Compressed Sensing with Sampling and Dictionary Uncertainties: An ℓ_0 + Ellipsoid Optimization

Yipeng Liu, *Member, IEEE*, Maarten De Vos, *Member, IEEE*, and Sabine Van Huffel, *Fellow, IEEE*

***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 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.**

I. Introduction

Classical compressed sensing (CS) theory assumes that the representation matrix (dictionary) and sampling (measurement) matrix are known exactly in advance [1] [2] [3]. 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 [4]. 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 [5]. The classical sparse signal model did not consider these uncertainties; and the corresponding sparse signal recovery methods can suffer performance degeneration because of signal model mismatch.

Some papers have addressed related problems recently. [6], [7], [8] and [9] analyzed signal recovery by basis pursuit (BP) and greedy algorithms with perturbations in either measurement matrix or representation matrix. To deal with the performance degeneration, several methods were proposed [10]-[11]. Instead of a fixed basis, [10] used a tree-structured dictionary of bases and the best bases were estimated with an iteratively processed recovery of the signal. [5] and [12] relaxed the distortionless constraint to allow entry-wise sampling error by a series of large inequality zoom operations. Similarly, [13] and [14] generalized the approximate message passing (AMP) algorithm to hold the sampling matrix uncertainty with several parameters to be tuned or predefined. [15] proposed a way only for the structured sensing matrix perturbation. In [16] [17], two non-convex methods were proposed to deal with uncertainty in data in the sparse linear regression problem. 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. [11] introduced a sparsely regularized total least-squares (SRTLS) method to deal with the uncertainty in the representation matrix. But 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, 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 generalize the sparse signal model containing both measurement and representation errors. 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 ℓ_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 ℓ_0 norm to the ℓ_1 norm to obtain a convex programming problem; and the other one takes a greedy algorithm approach. 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

This work was supported by Research Council KUL GOA/10/09 MaNet, CoE PFV/10/002 (OPTEC); FWO project G.0108.11 (Compressed Sensing);iMinds Medical Information Technologies SBO 2014; Belgian Federal Science Policy Office: IUAP P7/19/ (DYSCO, 'Dynamical systems, control and optimization', 2012-2017); EU: The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013), ERC Advanced Grant: BIOTENSORS (n 339804).

Yipeng Liu and Sabine Van Huffel are with ESAT-STADIUS and i-Minds medical IT Department, Dept. of Electrical Engineering, KU Leuven, Kasteelpark Arenberg 10, box 2446, 3001 Leuven, Belgium. (e-mail:yipeng.liu@esat.kuleuven.be; sabine.vanhuffel@esat.kuleuven.be)

Maarten De Vos is with Institute of Biomedical Engineering, Department of Engineering, University of Oxford, Oxford, United Kingdom. (e-mail:maarten.devos@eng.ox.ac.uk)

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

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

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x}, \tag{1}$$

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

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 $\mathbf{\Psi} \in \mathbb{R}^{N\times N}$. It can be formulated as:

$$\mathbf{x} = \mathbf{\Psi}\theta, \tag{2}$$

where $\theta \in \mathbb{R}^{N\times 1}$ is the representative vector with most of its entries are zero. When most of the entries are not strictly zero but trivial, or only a few of the entries are significant, strictly we should call the vector is compressible, but sometimes we say it be sparse too. The number of nonzero or significant entries are K.

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

$$\mathbf{y} = \mathbf{\Phi}\mathbf{\Psi}\theta = \mathbf{A}\theta, \tag{3}$$

where

$$\mathbf{A} = \mathbf{\Phi}\mathbf{\Psi}, \tag{4}$$

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

To further consider the errors in the data, an additive noise term is included into the signal model as [18]:

$$\mathbf{y} = \mathbf{A}\theta + \mathbf{n}, \tag{5}$$

where $\mathbf{n} \in \mathbb{R}^{M\times 1}$ is the additive white Gaussian noise (AWGN) with zero mean and covariance matrix $\sigma^2\mathbf{I}$ [19].

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:

$$\mathbf{\Phi} = \bar{\mathbf{\Phi}} + \mathbf{E}_1, \tag{6}$$

where $\bar{\mathbf{\Phi}}$ is the uncertainty-free sampling matrix which is known in advance or can be estimated by training data, and $\mathbf{E}_1$ is the sampling matrix error. The exact information about $\mathbf{E}_1$ cannot be available. We can approximately treat it as a random Gaussian variable matrix or some deterministic unknown variable matrix [4] [19].

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:

$$\mathbf{\Psi} = \bar{\mathbf{\Psi}} + \mathbf{E}_2, \tag{7}$$

where $\bar{\mathbf{\Psi}}$ is the uncertainty-free representation matrix which is known in advance or can be estimated by training data, and $\mathbf{E}_2$ 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 [20].

