Robust Sparse Signal Recovery for Compressed Sensing with Sampling and Dictionary Uncertainties

Yipeng Liu*, Maarten De Vos, and Sabine Van Huffel

Abstract

Compressed sensing (CS) shows that a signal having a sparse or compressible representation can be recovered from a small set of linear measurements. In classical CS theory, the sampling matrix and dictionary are assumed to be known exactly in advance. However, uncertainties exist due to sampling distortion, finite grids of the parameter space of dictionary, etc. In this paper, we take a generalized sparse signal model, which simultaneously considers the sampling and dictionary uncertainties. Based on the new signal model, a new optimization model for robust sparse signal recovery is proposed. This optimization model can be deduced with stochastic robust approximation analysis. Both convex relaxation and greedy algorithms are used to solve the optimization problem. For the convex relaxation method, a sufficient condition for recovery by convex relaxation method is given; For the greedy algorithm, it is realized by the introduction of a pre-processing of the sensing matrix and the measurements. In numerical experiments, both simulated data and real-life ECG data based results show that the proposed method has a better performance than the current methods.

Index Terms

compressed sensing, robust sparse signal recovery, sampling uncertainty, dictionary uncertainty.

Yipeng Liu and Sabine Van Huffel are with ESAT-STADIUS and iMinds Future Health Department, Dept. of Electrical Engineering, KU Leuven, Kasteelpark Arenberg 10, box 2446, 3001 Leuven, Belgium. (email:yipeng.liu@esat.kuleuven.be; sabine.vanhuffel@esat.kuleuven.be)

Maarten De Vos is with Methods in Neurocognitive Psychology Lab, Department of Psychology, Cluster of excellence Hearing4all, European Medical School, Carl von Ossietzky University, Oldenburg, Germany; and he is also with Research Center Neurosensory Science, Carl von Ossietzky University, Oldenburg, Germany. (email:maarten.de.vos@uni-oldenburg.de)

Manuscript received Month Day, 2014; revised Month Day, Year.
I. INTRODUCTION

Compressed sensing (CS) has received a great deal of attention over the recent years [1]. It shows that if a signal is sparse or compressible with respect to some basis, it can be accurately reconstructed with large probability from random measurements by some nonlinear sparse signal recovery methods. Rather than first sampling at a high Nyquist rate and then compressing the sampled data, it directly senses the data in a compressed form with a lower sub-Nyquist sampling rate. CS has wide applications in diverse fields, e.g., imaging [2] [3], biomedical signal analysis [4], wireless communications [5], etc.

Classical CS theory assumes the representation matrix (dictionary) and sampling (measurement) matrix are known exactly in advance [1] [6] [7]. However, some uncertainty or possible inaccuracy can affect them in many applications. For example, in the sparse representation of the signal, the assumed basis typically corresponds to a gridding of the parameter space, e.g., a discrete Fourier transformation (DFT) grid [8]. But in reality no physical field is exactly sparse in the DFT basis. No matter how finely the parameter space is gridded, the signal may not lie perfectly on the sampling points. This leads to mismatch between the assumed and the actual bases, which results in the uncertainty in the representation matrix. The sampling of the analogue signal’s circuit noise and other non-linear effects can induce uncertainty in the sampling matrices [9]. The classical sparse signal model did not consider these kinds of uncertainties; and the corresponding sparse signal recovery methods can suffer performance degeneration because of signal model mismatch.

Some papers have addressed related problems recently. In [10], the authors analyzed the basis pursuit (BP) recovery of signals with general perturbations in the measurement matrix. In [11] and [12], they evaluated the recovery performance of the BP when uncertainty exists in the representation matrix. Instead of a fixed basis, [13] made use of a tree-structured dictionary of bases and the best bases were estimated with an iteratively processed recovery of the signal. [9] investigated the uncertainty in a sampling matrix. A robust sparse spectrum estimation method was proposed by relaxation of the distortionless constraint. However, the paper only addressed the entry-wise sampling error with too much relaxation, which was
deduced by a series of large inequality zoom operations. [14] proposed a way for the structured sensing matrix perturbation. In [15] [16], two non-convex methods were proposed to deal with uncertainty in data in the sparse linear regression problem. [15] developed an expectation-maximization (EM) algorithm to cope with missing data in the linear regression problem, which is non-convex and no global convergence is guaranteed. To improve the EM algorithm, [16] proposed a projected gradient descent algorithm. It can not find the global optimal solution but a near-optimal one. The non-convexity requires knowledge of the L1 norm of the unknown sparse signal in order to maintain bounded iterates, which is not available in many applications. Besides, they focus on the linear regression problem, and neither of them focused their discussion on the CS background. Additionally, there is no separate model or analysis of the measurement matrix or representation matrix uncertainty. [17] introduced a sparsely regularized total least-squares (SRTLS) method to deal with the uncertainty in the sparse representation matrix. But unfortunately the paper did not explicitly model and discuss the sampling matrix uncertainty. The total uncertainty resulting from both sampling matrix uncertainty and representation matrix is not mentioned. In addition, its solver needs a number of iterations between the sparse signal estimation and the uncertainty matrix estimation, which implies a large computational burden. In summary, to our knowledge, previous publications have not fully analyzed the resulting total uncertainty from both sampling and representation uncertainties. Furthermore, no algorithm of low computational complexity exists for sparse signal recovery in the presence of either sampling uncertainty or representation uncertainty.

In this paper, we explicitly generalize the sparse signal model containing both measurement and representation errors. A brief discussion about the measurement and representation errors is given. Based on the generalized sparse signal model and possible statistical prior knowledge about the measurement and representation errors, a new data fitting constraint is deduced with stochastic uncertainty. We combine it with the \( \ell_0 \) norm minimization based sparsity-inducing constraint, and obtain an optimization model for robust sparse signal recovery. Two approaches are used to solve the optimization problem. One relaxes the \( \ell_0 \) norm to the \( \ell_1 \) norm to obtain a convex programming problem; and the other takes a greedy algorithm
way. For convex programming, we give a sufficient condition for successful recovery; and for the greedy algorithm, we prove it can be solved by regular greedy algorithms with transformations on sensing matrix and measurements. Numerical results show the performance of the proposed method with both simulated data and real-life ECG signals.

The rest of the paper is organized as follows. Section II gives the generalized sparse signal model. In section III the corresponding optimization model for robust sparse signal recovery is deduced. In Section IV both convex relaxation and greedy algorithm are used to solve the optimization model. Section V demonstrates the performance of the proposed method by numerical experiments. Finally, section VI presents the conclusions of this work.

