "Highly robust error correction by convex programming," *Information Theory, IEEE Transactions on*, vol. 54, no. 7, pp. 2829 –2840, july 2008.

[5] R. Chartrand and Wotao Yin, "Iteratively reweighted algorithms for compressive sensing," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, 31 2008-april 4 2008, pp. 3869 –3872.

[6] R. E. Carrillo, K. E. Barner, and T. C. Aysal, "Robust sampling and reconstruction methods for sparse signals in the presence of impulsive noise," *IEEE J. of Sel. Top. Signal Process.*, vol. 4, no. 2, pp. 392–408, 2010.

[7] H. Fu, M. Ng, M. Nikolova, and J. Barlow, "Efficient minimization methods of mixed l2-l1 and l1-l1 norms for image restoration," *SIAM Journal on Scientific Computing*, vol. 27, no. 6, pp. 1881–1902, 2006.

[8] Jian Wang, Seokbeop Kwon, and Byonghyo Shim, "Near optimal bound of orthogonal matching pursuit using restricted isometric constant," *EURASIP Journal on Advances in Signal Processing*, vol. 2012, pp. 1–7, 2012.

[9] D. Needell and J.A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301 – 321, 2009.

[10] Jian Wang, Seokbeop Kwon, and Byonghyo Shim, "Generalized orthogonal matching pursuit," *Signal Processing, IEEE Transactions on*, vol. 60, no. 12, pp. 6202 –6216, dec. 2012.

[11] Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, pp. 253–263, 2008.

[12] Sanjoy Dasgupta and Anupam Gupta, "An elementary proof of a theorem of johnson and lindenstrauss," *Random Structures & Algorithms*, vol. 22, no. 1, pp. 60–65, 2003.

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

[14] A. Maleki and D.L. Donoho, "Optimally tuned iterative reconstruction algorithms for compressed sensing," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, no. 2, pp. 330 –341, april 2010.