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

$$\mathbf{A} = \left(\bar{\mathbf{\Phi}} + \mathbf{E}_1\right)\left(\bar{\mathbf{\Psi}} + \mathbf{E}_2\right) = \overline{\mathbf{A}} + \mathbf{E}, \tag{8}$$

where

$$\overline{\mathbf{A}} = \bar{\mathbf{\Phi}}\bar{\mathbf{\Psi}}, \tag{9}$$

$$\mathbf{E} = \bar{\mathbf{\Phi}}\mathbf{E}_2 + \mathbf{E}_1\bar{\mathbf{\Psi}} + \mathbf{E}_1\mathbf{E}_2. \tag{10}$$

$\mathbf{E}$ is the sensing matrix error. As can be seen in (10), the correlation between measurement error $\mathbf{E}_1$ and representation error $\mathbf{E}_2$ affects the estimation of $\mathbf{E}$ mainly by the term $\mathbf{E}_1\mathbf{E}_2$.

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:

$$\mathbf{y} = \mathbf{A}\theta + \mathbf{n}, \quad \mathbf{A} = \overline{\mathbf{A}} + \mathbf{E}. \tag{11}$$

III. Optimization Model for Robust Sparse Signal Recovery

A. Classical methods

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

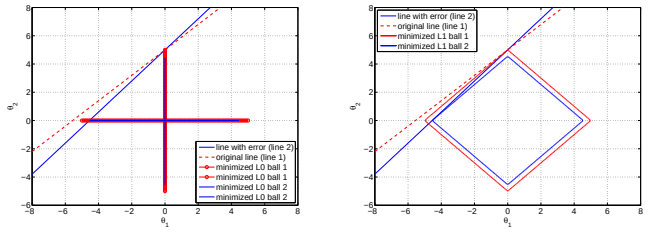
$$\min_{\theta} \|\theta\|_0, \quad \text{s. t.} \quad \mathbf{y} = \mathbf{A}\theta, \tag{12}$$

where $\|\theta\|_0$ is the ℓ_0 norm which counts the number of the nonzero entries of the vector θ. It should be noted that the ℓ_0 norm is not a full-fledged norm. Solving (12) is NP-hard.

The basis pursuit denoising (BPDN) uses the ℓ_1 norm to replace the ℓ_0 norm to make it convex and relaxes the data fitting constraint to deal with the additive noise, where the the ℓ_1 norm of the vector θ is defined as $\|\theta\|_1 = \sum_{n=1}^{N} |\theta_n|$.

BPDN can recover compressible signal with additive noise. However, it cannot allow the multiplicative error as in (8) which is caused by sampling and representation uncertainties. In fact, the relaxed data fitting constraint 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 [6] [8] [9].

To explain why classical ℓ_0 pseudo norm and ℓ_1 norm based optimization methods 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 $\mathbf{y} = \mathbf{A}\theta$ is $\theta_2 = 0.9\theta_1+5$ (i.e. $5 = \begin{bmatrix} -0.9 & 1 \end{bmatrix}\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$) 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. $5 = \begin{bmatrix} -1.2 & 1 \end{bmatrix}\begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$). In Fig. 1a and Fig. 1b, we can see that the tangent points of the minimized ℓ_0 and ℓ_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 ℓ_0 and ℓ_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.



(a) (b)

Fig. 1: (a). The contour map of minimized ℓ_0 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 ℓ_1 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. Robust ℓ_0 optimization

To robustly recover this generalized sparse signal, a new data fitting constraint, other than the one $\left\|\mathbf{y} - \overline{\mathbf{A}}\theta\right\|_2^2 \leq \varepsilon$ of BPDN, should be deduced. We assume the uncertainty term $\mathbf{E}$ in (8) is a random variable matrix, and $\mathbf{P}$ is the covariance matrix $\mathbf{P} = \mathrm{E}\left(\mathbf{E}^T\mathbf{E}\right)$, which is positive-semidefinite and symmetric. We refer to the stochastic robust approximation [21]. $\mathbf{y}_0 \in \mathbb{R}^{M\times 1}$ denotes the assumed measurement vector obtained by the signal model (11), i. e. $\mathbf{y}_0 = \left(\overline{\mathbf{A}} + \mathbf{E}\right)\theta + \mathbf{n}$, and part of its parameters ($\overline{\mathbf{A}}$ and $\mathbf{P}$) are known in advance; and $\mathbf{y} \in \mathbb{R}^{M\times 1}$ denotes the measurement vector obtained in practice without any knowledge of the signal model. We use the signal model to fit the practical measurements $\boldsymbol{y}$. We try to fit the obtained measurement vector with the generalized sparse signal model (11), and the expected value of the data fitting error can be formulated as:

$$\mathrm{E}[\|\mathbf{y} - \mathbf{y}_0\|_2^2]. \tag{13}$$

Incorporating the generalized sparse signal model (8), we can get

$$\begin{aligned} &\mathrm{E}\left[\|\mathbf{y} - \mathbf{y}_0\|_2^2\right] = \mathrm{E}\left[(\mathbf{y} - \mathbf{y}_0)^T(\mathbf{y} - \mathbf{y}_0)\right] \\ &= \mathrm{E}\left[\left[\left(\overline{\mathbf{A}} + \mathbf{E}\right)\theta + \mathbf{n} - \mathbf{y}\right]^T\left[\left(\overline{\mathbf{A}} + \mathbf{E}\right)\theta + \mathbf{n} - \mathbf{y}\right]\right] \\ &= \mathrm{E}\left[\left\|\overline{\mathbf{A}}\theta - \mathbf{y}\right\|_2^2 + \left(\overline{\mathbf{A}}\theta - \mathbf{y}\right)^T(\mathbf{E}\theta + \mathbf{n})\right. \\ &\quad \left.+(\mathbf{E}\theta + \mathbf{n})^T\left(\overline{\mathbf{A}}\theta - \mathbf{y}\right) + \|\mathbf{E}\theta + \mathbf{n}\|_2^2\right]. \end{aligned} \tag{14}$$

Here we assume that all the entries in $\mathbf{n}$ are i.i.d Gaussian with $\mathrm{E}\left(\mathbf{n}^T\mathbf{n}\right) = M\sigma^2$, and $\mathbf{n}$ is independent from $\mathbf{E}$. Thus we can get:

$$\begin{aligned} \mathrm{E}\,\|\mathbf{y} - \mathbf{y}_0\|_2^2 &= \mathrm{E}\left[\left\|\overline{\mathbf{A}}\theta - \mathbf{y}\right\|_2^2 + \|\mathbf{E}\theta + \mathbf{n}\|_2^2\right] \\ &= \left\|\overline{\mathbf{A}}\theta - \mathbf{y}\right\|_2^2 + \theta^T\mathbf{P}\theta + M\sigma^2 \end{aligned}. \tag{15}$$

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

$$\left\|\overline{\mathbf{A}}\theta - \mathbf{y}\right\|_2^2 + \theta^T\mathbf{P}\theta \leqslant \eta. \tag{16}$$

Combining (16) with the ℓ_0 norm minimization yields the optimization model for recovering a generalized sparse signal with sampling and representation uncertainties:

$$\min_{\theta} \|\theta\|_0, \quad \text{s. t. } \left\|\overline{\mathbf{A}}\theta - \mathbf{y}\right\|_2^2 + \theta^T\mathbf{P}\theta \leqslant \eta. \tag{17}$$

It can be further generalized to:

$$\min_{\theta}\left(\lambda_1\|\theta\|_0 + \lambda_2\theta^T\mathbf{P}\theta + \left\|\overline{\mathbf{A}}\theta - \mathbf{y}\right\|_2^2\right), \tag{18}$$

where λ_1 and λ_2 are nonnegative parameters balancing the constraints, which can be tuned using cross validation, regularization path 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:

$$\min_{\theta}\left\|\overline{\mathbf{A}}\theta - \mathbf{y}\right\|_2, \quad \text{s. t. } \|\theta\|_0 \leqslant \omega_1,\ \theta^T\mathbf{P}\theta \leqslant \omega_2^2, \tag{19}$$

where ω_1 and ω_2 are parameters too.

We assume that the covariance matrix $\mathbf{P}$ is *a priori* known in the RL0 optimization. On one hand, we can model $\mathbf{P}$ on the basis of an analysis of the CS setup. For example, we can make a possible assumption that the sampling matrix error is Gaussian as done in [6]; or, when the dictionary error is caused by finite gridding of the parameters (dictionary error corresponding to the quantization of the sparse vector) we can assume a uniform distribution of the gridding parameter, etc. On the other hand, we can estimate $\mathbf{P}$ as addressed in the errors-in-variables modeling literature, see e.g. [22][23][24][25]. As far as we know, the best way to estimate P is via replicated observations

[22][23][24]. Assuming that independent repeated measurements are available for each variable observed with error, this type of replications provides enough information about the error covariance matrix to derive a consistent unbiased estimate of $\mathbf{P}$. A simple way to calculate $\mathbf{P}$ exists provided we can assume that the entries of the error matrix $\begin{bmatrix} \mathbf{E} & \mathbf{n} \end{bmatrix}$ are i.i.d with mean zero and unknown variance δ^2. The estimated covariance matrix is then $\mathbf{P}_{est} = \delta^2\mathbf{I}$. A consistent estimate of δ^2 is provided by the squared minimal singular value of $\begin{bmatrix} \bar{\mathbf{A}} & \mathbf{y} \end{bmatrix}$ [24]. To address more general error distributions, we refer to the extended errors-in-variables modeling literature, see e.g. [23] [24]. Using these consistent estimates instead of the true covariance matrix does not change the consistency properties of the parameter estimators for linear errors-in-variables models, of which the generalized sparse signal model (11) is a special case [22] [25].

The newly proposed optimization model finds the sparsest solution among all possible solutions satisfying (16). It can be formulated in a more generalized form as (18) which uses regularization parameters to balance different constraints.