II. Generalized Sparse Signal Model

In CS, instead of acquiring the signal $x \in \mathbb{R}^{N \times 1}$ directly according to the Nyquist sampling, a measurement matrix $\Phi \in \mathbb{R}^{M \times N}$ is used to sample the signal with $M \ll N$, which can be formulated as:

$$y = \Phi x$$  \hspace{1cm} (1)

where the obtained vector $y \in \mathbb{R}^{M \times 1}$ contains the sub-Nyquist-sampled random measurements.

To enable CS, the measurement matrix $\Phi$ should satisfy one of the sufficient conditions, such as restricted isometry property (RIP) [18], coherence condition [19], null space property (NSP) [20], constrained minimal singular values (CMSV) condition [21]. Usually three kinds of measurement matrices are used, namely a Gaussian matrix, Bernoulli matrix or partial Fourier matrix [1].

In order to recover the signal from sub-Nyquist measurements, sparsity should be exploited in CS. Sparsity widely exists in many natural and man-made signals. It means that many of the representative coefficients are close to or equal to zero, when the signal is represented in a dictionary $\Psi \in \mathbb{R}^{N \times N}$. It can be formulated as:

$$x = \Psi \theta$$  \hspace{1cm} (2)
where $\theta \in \mathbb{R}^{N \times 1}$ is the representative vector, and most of its entries are zero or nearly zero. The number of nonzero or significant entries are $K$.

Combining (1) and (2), we can get:

$$y = \Phi \Psi \theta = A \theta$$

where

$$A = \Phi \Psi$$

where $A$ is called sensing matrix. Based on the standard sparse signal model (3), classical CS proves that the signal can be successfully recovered by a series of sparse signal recovery methods [1].

To further consider the errors in the data, an additive noiseterm is included into the signal model as:

$$y = A \theta + n$$

where $n \in \mathbb{R}^{M \times 1}$ is the additive white Gaussian noise (AWGN) with zero mean and variance $\sigma^2$ [22].

However, in many practical scenarios, uncertainty in the sampling matrix exists. When sampling the analogue signals, uncertainty can result from various types of non-ideal effects, such as aliasing, aperture effect, jitter and deviation from the precise sample timing intervals, noise, and other non-linear effects. After sampling, uncertainty can also be introduced by an inconsistent channel effect, channels’ coupling effect, and so on. Here we can model the sampling matrix with uncertainty as:

$$\Phi = \Phi + E_1$$

where $\Phi$ is the uncertainty-free sampling matrix which is known in advance or can be estimated by training data, and $E_1$ is the sampling matrix error. The exact information about $E_1$ cannot be available. We can approximately treat it as a random Gaussian variable matrix or some deterministic unknown variable matrix [8].

There is uncertainty in the representation matrix (dictionary) too. It can result from the quantification of the representation matrix, such as the gridding of the parameter space of dictionary, the mismatch
between the assumed dictionary for sparsity and the actual dictionary in which the signal is sparse, and so on. Similarly we model the representation matrix with uncertainty as:

$$\Psi = \tilde{\Psi} + E_2$$  

(7)

where $\tilde{\Psi}$ is the uncertainty-free representation matrix which is known in advance or can be estimated by training data, and $E_2$ is the representation matrix error. We can approximately treat it as a random Gaussian variable matrix, or a random variable matrix in uniform distribution or some deterministic unknown variable matrix [23].

To take the errors in both sampling and representation into consideration, we can reformulate (4) as:

$$A = (\Phi + E_1)(\Psi + E_2)$$

$$= \Phi\tilde{\Psi} + \Phi E_2 + E_1 \Psi + E_1 E_2$$

$$= \Phi\tilde{\Psi} + (\Phi E_2 + E_1 \Psi + E_1 E_2)$$

$$= \tilde{\Phi} + E$$

(8)

where

$$\tilde{\Phi} = \Phi\tilde{\Psi}$$  

(9)

$$E = \Phi E_2 + E_1 \Psi + E_1 E_2$$  

(10)

$E$ is the sensing matrix error.

Based on the discussed model above, we can set up the sparse signal model with sampling and representation uncertainties and the additive noise. The generalized sparse signal model can be formulated as:

$$y = A\theta + n$$

$$A = \bar{A} + E$$

(11)
III. Optimization Model for Robust Sparse Signal Recovery

A. Classical methods

Given the measurement vector $y$ and the matrix $A$, we need to recover the sparse representative vector $\theta$. In CS, to find the sparsest signal that yields the measurements, we can solve the optimization problem:

$$
\min_{\theta} \|\theta\|_0 \quad \text{s. t. } y = A\theta
$$

(12)

where $\|\theta\|_0$ is the $\ell_0$ norm which counts the number of the nonzero entries of the vector $\theta$. It should be noted that the $\ell_0$ norm is not a full-fledged norm. Solving (12) is NP-hard.

To solve the $\ell_0$ programming problem (12), four groups of sparse signal recovery methods can be used. The first one is $\ell_q$ ($0 < q < 1$) optimization [24] [25], which can be solved in several ways, such as iteratively re-weighted least squares (IRLS) [26], and reweighted L1 optimization [27], firm thresholding algorithms [28]. The second one contains convex optimization algorithms, such as basis pursuit denoising (BPDN) and Dantzig selector (DS); the third one constitutes of greedy algorithms, such as matching pursuit (MP), orthogonal matching pursuit (OMP), and orthogonal multiple matching pursuit (OMMP); the fourth one includes hybrid methods, such as CoSaMP, and subspace pursuit (SP). Generally $\ell_q$ ($0 < q < 1$) optimization may result in convergence to local minima. Moreover, the relative performance of these non-convex algorithms may vary from one set of randomly selected measurements to another [29]. In the other three groups, convex optimization achieves the best reconstruction accuracy while greedy algorithms are computationally most efficient; hybrid methods keep balance between the reconstruction accuracy and the computational complexity [30].

BPDN is a very popular convex method for standard sparse signal recovery. It uses the $\ell_1$ norm to replace the $\ell_0$ norm used in (12) and the data fitting constraint is relaxed to deal with the additive noise. The obtained convex programming model can be formulated in the form of Morozov regularization as:
\[
\min_{\theta} \|\theta\|_1
\]
\[
\text{s. t. } \|y - A\theta\|_2^2 \leq \varepsilon
\]
\[\text{(13)}\]

where \(\varepsilon\) is an appropriately chosen parameter bounding the noise power, \(\|\theta\|_1 = \sum_{n=1}^{N} |\theta_n|\) is the \(\ell_1\) norm of the vector \(\theta\). The \(\ell_1\) norm minimization constraint encourages sparse distribution of entries of \(\theta\), while the constraint \(\|y - A\theta\|_2^2 \leq \varepsilon\) tries to find the best \(\theta\) that fits the linear sampling model with an AWGN.

When BPDN is used in classical CS, the matrix \(A\) is usually assumed to be known exactly in advance. But if uncertainty exists in the sampling and representation as in (8), we can not know \(A\) but only know \(\overline{A}\). The BPDN for a generalized sparse signal in the form of the Tikhonov regularization should be formulated as:

\[
\min_{\theta} \left( \nu \|\theta\|_1 + \|y - \overline{A}\theta\|_2^2 \right)
\]
\[\text{(14)}\]

where \(\nu\) is the parameter which tries to balance the sparsity constraint and data fitting error minimization constraint. A number of efficient solvers can solve the BPDN problem [30].

BPDN is used to recover a sparse signal with additive noise. However, it cannot allow the multiplicative error as in (8) caused by sampling and representation uncertainties. In fact, the data fitting constraint \(\|y - \overline{A}\theta\|_2^2 \leq \varepsilon\) of BPDN matches the sparse signal model with additive noise but does not match the generalized sparse signal model with sampling and representation uncertainties (11). The performance degradation of BPDN has been investigated in [10] [11] [12].

To explain why classical \(\ell_0\) pseudo norm and \(\ell_1\) norm based optimization methods (12) and (13) could lead to incorrect solution with large error, we give an example to illustrate the situation in the presence of sampling and dictionary uncertainty, as shown in Fig. 1. The designed data fitting constraint \(y = A\theta\) is \(\theta_2 = 0.9\theta_1 + 5\) (i.e. \(5 = \begin{bmatrix} -0.9 & 1 \end{bmatrix} \theta_1 \theta_2\) where \(\theta = [\theta_1, \theta_2]^T\). Because of multiplicative noise, the real data constraint in practice is \(\theta_2 = 1.2\theta_1 + 5\) (i.e. \(5 = \begin{bmatrix} -1.2 & 1 \end{bmatrix} \theta_1 \theta_2\)). In Fig. 1a and Fig. 1b, we can see that the tangent points of the minimized \(\ell_0\) and \(\ell_1\) balls with the observed line are on the coordinate
axes, which means the corresponding solutions are sparse. But they are far away from the ones of the
minimized $\ell_0$ and $\ell_1$ balls with the original line which are the true solutions, which means that the error
of the solutions are very large and they are not robust.

To deal with the uncertainty in the dictionary, [17] casts the problem into an SRTLS framework. It can
be formulated as:

$$\min_{\theta, n, E} \|\{E, n\}\|_F^2 + \lambda \|\theta\|_1$$

(15)
s. t. $y = (\bar{A} + E)\theta + n$

for $\lambda > 0$. To solve it, an alternating descent algorithm is used to iterate between

$$\min_{\theta, n} \left(\|n\|_2^2 + \lambda \|\theta\|_1\right)$$

(16)
s. t. $y = [A + E(k - 1)]\theta + n$

to obtain the local estimate $\theta(k)$, and

$$E(k) = [y - A\theta(k)]\theta^T(k)[I + \theta(k)\theta^T(k)]^{-1}$$

(17)
to get the local estimate $E(k)$ in the $k$-th step. The computational complexity of the used alternating
descent algorithm is largely due to the iterations, and the global optimum can not be guaranteed.

B. Robust $\ell_0$ optimization

To robustly recover this generalized sparse signal in a convex way, a new data fitting constraint, other
than the one $\|y - \bar{A}\theta\|_2^2 \leq \varepsilon$ of BPDN, should be deduced, as motivated below. Here we propose a robust
convex programming formulation as:

$$\min_{\theta} \left(\lambda_1\|\theta\|_0 + \lambda_2\theta^T P\theta + \|\bar{A}\theta - y\|_2^2\right)$$

(18)

where $P$ is the error covariance matrix $P = E\{EE^T\}$, which is positive-definite; $\lambda_1$ and $\lambda_2$ are nonnegative
parameters balancing the constraints, which can be tuned using cross validation, regularization path
(a) The contour map of minimized L0 balls which are tangent to the accurate line and the line which has error on slope (multiplicative noise): correct solution (red point of intersection): $[0, 5]^T$; incorrect solution (blue point of intersection): $[-4.1667, 0]^T$.

(b) The contour map of minimized L1 balls which are tangent to the accurate line and the line which has error on slope (multiplicative noise): correct solution (red point of intersection): $[0, 5]^T$, incorrect solution (blue point of intersection): $[-4.1667, 0]^T$.

Fig. 1: The contour map of the minimized balls which are tangent to the accurate line and the line which has error on slope (multiplicative noise).
following etc. The proposed optimization (18) is called robust ℓ_0 (RL0) optimization because it has robustness against the measurement and dictionary uncertainties. One of its equivalent forms is:

$$\min_\theta \|A\theta - y\|_2$$

s.t. $\|\theta\|_0 \leq \omega_1, \ \theta^T P \theta \leq \omega_2^2$

(19)

where $\omega_1$ and $\omega_2$ are parameters too. The robust data fitting constraint with sampling and dictionary uncertainties can be derived as follows.

We can assume that the covariance matrix $P$ is a priori known in the RL0 optimization. If $P$ is not known, we can estimate it by training data. In practice if the data is measured by hardware, as the real measurement matrix and representation matrix are unknown, we may use repeated trials to get the expectation of the uncertainty $E_l$, $l=1, 2, \ldots, L$, where $E_l$ contains the estimated values of uncertainty in (10) in the $l$-th trial. The covariance matrix $P$ can be estimated as $1/L \sum_{l=1}^{L} E_l E_l^T$. That is to say, given $\bar{A}$ and $\theta$, we can get a theoretical measurement vector $y_1 = \bar{A}\theta + n_1$. But the actually measured data is $y_2 = (\bar{A} + E)\theta + n_2$. We can get $z = y_2 - y_1 = E\theta + (n_2 - n_1) = E\theta + n_0$. We use $T$ different values of $\theta$ ($\theta_t$, $t = 1, 2, \ldots, T$) to get the different values of $z$ ($z_t$, $t = 1, 2, \ldots, T$). Denoting $\Xi = \begin{bmatrix} \theta_1 & \theta_2 & \cdots & \theta_T \end{bmatrix}$ and $Z = \begin{bmatrix} z_1 & z_2 & \cdots & z_T \end{bmatrix}$, we have $Z = E\Xi + V$. Letting $T \geq N$ and rank($\Xi$) = $N$, this is a least squares (LS) problem and we can get a good estimation of $E$. One simple choice of $\Xi$ can be the identity matrix. Repeating this kind of trials for $L$ times, we can get a group of $E_l$, $l=1, 2, \ldots, L$, and estimate the covariance matrix $P$.