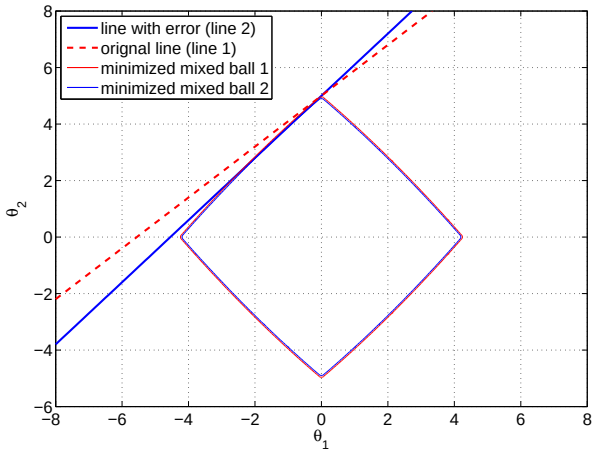


Fig. 2: The contour of minimized mixed balls (L1 ball + ellipsoid) which are tangent to the accurate line and the line which has error on slope (multiplicative noise)): correct solution (red point of intersection): $[0,\ 5]^T$, incorrect solution (blue point of intersection): $[-0.5554,\ 4.3891]^T$.

If we further assume the random elements of $\mathbf{E}$ are uncorrelated, $\mathbf{P}$ can be a diagonal matrix. If we assume the entries in the multiplicative uncertainty matrix $\mathbf{E}$ are uncorrelated random variables with the variances $\delta_1, \delta_2, \cdots, \delta_N$, (18) can be simplified as

$$\min_{\theta} \lambda_1 \|\theta\|_0 + \left\|\bar{\mathbf{A}}\theta - \mathbf{y}\right\|_2^2 + \lambda_2 \left\|\mathbf{\Delta}\theta\right\|_2^2, \tag{20}$$

where

$$\mathbf{\Delta} = diag\begin{pmatrix} \sqrt{\delta_1} & \sqrt{\delta_2} & \cdots & \sqrt{\delta_N} \end{pmatrix}. \tag{21}$$

When we further assume the multiplicative uncertainties are equal with the same variances, i.e. $\delta = \delta_1 = \delta_2 = \cdots = \delta_N$, (20) is further simplified in another form of the elastic net [26]:

$$\min_{\theta} \lambda_1 \|\theta\|_0 + \left\|\overline{\mathbf{A}}\theta - \mathbf{y}\right\|_2^2 + \lambda_2 \delta \left\|\theta\right\|_2^2. \tag{22}$$

Its performance for CS was evaluated in [27] 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 it.

A. Convex relaxation

A natural way relaxes the ℓ_0 norm into the ℓ_1 norm in (18), which achieves a convex optimization model:

$$\min_{\theta} \left(\lambda_1 \|\theta\|_1 + \lambda_2 \theta^T \mathbf{P} \theta + \left\|\overline{\mathbf{A}}\theta - \mathbf{y}\right\|_2^2\right). \tag{23}$$

This newly formed one is called convex robust ℓ_1 (CR-L1) optimization. Another equivalent formulation, which is also the convex relaxation of (19), is:

$$\min_{\theta} \left\|\overline{\mathbf{A}}\theta - \mathbf{y}\right\|_2, \quad \text{s. t. } \ \|\theta\|_1 \leqslant \omega_1, \ \theta^T \mathbf{P} \theta \leqslant \omega_2^2. \tag{24}$$

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 (24) is robust to multiplicative noise, we use the same example as Fig. 1. Thus the covariance matrix is $\mathbf{P} = \begin{bmatrix} 0.09 & 0 \\ 0 & 0 \end{bmatrix}$. One example of the simplified CR-L1 optimization is:

$$\min_{\theta} \left(\|\theta\|_1 + \theta^T \mathbf{P} \theta\right), \quad \text{s. t. } \ \mathbf{y} = \overline{\mathbf{A}}\theta \ . \tag{25}$$

To combine the ellipsoid constraint's robustness to multiplicative noise and ℓ_1 ball constraint's sparsity, we use the mixed ball in Fig. 2. The tangent point of the minimized mixed ball (ℓ_1 ball + ellipsoid) with the observed line is quite near the one of the minimized ℓ_1 ball with the original line, and they are near the coordinate axes too. The additional quadratic term induces a slight compressibility loss but brings robustness to multiplicative noise in terms of better accuracy. Therefore, we can see that the proposed mixed ball can achieve a robust compressible solution.