To deduce of the newly formed data fitting constraint, we assume the uncertainty term $E$ in (8) is a random variable matrix. We refer to the stochastic robust approximation [31]. $y_0 \in \mathbb{R}^{M \times 1}$ denotes the assumed measurement vector obtained by a certain signal model where the model and its parameters are known in advance; and $y \in \mathbb{R}^{M \times 1}$ denotes the practical measurement vector obtained without any knowledge of the signal model. We use the signal model to fit the practical measurements $y$. The expected value of the data fitting error can be formulated as:
Incorporating the generalized sparse signal model (8), we can get

\[
E[\|y - y_0\|_2^2] = E \left[ (\overline{A} + E) \theta + n - y \right]^T \left[ (\overline{A} + E) \theta + n - y \right] = E \left[ (\overline{A} \theta - y)^T + (E \theta + n)^T \right] \left[ (\overline{A} \theta - y) + (E \theta + n) \right] \\
= E \left[ \|\overline{A} \theta - y\|_2^2 + (\overline{A} \theta - y)^T (E \theta + n) \right] \\
+ (E \theta + n)^T (\overline{A} \theta - y) + \|E \theta + n\|_2^2
\] (21)

Here we assume that \(n\) is a random vector with zero mean and variance \(\sigma^2\), \(E\) is a random matrix with zero mean, \(P\) is its covariance matrix, and \(n\) is independent from \(E\). Thus we can get:

\[
E \|y - y_0\|_2^2 = E \left[ \|\overline{A} \theta - y\|_2^2 + \|E \theta + n\|_2^2 \right] = \|\overline{A} \theta - y\|_2^2 + \theta^T P \theta + \sigma^2
\] (22)

Bounding this data fitting error expectation with a parameter \(\eta\) would give a new constraint which matches the generalized sparse signal model (8) as:

\[
\|\overline{A} \theta - y\|_2^2 + \theta^T P \theta \leq \eta
\] (23)

Combining (23) with the \(\ell_0\) norm minimization yields the optimization model for recovering a generalized sparse signal with sampling and representation uncertainties:

\[
\min_{\theta} \|\theta\|_0 \\
\|\overline{A} \theta - y\|_2^2 + \theta^T P \theta \leq \eta
\] (24)

The newly proposed optimization model finds the sparsest solution in all the possible solutions satisfying (23). It can be formulated in a more generalized form as (18) which uses regularization parameters to
balance different constraints. According to the derivation, we can say that it can not only deal with the uniform, Gaussian random uncertainties, but also with any other kind of random uncertainties, provided that the sampled covariance matrix is available.

(a) The contour of minimized L2 balls which are tangent (b) The contour of minimized mixed balls (L1 ball + ellipsoid) to the accurate line and the line which has error on slope (multiplicative noise): correct solution (red point of intersection): \([-2.4862, 2.7624]^T\), incorrect solution (blue point of intersection): \([-2.5594, 4.3891]^T\). Fig. 2: The contour of the minimized balls which are tangent to the accurate line and the line which has error on slope (multiplicative noise)

If we further assume the entries in the multiplicative uncertainty matrix \(E\) are uncorrelated random variables with the same variance \(\delta\), (18) can be simplified as

\[
\min_{\theta} \left( \lambda_1 \|\theta\|_0 + \|\mathbf{A}\theta - \mathbf{y}\|_2^2 + \lambda_2 \delta \|\theta\|_2^2 \right)
\]

(25)

Its performance for compressive sensing was evaluated in [32] recently.

IV. Solutions

Similarly to the classical sparse signal recovery methods, several kinds of methods can solve the optimization model (18). In this section, convex relaxation and a greedy algorithm are used to solve
A. Convex relaxation

A natural way relaxes the $\ell_0$ norm into the $\ell_1$ norm in (18), which achieves a convex optimization model:

$$\min_{\theta} \left( A_1 \|\theta\|_1 + A_2 \theta^T P \theta + \|A\theta - y\|_2^2 \right)$$  \hspace{1cm} (26)

This newly formed one is called convex robust $\ell_1$ (CR-L1) optimization. Another equivalent formulation, which is also the convex relaxation of (19), is:

$$\min_{\theta} \|A\theta - y\|_2$$ \hspace{1cm} s.t. $\|\theta\|_1 \leq \omega_1, \theta^T P \theta \leq \omega_2^2$  \hspace{1cm} (27)

Several approaches exist to solve the CR-L1 optimization, such as interior-point methods, subgradient methods, splitting Bregman algorithm, etc. The convergence can be guaranteed because of its convexity.

To explain why the proposed CR-L1 optimization (27) is robust to multiplicative noise, we use the same example as Fig. 1. Thus the covariance matrix is $P = \begin{bmatrix} 0.09 & 0 \\ 0 & 0 \end{bmatrix}$. Fig. 2 shows the situation when the $\ell_2$ ball and the mixed ball are tangent to the linear constraints. We assume there is no additive noise and $\lambda_1 = \lambda_2 = 1$. One example of the simplified CR-L1 optimization is:

$$\min_{\theta} \left( \|\theta\|_1 + \theta^T P \theta \right)$$ \hspace{1cm} s. t. $y = A\theta$  \hspace{1cm} (28)

In Fig. 2a, we can see that the tangent point of a minimized $\ell_2$ ball with the observed line is not far away from the one of a minimized $\ell_1$ ball with the original line. We can see that $\ell_2$ ball based constraint is more robust to multiplicative noise. But they are not near the coordinate axes, which means that the corresponding solutions are not sparse. We can see clearly in Fig. 2a that the sparsity of the solutions are poor. To combine the ellipsoid constraint’s robustness to multiplicative noise and $\ell_1$ ball constraint’s sparsity, we use the mixed ball in Fig. 2b. The tangent point of the minimized mixed ball ($\ell_1$ ball + ellipsoid) with the observed line is quite near the one of the minimized $\ell_1$ ball with the original line,
and they are near the coordinate axes too. The additional quadratic term induces a slight sparsity loss but brings robustness to multiplicative noise. Therefore, we can see that the proposed mixed ball can achieve a robust sparse solution.