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

$$\min_{\theta} \|\theta\|_0 + \lambda_2 \sqrt{\theta^T \mathbf{P} \theta} \ , \quad \text{s. t. } \mathbf{y} = \overline{\mathbf{A}}\theta. \tag{26}$$

Similarly, one equivalent form of (24), which is also the convex relaxation of (26), is:

$$\min_{\theta} \|\theta\|_1 + \lambda_2 \sqrt{\theta^T \mathbf{P} \theta} \ , \quad \text{s. t. } \mathbf{y} = \overline{\mathbf{A}}\theta. \tag{27}$$

Theorem 1 (sufficient condition): Assuming there is no additive noise, the CR-L1 optimization (27) can solve RL0 optimization (26) provided that

$$M \geqslant \frac{\left(2\sqrt{K} + \lambda_2 C_2\right)^2}{C_1^2} \log N, \tag{28}$$

where C_1 is a constant independent of the dimensions; $C_2 = \mathrm{E}\|\mathbf{E}\|_2$ is the expectation of the compatible matrix norm of the ℓ_2 vector norm.

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

$$\alpha \in \arg\min_{\theta} \|\theta\|_0 + \lambda_2 \sqrt{\theta^T \mathbf{P} \theta},\ \text{s. t. } \mathbf{y} = \overline{\mathbf{A}}\theta \tag{29}$$

and

$$\beta \in \arg\min_{\theta} \|\theta\|_1 + \lambda_2 \sqrt{\theta^T \mathbf{P} \theta},\ \text{s. t. } \mathbf{y} = \overline{\mathbf{A}}\theta, \tag{30}$$

the solutions of (29) can solve (30), if

$$\|\alpha + \mathbf{v}\|_1 + \lambda_2 \sqrt{(\alpha+\mathbf{v})^T \mathbf{P} (\alpha+\mathbf{v})} \geqslant \|\alpha\|_1 + \lambda_2 \sqrt{\alpha^T \mathbf{P} \alpha}, \forall \mathbf{v} \in \ker(\overline{\mathbf{A}}), \tag{31}$$

recalling

$$\ker\left(\overline{\mathbf{A}}\right) = \left\{\theta \in \mathbb{R}^N : \overline{\mathbf{A}}\theta = 0\right\} \tag{32}$$

is the kernel (null space) of $\overline{\mathbf{A}}$. (31) means that in all the possible solutions of $\mathbf{y} = \overline{\mathbf{A}}\theta$, α also achieves the smallest value of $\|\theta\|_1 + \lambda_2 \theta^T \mathbf{P} \theta$.

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

$$\begin{aligned} \|\alpha + \mathbf{v}\|_1 &= \left\|\alpha_S + \alpha_{\overline{S}} + \mathbf{v}_S + \mathbf{v}_{\overline{S}}\right\|_1 \\ &= \|\alpha_S + \mathbf{v}_S\|_1 + \left\|\mathbf{v}_{\overline{S}}\right\|_1 \\ &\geqslant \|\alpha_S\|_1 - \|\mathbf{v}_S\|_1 + \left\|\mathbf{v}_{\overline{S}}\right\|_1 \\ &= \|\alpha_S\|_1 + \left\|\mathbf{v}_{\overline{S}}\right\|_1 + \|\mathbf{v}_S\|_1 - 2\|\mathbf{v}_S\|_1 \\ &= \|\alpha\|_1 + \|\mathbf{v}\|_1 - 2\|\mathbf{v}_S\|_1 \\ &\geqslant \|\alpha\|_1 + \|\mathbf{v}\|_1 - 2\sqrt{K}\|\mathbf{v}\|_2, \end{aligned} \tag{33}$$

where $\mathbf{v}_S$ keeps its entries corresponding to the support S and let the others be zero; and $\mathbf{v}_{\overline{S}}$ keeps its entries corresponding to the support $\overline{S}$ and let the others be zero.

Furthermore, we have

$$\begin{aligned} \sqrt{(\alpha+\mathbf{v})^T \mathbf{P} (\alpha+\mathbf{v})} &= \sqrt{(\alpha+\mathbf{v})^T \mathrm{E}\left(\mathbf{E}^T\mathbf{E}\right)(\alpha+\mathbf{v})} \\ &= \mathrm{E}\|\mathbf{E}(\alpha+\mathbf{v})\|_2 \end{aligned} \tag{34}$$

and

$$\begin{aligned} \mathrm{E}\|\mathbf{E}(\alpha+\mathbf{v})\|_2 &= \mathrm{E}\|\mathbf{E}\alpha + \mathbf{E}\mathbf{v}\|_2 \\ &\geqslant \mathrm{E}\|\mathbf{E}\alpha\|_2 - \mathrm{E}\|\mathbf{E}\mathbf{v}\|_2 \\ &\geqslant \sqrt{\alpha^T \mathbf{P} \alpha} - \mathrm{E}\|\mathbf{E}\|_2 \|\mathbf{v}\|_2 \\ &= \sqrt{\alpha^T \mathbf{P} \alpha} - C_2 \|\mathbf{v}\|_2. \end{aligned} \tag{35}$$