For analysis’ convenience, we assume there is no additive noise. Therefore, one equivalent form of (19) is:

\[
\min_{\theta} \|\theta\|_0 + \lambda_2 \sqrt{\theta^T P \theta} \\
\text{s. t. } y = A \theta
\]  

(29)

Similarly, one equivalent form of (27), which is also the convex relaxation of (29), is:

\[
\min_{\theta} \|\theta\|_1 + \lambda_2 \sqrt{\theta^T P \theta} \\
\text{s. t. } y = A \theta
\]  

(30)

**Theorem 1 (sufficient condition):** Assuming there is no additive noise, the CR-L1 optimization (30) can solve RL0 optimization (29) provided that

\[
M \geq \frac{(2 \sqrt{K} + \lambda_2 C_2)^2}{C_1^2} \log N
\]  

(31)

where \(C_1\) is a constant independent of the dimensions; \(C_2 = E\|E\|_2\) is the expectation of the compatible matrix norm of the \(\ell_2\) vector norm.

**Proof:** Assuming the optimal solutions of the simplified RL0 optimization and CR-L1 optimization are:

\[
\alpha \in \arg \min_{\theta} \|\theta\|_0 + \lambda_2 \sqrt{\theta^T P \theta}, \text{ s. t. } y = A \theta
\]  

(32)

and

\[
\beta \in \arg \min_{\theta} \|\theta\|_1 + \lambda_2 \sqrt{\theta^T P \theta}, \text{ s. t. } y = A \theta
\]  

(33)

the solutions of (32) can solve (33), if

\[
\|\alpha + v\|_1 + \lambda_2 \sqrt{(\alpha + v)^T P (\alpha + v)} \geq \|\alpha\|_1 + \lambda_2 \sqrt{\alpha^T P \alpha}, \forall v \in \ker(A)
\]  

(34)

recalling

\[
\ker(A) = \{\theta \in \mathbb{R}^N : A \theta = 0\}
\]  

(35)
is the kernel (null space) of $A$. (34) means that in all the possible solutions of $y = A\theta$, $\alpha$ also achieves the smallest value of $\|\theta\|_1 + \lambda_2 \theta^T P \theta$.

Let $S$ be the support set $S = \{ n : \alpha_n \neq 0, n = 1, 2, \cdots, N \}$ and $\overline{S} = \{ 1, \cdots, N \} \setminus S$ where $\alpha = [\alpha_1, \alpha_2, \cdots, \alpha_N]^T$, i.e. $S$ is the support of the nonzero entries of $\alpha$; and $\overline{S}$ is the support of the zero entries of $\alpha$. Then,

$$
\|\alpha + v\|_1 = \|\alpha_S + \alpha_{\overline{S}} + v_S + v_{\overline{S}}\|_1 \\
= \|\alpha_S + v_S\|_1 + \|v_{\overline{S}}\|_1 \\
\geq \|\alpha_S\|_1 - \|v_S\|_1 + \|v_{\overline{S}}\|_1 \\
= \|\alpha_S\|_1 + \|v_{\overline{S}}\|_1 + \|v_S\|_1 - 2\|v_S\|_1 \\
= \|\alpha\|_1 + \|v\|_1 - 2\|v_S\|_1 \\
\geq \|\alpha\|_1 + \|v\|_1 - 2\sqrt{K}\|v\|_2
$$

(36)

where $v_S$ keeps its entries corresponding to the support $S$ and let the others be zero; and $v_{\overline{S}}$ keeps its entries corresponding to the support $\overline{S}$ and let the others be zero.

Furthermore, we have

$$
\sqrt{(\alpha + v)^T P (\alpha + v)} = \sqrt{(\alpha + v)^T E (EE^T) (\alpha + v)} \\
= E\|E^T (\alpha + v)\|_2
$$

(37)

and

$$
E\|E^T (\alpha + v)\|_2 = E\|E^T \alpha + E^T v\|_2 \\
\geq E\|E^T \alpha\|_2 - E\|E^T v\|_2 \\
\geq \sqrt{\alpha^T P \alpha} - E\|E\|_2\|v\|_2 \\
= \sqrt{\alpha^T P \alpha} - C_2\|v\|_2
$$

(38)

Combining (36) and (38) results in:

$$
\|\alpha + v\|_1 + \lambda_2 \sqrt{(\alpha + v)^T P (\alpha + v)} \\
\geq \|\alpha\|_1 + \|v\|_1 - 2\sqrt{K}\|v\|_2 + \lambda_2 \sqrt{\alpha^T P \alpha} - \lambda_2 C_2\|v\|_2 \\
= \|\alpha\|_1 + \lambda_2 \sqrt{\alpha^T P \alpha} + \|v\|_1 - \left(2\sqrt{K} + \lambda_2 C_2\right)\|v\|_2
$$

(39)
From (39), we can see that (34) holds provided that \( \|v\|_1 \geq (2\sqrt{K} + \lambda_2 C_2)\|v\|_2 \). In general we have
\[
1 \leq \frac{\|v\|_1}{\|v\|_2} \leq \sqrt{N}.
\]
However, if the elements of \( A \in \mathbb{R}^{M \times N} \) are sampled i.i.d. from Gaussian process with zero mean and unit variance, with high probability, we have
\[
\frac{\|v\|_1}{\|v\|_2} \geq \frac{C_1 \sqrt{M}}{\sqrt{\log \frac{N}{M}}}, \quad \text{for all } v \in \ker(A)
\]
where \( C_1 \) is a constant. When \( A \) is Gaussian, with high probability, we have (34) holds if
\[
M \log M \geq \frac{(2\sqrt{K} + \lambda C_2)^2}{C_1^2} \log N
\]
Generally, \( M \) is larger than the constant of the natural logarithm, and we have \( M \log M \geq M \). Therefore, Theorem 1 is proved.

The similar sufficient condition for convex relaxation of (25) can be obtained if we let \( C_2 = 1 \) in (29). Furthermore, if \( C_2 = 0 \) which means no noise in the model, the sufficient condition for standard BP for CS can be obtained, and the resulted condition agrees with previous conclusions too [1] [6] [7].