Combining (33) and (35) results in:

$$\begin{aligned} &\|\alpha+\mathbf{v}\|_1 + \lambda_2 \sqrt{(\alpha+\mathbf{v})^T \mathbf{P} (\alpha+\mathbf{v})} \\ &\geqslant \|\alpha\|_1 + \|\mathbf{v}\|_1 - 2\sqrt{K}\|\mathbf{v}\|_2 + \lambda_2 \sqrt{\alpha^T \mathbf{P} \alpha} - \lambda_2 C_2 \|\mathbf{v}\|_2 \\ &= \|\alpha\|_1 + \lambda_2 \sqrt{\alpha^T \mathbf{P} \alpha} + \|\mathbf{v}\|_1 - \left(2\sqrt{K} + \lambda_2 C_2\right)\|\mathbf{v}\|_2. \end{aligned} \tag{36}$$

From (36), we can see that (31) holds provided that $\|\mathbf{v}\|_1 \geq (2\sqrt{K} + \lambda_2 C_2)\|\mathbf{v}\|_2$. In general we have $1 \leq \|\mathbf{v}\|_1/\|\mathbf{v}\|_2 \leq \sqrt{N}$. However, if the elements of $\overline{\mathbf{A}} \in \mathbb{R}^{M \times N}$ are sampled i.i.d. from Gaussian process with zero mean and unit variance, with high probability, we have

$$\frac{\|\mathbf{v}\|_1}{\|\mathbf{v}\|_2} \leq \frac{C_1 \sqrt{M}}{\sqrt{\log \frac{N}{M}}},\ \text{for all}\ \ \mathbf{v} \in \ker\left(\overline{\mathbf{A}}\right), \tag{37}$$

where C_1 is a constant [28] [29]. When $\overline{\mathbf{A}}$ is Gaussian, with high probability, we have (31) holds if

$$M \geq (\log N - \log M)\left(\frac{2\sqrt{K} + \lambda_2 C_2}{C_1}\right)^2 \tag{38}$$

is satisfied.

Obviously we have $M \geq 1$. (38) can be met if (28) holds. Therefore, Theorem 1 is proved. ■

The similar sufficient condition for convex relaxation of (22) can be obtained if we let $C_2 = 1$ in (28). 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] [2] [3]. In order to suppress the multiplicative noise, the proposed new method has a larger lower bound on the required number of measurements than that of the standard BP with no multiplicative noise. The requirement of additional measurements is the price paid for multiplicative noise suppression.

The introduction of the ℓ_2 norm constraint will not only enhance robustness, but also smooth the ℓ_1 norm based penalty function, as can be seen in Fig. 2 [30]. The convergence of the sub-gradient algorithm would be accelerated.

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 $\left\|\mathbf{y} - \overline{\mathbf{A}}\theta\right\|_2^2$, we update by choosing the ones to minimize

$$\begin{aligned} f(\theta) &= \left\|\mathbf{y} - \overline{\mathbf{A}}\theta\right\|_2^2 + \theta^T \mathbf{P} \theta \\ &= \left(\mathbf{y} - \overline{\mathbf{A}}\theta\right)^T \left(\mathbf{y} - \overline{\mathbf{A}}\theta\right) + \theta^T \mathbf{P} \theta \\ &= \mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T\overline{\mathbf{A}}\theta + \theta^T\left(\overline{\mathbf{A}}^T\overline{\mathbf{A}} + \mathbf{P}\right)\theta. \end{aligned} \tag{39}$$

To find the minimum with different values of θ, we can let

$$\begin{aligned} \frac{\partial f(\theta)}{\partial \theta} &= -2\mathbf{y}^T\overline{\mathbf{A}} + \theta^T\left[\left(\overline{\mathbf{A}}^T\overline{\mathbf{A}} + \mathbf{P}\right) + \left(\overline{\mathbf{A}}^T\overline{\mathbf{A}} + \mathbf{P}\right)^T\right] \\ &= -2\mathbf{y}^T\overline{\mathbf{A}} + 2\theta^T\left(\overline{\mathbf{A}}^T\overline{\mathbf{A}} + \mathbf{P}\right) \\ &= 0, \end{aligned} \tag{40}$$

which results in a new equation:

$$\mathbf{B}\theta = \mathbf{z}, \tag{41}$$

where

$$\mathbf{B} = \overline{\mathbf{A}}^T\overline{\mathbf{A}} + \mathbf{P} = \left(\overline{\mathbf{\Phi}\mathbf{\Psi}}\right)^T\overline{\mathbf{\Phi}\mathbf{\Psi}} + \mathbf{P}, \tag{42}$$

$$\mathbf{z} = \overline{\mathbf{A}}^T\mathbf{y}. \tag{43}$$

Therefore, in greedy algorithms, we can find one or several atoms which give the minimum residual for each iteration. i.e. We use the new "sensing matrix" $\mathbf{B}$ and "measurements" $\mathbf{z}$ instead of $\overline{\mathbf{A}}$ and $\mathbf{y}$. With these transformations of sensing matrix and measurement, we can use all the greedy algorithms for CS as before [31]. We should note that the new sensing matrix $\mathbf{B}$ should not be fully random but partly random. The component $\mathbf{P}$ is deterministic and may prevail when the

multiplicative error is strong and its corresponding covariance matrix has a bad CS performance, such as a large coherence [32], a large restricted isometry constant (RIC) [33]. It can require a stricter condition for successful recovery.