B. Greedy algorithm

To reduce the computational complexity, greedy algorithms can be used to solve the RL0 optimization model. In contrast to the classical OMP which greedily chooses the atoms giving the minimum \( \|y - A\theta\|_2^2 \), we update by choosing the ones to minimize
\[
f(\theta) = \|y - A\theta\|_2^2 + \theta^T P \theta
\]

\[
= (y - A\theta)^T (y - A\theta) + \theta^T P \theta
\]

\[
= y^T y - 2y^T A \theta + \theta^T (A^T A + P) \theta
\]

To find the minimum with different values of \( \theta \), we can let
\[
\frac{\partial f(\theta)}{\partial \theta} = -2y^T A + \theta^T \left[ (A^T A + P) + (A^T A + P)^T \right]
\]

\[
= -2y^T A + 2\theta^T (A^T A + P)
\]

\[
= 0
\]
which results in a new equation:

\[ B\theta = z \]  

(44)

where

\[ B = A^T A + P = (\Phi \Psi)^T \Phi \Psi + P \]  

(45)

\[ z = A^T y \]  

(46)

Therefore, in greedy algorithms, we can find one or several atoms which gives the minimum residual for each iteration. i.e. We use the new ”sensing matrix” \( B \) and ”measurements” \( z \) instead of \( A \) and \( y \). With these transformations of sensing matrix and measurement, we can use all the greedy algorithms for CS as before [35]. Therefore the convergence conditions are the same. Because the new sensing matrix \( B \) is random, the condition of successful recovery is the same too [36].

A generalized OMP, which is also called OMMP, is used to realize the robust greedy algorithm [37][38]. It is in the sense that multiple indices are identified in each iteration. When the number of identified indices is \( \rho = 1 \), OMMP is equivalent to OMP. The proposed robust orthogonal multiple matching pursuit (ROMMP) algorithm is summarized in Algorithm 1. The algorithm can be stopped when the residual is larger than a threshold \( \epsilon \) which is proportional to \( \sigma \sqrt{N + 2 \sqrt{N \log N}} \) for Gaussian noise [39].

It should be noted that other solutions for the RL0 optimization (18) are available as well. For example, alternating directions methods of multipliers (ADMM) can be used when the application is a large-scale data processing problem [40].

V. Numerical Experiments

Numerical experiments are given to demonstrate the performance improvement of the proposed method for generalized sparse signals. In Section V-A using simulated data, the performance of the proposed CR-L1 optimization (27) is evaluated in comparison with the BPDN method and SRTLS method. In Section V-B, the performance of the proposed ROMMP is evaluated with the measured ECG signals. To make
Algorithm 1: robust orthogonal multiple matching pursuit

- **Input:** \( \Psi, B \) in (45), \( z \) in (46) and \( \rho \)
- **Output:** \( \hat{x} \)

1. Initial Iteration: \( t := 0 \)
2. Initial Support: \( \hat{\Omega}_t = \emptyset \)
3. Initial Residual: \( r_t = z \)

repeat
   1. Set \( t := t + 1 \)
   2. Update Support: \( \hat{\Omega}_t = \hat{\Omega}_{t-1} \cup \{ \text{argmax}_{i \in \hat{\Omega}_{t-1}} |(B_i^T r_t)| \} \)
   3. Update Coefficients: \( \hat{\theta}_t = \text{argmin}_\theta ||z - (B)_{\hat{\Omega}_t} \theta||_2 \)
   4. Calculate Residual: \( r_t = z - B\hat{\theta}_t \)

until \( ||r_t||_2 \leq \epsilon \);

- \( \hat{x} = \Psi\hat{\theta}_t \)

the numerical experiments on varieties of conditions, we use different kinds of settings in each group of experiments.

A. Simulated data

In the numerical experiments with simulated data, the length of the sparse signal \( \theta \) is \( N = 200 \). It contains only a few nonzero entries. Here the number of nonzero entries is denoted as \( K \). The locations of the nonzero entries vary randomly. It is normalized by its \( \ell_2 \) norm. The signal is sparse with respect to the canonical basis of the Euclidean space, i.e. \( \Psi = I_{N \times N} \); and the sampling matrix \( \Phi \) is Gaussian distributed. As shown in the signal model (10) and (11), the matrix \( A \) can be assumed to be generated by

\[
A = \tilde{A} + \tau U \tag{47}
\]

where \( \tilde{A} \in \mathbb{R}^{M \times N} \) is generated by sampling a white Gaussian distribution with zero mean, and every column of it is normalized by its \( \ell_2 \) norm; \( \tau \in \mathbb{R} \) is the uncertainty parameter; \( U \) is a random matrix, whose columns are normalized with the \( \ell_2 \) norm. \( U \) can be generated by different distributions to simulate different kinds of uncertainties. here \( U \) is generated to be uniformly distributed. The standard deviation
of the AWGN $n$ is $\sigma = 0.1$.

All the parameters in all three methods are chosen to give the best performance based on advanced searching. Here $\nu$, $\lambda$, $\lambda_1$ and $\lambda_2$ are chosen to achieve the best accuracy performance. The number of iterations for SRTLS is 20, which can guarantee the convergence of the algorithms. The matrix $P$ is chosen as the sampled covariance matrix $P = \sum_{l=1}^{L} U_l^T U_l / L$, where $L$ is the number of Monte Carlo simulations, which is chosen to be $L = 500$.

To quantify the performance of signal recovery, the estimation error is calculated via the mean L-b error:

$$e_b = \frac{1}{L} \sum_{l=1}^{L} ||x_l - \hat{x}_l||_b$$

and the mean coherence:

$$c = \frac{1}{L} \sum_{l=1}^{L} \frac{|x_l^H \hat{x}_l|}{||x_l||_2 ||\hat{x}_l||_2}$$

where $x_l$ and $\hat{x}_l$ are the real and estimated signals in the $l$-th experiment; $b \in \{1, 2\}$ indicates different criteria for the evaluation of the estimation performance; when $b = 1$, we call $e_1$ the mean L1 error; and when $b = 2$, $e_2$ is called the mean L2 error.

Fig. 3 - Fig. 5 demonstrate the signal reconstruction performance with the mean L1 error, L2 error and coherence. Fig. 3 gives the mean L1 error, L2 error and coherence with the number of measurements ranging from $M = 10$ to 200, when the uncertainty parameter $\tau = 0.3$ and the number of nonzero entries $K = 10$; Fig. 4 gives the mean L1 error, L2 error and coherence with the number of nonzero entries of the sparse signal ranging from $K = 2$ to 20, when the uncertainty parameter $\tau = 0.3$ and the number of measurements $M = 100$; Fig. 5 gives the mean L1 error, L2 error and coherence with the uncertainty parameter ranging from $\tau = 0.1$ to $\tau = 1$, when the number of nonzero entries of the sparse signal $K = 10$ and the number of measurements $M = 100$.

From Fig. 3 we can see that the CR-L1 optimization outperforms or at least share the same performance with the BPDN and SRTLS with all the possible number of measurements, especially when the number of measurements is large. From Fig. 4 we can see that the CR-L1 optimization outperforms BPDN and
Fig. 3: Experiments of Group A: The mean L1 error, L2 error and coherence versus the number of measurements when the uncertainty is uniformly distributed, the uncertainty parameter $\tau = 0.3$ and the number of nonzero entries $K = 10$. 
Fig. 4: Experiments of Group A: The mean L1 error, L2 error and coherence versus the number of nonzero elements of the sparse signal when the uncertainty is uniformly distributed, the uncertainty parameter
Fig. 5: Experiments of Group A: The mean L1 error, L2 error and coherence versus the uncertainty parameter when the uncertainty is uniformly distributed, the number of nonzero entries of the sparse signal $K=10$ and the number of measurements $M=100$.
SRTLS method for all possible number of nonzero elements are not large. We can see that the CR-L1 optimization outperforms or has nearly the same performance in terms of mean L2 error and mean coherence as the BPDN or SRTLS when the number of nonzero elements are smaller than 14. When the number is larger than 14, the mean coherence performance of the CR-L1 optimization decreases dramatically. This means that the CR-L1 optimization can not have better mean coherence performance than BPDN or SRTLS when the signal is not really sparse. However, as the sparsity is the basic requirement for CS, the number of nonzero elements should not be too large. Fig. 5 also shows that the CR-L1 optimization performs the best, especially when the uncertainties are strong.

In summary, while the mean L2 error of the CR-L1 optimization is similar to the better one of BP and SRTLS, we show that in terms of both mean L1 error and mean coherence, the CR-L1 optimization outperforms the other methods. Therefore, we can make the conclusion that the CR-L1 optimization’s overall performance is better in general.

B. ECG data

To test the proposed method for real-life data, we use ECG data which is obtained from the Physiobank database [41]. Mobile ECG monitoring is one of the most popular applications in compressed sensing of ECG signals [42]. In this application, the computational complexity should be as low as possible. Therefore, in this group of numerical experiments, the greedy algorithms, i.e. OMP, OMMP with $\rho = 4$, and the proposed ROMMP with $\rho = 4$, are compared. The utilized representation matrix is given by the orthogonal Daubechies wavelets (db 10) which is reported to be the most popular wavelet family for ECG compression [43]. The ECG data has 15 channels with 37888 samples for each channel. In each channel, the data are divided into 37 segments, i.e. the length of the signal in each reconstruction is 1024. The generation of multiplicative sampling uncertainty is similar to (47) except that $U$ is Gaussian distributed with zero mean and unit variance. In addition, additive noise is generated with zero mean and variance $\tau^2$.

In Fig. 6, the mean L1 error, mean L2 error and mean coherence for the reconstruction of ECG
signal from the noisy compressive measurements with different number of measurements is shown when the uncertainty parameter is $\tau = 0.3$; and Fig. 7 shows the mean L1 error, mean L2 error and mean coherence for the reconstruction of the ECG signal from the noisy compressive measurements with different uncertainty parameters when the number of measurements are 512. It can be seen that OMP and OMMP have almost the same reconstruction accuracy. However, ROMMP improves much more over various number of measurements and uncertainty degrees compared to OMP and OMMP.

To compare the computational complexity, Fig. 8 shows the mean number of iterations at various percentages of additive and multiplicative noise and various values of CR. We can see that the regular OMP needs the largest number of iterations. The proposed ROMMP needs less iterations than OMMP. Considering that the computational complexity of each iteration of OMP, OMMP, and ROMMP is almost the same, we can conclude that the ROMMP has the least computational complexity.

VI. Conclusion

In this paper, we discuss the sampling and representation uncertainties in CS. A generalized sparse signal model which has both multiplicative noise and additive noise is given. Based on this model, we propose a new optimization model for the robust recovery of this kind of sparse signal with uncertainties. The proposed method can be deduced by a stochastic analysis. To solve the optimization model, both convex relaxation and a greedy algorithm are used. The sufficient condition for successful recovery is analyzed too. Numerical experiments show that the proposed RLO optimization based algorithms are in general superior to the previous ones.

In the future, the proposed method can be combined with the best basis search method to further enhance its performance. The same idea can be extended to handle structured sparse signals with noise. In practice, the dictionary uncertainty may be coherent with the dictionary. In this case the exploitation of the coherence may help to further improve the performance of robust sparse signal recovery.
(a) mean L2 error vs number of measurements

(b) mean L1 error vs number of measurements

(c) mean coherence vs number of measurements

Fig. 6: Experiments of Group B: The mean L1 error, L2 error and coherence values with various number of measurements for a fixed uncertainty parameter.
Fig. 7: Experiments of Group B: The mean L1 error, L2 error and coherence with various uncertainty parameter for a fixed number of measurements.
Fig. 8: Experiments of Group B: The mean number of iterations at various uncertainty parameters of additive and multiplicative noise and various number of measurements.

REFERENCES

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

[2] M. Lustig, D. Donoho, and J. M. Pauly, “Sparse mri: The application of compressed sensing for rapid mr imaging,” *Magnetic resonance in medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[3] R. Baraniuk and P. Steeghs, “Compressive radar imaging,” in *Radar Conference, 2007 IEEE*. IEEE, 2007, pp. 128–133.

[4] Y. Liu, I. Gligorijevic, V. Matic, M. De Vos, and S. Van Huffel, “Multi-sparse signal recovery for compressive sensing,” in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. IEEE, 2012, pp. 1053–1056.
[5] Y. Liu and Q. Wan, “Enhanced compressive wideband frequency spectrum sensing for dynamic spectrum access,” EURASIP Journal on Advances in Signal Processing, vol. 2012, no. 1, pp. 1–11, 2012.

[6] E. J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” Information Theory, IEEE Transactions on, vol. 52, no. 2, pp. 489–509, 2006.