A generalized OMP (orthogonal matching pursuit), which is also called OMMP (orthogonal multi-matching pursuit), is used to realize the robust greedy algorithm [34] [35]. 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 smaller than a threshold ϵ which is proportional to the standard deviation of its additive Gaussian noise [36].

Algorithm 1: robust orthogonal multiple matching pursuit

- **Input:** $\overline{\mathbf{\Psi}}$, $\mathbf{B}$ in (42), $\mathbf{z}$ in (43) and ρ
- **Output:** $\hat{\mathbf{x}}$
- Initial Iteration: $t := 0$
- Initial Support: $\hat{\Omega}_t = \emptyset$
- Initial Residual: $\mathbf{r}_t = \mathbf{z}$

repeat

- Set $t := t + 1$;
- Update Support: $\hat{\Omega}_t = \hat{\Omega}_{t-1} \cup \{ \underset{\mathbf{i} \notin \hat{\Omega}_{t-1}}{argmax} |(\mathbf{B})_{\mathbf{i}}^T \{\mathbf{r_t}\}| \}$
- Update Coefficients: $\hat{\theta}_{\mathbf{t}} = \underset{\vartheta}{argmin} \|\mathbf{z} - (\mathbf{B})_{\Omega_t} \vartheta\|_2$
- Calculate Residual: $\mathbf{r_t} = \mathbf{z} - \overline{\mathbf{B}} \theta_{\mathbf{t}}$

until $\|\mathbf{r}_t\|_2 \leq \epsilon$;

- $\hat{\mathbf{x}} = \overline{\mathbf{\Psi}} \hat{\theta}_{\mathbf{t}}$

V. Numerical Experiments

A. Simulated data

In the numerical experiments with simulated data, the length of the sparse signal θ is $N = 200$. It contains only a few nonzero entries. The number of nonzero entries of the sparse signal $K = 10$. The locations of the nonzero entries vary randomly. It is normalized by its ℓ_2 norm. The signal is sparse with respect to the canonical basis of the Euclidean space, i. e. $\overline{\mathbf{\Psi}} = \mathbf{I}_{N \times N}$; and the sampling matrix $\overline{\mathbf{\Phi}}$ is Gaussian distributed. As shown in the signal model (10) and (11), the matrix $\mathbf{A}$ can be assumed to be generated by

$$\mathbf{A} = \overline{\mathbf{A}} + \tau \mathbf{U}, \tag{44}$$

where $\overline{\mathbf{A}} \in \mathbb{R}^{M \times N}$ is generated by sampling a white Gaussian distribution with zero mean, and every column of it is normalized by its ℓ_2 norm; $\tau \in \mathbb{R}$ is the uncertainty parameter; $\mathbf{U}$ is a random matrix, whose columns are normalized with the ℓ_2 norm. $\mathbf{U}$ can be generated by different distributions to simulate different kinds of uncertainties. here $\mathbf{U}$ is generated to be uniformly distributed. The standard deviation of the AWGN $\mathbf{n}$ is $\sigma = 0.1$.

All the parameters in all three methods are chosen to give the best performance based on advanced searching. Here υ λ, λ_1 and λ_2 are chosen to achieve the best accuracy performance. Getting these optimal values for the parameters is not straightforward. Similarly to parameter estimation in BPDN, cross validation may be used resulting in additional computation burden. The number of iterations for SRTLS is 20. Before the number of iterations reach 20, the error does not vary much. The matrix $\mathbf{P}$ is chosen as the sampled covariance matrix $\mathbf{P} = \sum_{l=1}^{L} \mathbf{U}_l^T \mathbf{U}_l \big/ L$, 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 normalized mean L-b error:

$$e_b = \frac{1}{2L} \sum_{l=1}^{L} \frac{\|\mathbf{x}_l - \hat{\mathbf{x}}_l\|_b}{\|\mathbf{x}_l\|_b} \tag{45}$$

and the mean coherence:

$$c = \frac{1}{L} \sum_{l=1}^{L} \frac{\left|\mathbf{x}_l{}^H \hat{\mathbf{x}}_l\right|}{\|\mathbf{x}_l\|_2 \|\hat{\mathbf{x}}_l\|_2}, \tag{46}$$

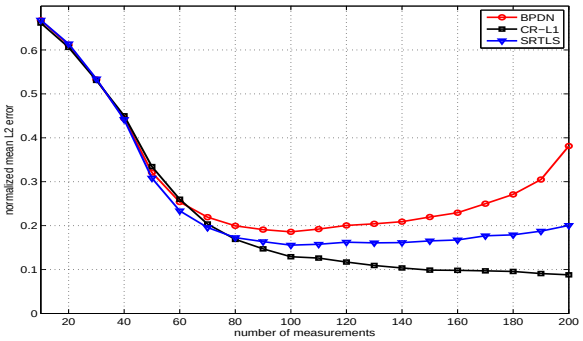
where $\mathbf{x}_l$ and $\hat{\mathbf{x}}_l$ are the real and estimated signals in the l-th experiment, and they are normalized by their ℓ_b norms; $b \in \{1, 2\}$ indicates different criteria for the evaluation of the estimation performance; when $b = 1$, we call e_1 the normalized mean L1 error;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