[7] D. L. Donoho, “Compressed sensing,” Information Theory, IEEE Transactions on, vol. 52, no. 4, pp. 1289–1306, 2006.

[8] A. V. Oppenheim, R. W. Schafer, J. R. Buck et al., Discrete-time signal processing. Prentice Hall Upper Saddle River, 1999, vol. 5.

[9] Y. Liu and Q. Wan, “Anti-sampling-distortion compressive wideband spectrum sensing for cognitive radio,” International Journal of Mobile Communications, vol. 9, no. 6, pp. 604–618, 2011.

[10] M. A. Herman and T. Strohmer, “General deviants: An analysis of perturbations in compressed sensing,” Selected Topics in Signal Processing, IEEE Journal of, vol. 4, no. 2, pp. 342–349, 2010.

[11] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, “Sensitivity to basis mismatch in compressed sensing,” Signal Processing, IEEE Transactions on, vol. 59, no. 5, pp. 2182–2195, 2011.

[12] D. H. Chae, P. Sadeghi, and R. A. Kennedy, “Effects of basis-mismatch in compressive sampling of continuous sinusoidal signals,” in Future Computer and Communication (ICFCC), 2010 2nd International Conference on, vol. 2. IEEE, 2010, pp. V2–739.

[13] G. Peyre, “Best basis compressed sensing,” Signal Processing, IEEE Transactions on, vol. 58, no. 5, pp. 2613–2622, 2010.

[14] Z. Yang, C. Zhang, and L. Xie, “Robustly stable signal recovery in compressed sensing with structured matrix perturbation,” Signal Processing, IEEE Transactions on, vol. 60, no. 9, pp. 4658–4671, 2012.

[15] N. Städler and P. Bühlmann, “Missing values: sparse inverse covariance estimation and an extension to sparse regression,” Statistics and Computing, vol. 22, no. 1, pp. 219–235, 2012.

[16] P.-L. Loh and M. J. Wainwright, “High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity,” The Annals of Statistics, vol. 40, no. 3, pp. 1637–1664, 2012.

[17] H. Zhu, G. Leus, and G. B. Giannakis, “Sparsity-cognizant total least-squares for perturbed compressive sampling,” Signal Processing, IEEE Transactions on, vol. 59, no. 5, pp. 2002–2016, 2011.

[18] E. J. Candès, “The restricted isometry property and its implications for compressed sensing,” Comptes Rendus Mathematique, vol. 346, no. 9, pp. 589–592, 2008.

[19] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” Information Theory, IEEE Transactions on, vol. 53, no. 12, pp. 4655–4666, 2007.

[20] A. Cohen, W. Dahmen, and R. DeVore, “Compressed sensing and best k-term approximation,” Journal of the American Mathematical Society, vol. 22, no. 1, pp. 211–231, 2009.

[21] G. Tang and A. Nehorai, “Performance analysis of sparse recovery based on constrained minimal singular values,” Signal Processing, IEEE Transactions on, vol. 59, no. 12, pp. 5734–5745, 2011.

[22] K. Kim and G. Shevlyakov, “Why gaussianity?” Signal Processing Magazine, IEEE, vol. 25, no. 2, pp. 102–113, 2008.

[23] J. Li and P. Stoica, Robust adaptive beamforming. Wiley Online Library, 2006.
[24] D. Wipf and S. Nagarajan, “Iterative reweighted l1 and l2 methods for finding sparse solutions,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 317–329, 2010.

[25] A. Majumdar and R. K. Ward, “An algorithm for sparse mri reconstruction by schatten p-norm minimization,” *Magnetic Resonance Imaging*, vol. 29, no. 3, pp. 408–417, 2011.

[26] B. D. Rao and K. Kreutz-Delgado, “An affine scaling methodology for best basis selection,” *Signal Processing, IEEE Transactions on*, vol. 47, no. 1, pp. 187–200, 1999.

[27] E. J. Candes, M. B. Wakin, and S. P. Boyd, “Enhancing sparsity by reweighted l1 minimization,” *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 877–905, 2008.

[28] M. Fornasier and H. Rauhut, “Iterative thresholding algorithms,” *Applied and Computational Harmonic Analysis*, vol. 25, no. 2, pp. 187–208, 2008.

[29] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[30] J. A. Tropp and S. J. Wright, “Computational methods for sparse solution of linear inverse problems,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 948–958, 2010.

[31] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust optimization*. Princeton University Press, 2009.

[32] M.-J. Lai and W. Yin, “Augmented l1 and nuclear-norm models with a globally linearly convergent algorithm,” *SIAM Journal on Imaging Sciences*, vol. 6, no. 2, pp. 1059–1091, 2013.

[33] B. S. Kashin, “Diameters of some finite-dimensional sets and classes of smooth functions,” *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya*, vol. 41, no. 2, pp. 334–351, 1977.

[34] A. Y. Garnaev and E. D. Gluskin, “The widths of a euclidean ball,” in *Dokl. Akad. Nauk SSSR*, vol. 277, no. 5, 1984, pp. 1048–1052.

[35] J. A. Tropp, “Greed is good: Algorithmic results for sparse approximation,” *Information Theory, IEEE Transactions on*, vol. 50, no. 10, pp. 2231–2242, 2004.

[36] J. A. Tropp and S. J. Wright, “Computational methods for sparse solution of linear inverse problems,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 948–958, 2010.

[37] Z. Xu, “The performance of orthogonal multi-matching pursuit under rip,” *arXiv preprint arXiv:1210.5323*, 2012.

[38] B. Shim, J. Wang, and S. Kwon, “Generalized orthogonal matching pursuit,” *Signal Processing, IEEE Transactions on*, vol. 60, no. 12, pp. 6202–6216, 2012.

[39] T. T. Cai and L. Wang, “Orthogonal matching pursuit for sparse signal recovery with noise,” *Information Theory, IEEE Transactions on*, vol. 57, no. 7, pp. 4680–4688, 2011.

[40] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[41] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, “Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals,” *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
[42] H. Mamaghanian, N. Khaled, D. Atienza, and P. Vandergheynst, “Compressed sensing for real-time energy-efficient ecg compression on wireless body sensor nodes,” *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 9, pp. 2456–2466, 2011.

[43] Y. Liu, M. De Vos, I. Gligorijevic, V. Matic, Y. Li, and S. Van Huffel, “Multi-structural signal recovery for biomedical compressive sensing,” *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2794–2805, 2013.