Fig. 3 - Fig. 4 demonstrate the signal reconstruction performance with the normalized mean L1 and L2 errors, and mean coherence. Fig. 3 gives the normalized mean L1 and L2 errors and mean coherence with the number of measurements ranging from $M = 10$ to 200, when the uncertainty parameter $\tau = 0.3$; Fig. 4 gives the normalized mean L1 and L2 errors and mean coherence with the uncertainty parameter ranging from $\tau = 0.1$ to $\tau = 1$, when the number of measurements $M = 100$.

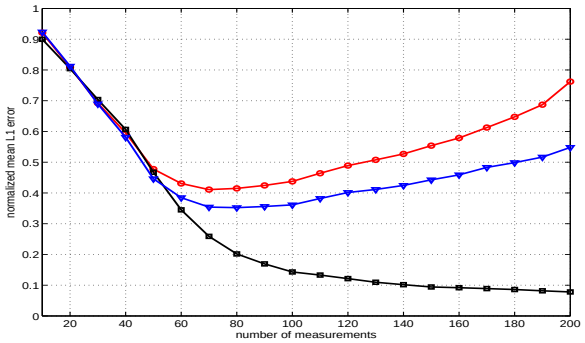
From Fig. 3, we can see that the CR-L1 optimization outperforms or at least share the similar performance with the BPDN and SRTLS with all the possible number of measurements in terms of the normalized mean L1 and L2 errors and mean coherence, especially when the number of measurements is large. Fig. 4 also shows that the CR-L1 optimization performs the best or achieves similar performance, and the performance improvement is especially obvious when the uncertainties are strong.

B. ECG data

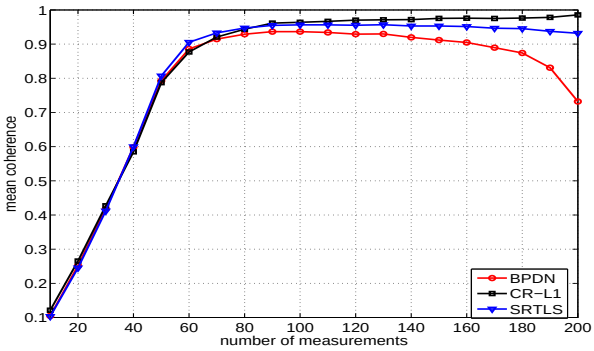
To test the proposed method for real-life data, we use ECG data which is obtained from the Physiobank database [37] [38] [39]. Mobile ECG monitoring is one of the most popular applications in CS of ECG signals. In this application, the computational complexity should be as low as possible. Therefore, 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 measurement matrix is the Gaussian matrix. The utilized representation matrix is given by the orthogonal Daubechies wavelets (db 10) with the decomposition level 5 which is one of the most popular wavelet families for ECG compression [39]. 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 $N = 1024$. The sensing matrix error is generated as $\mathbf{E} = \tau \mathbf{U}$ which is same as in (44) too, where $\mathbf{U}$ is composed by entries in multivariate normal



(a) normalized mean L2 error vs number of measurements

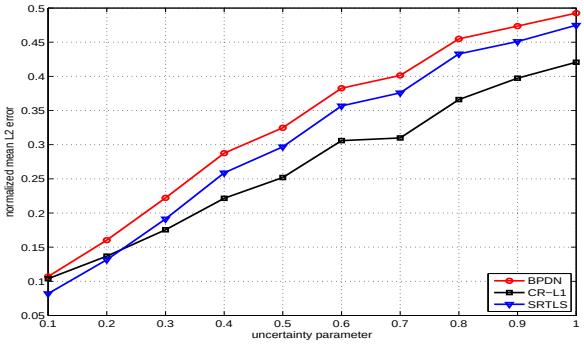


(b) normalized mean L1 error vs number of measurements

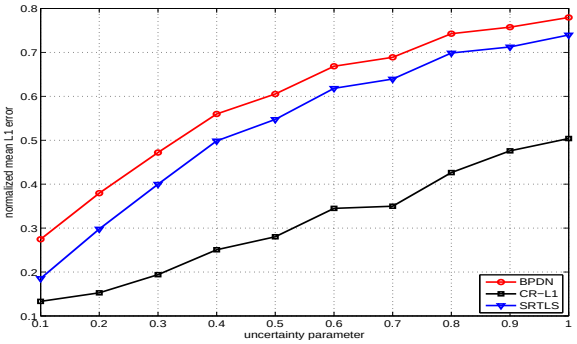


(c) mean coherence vs number of measurements

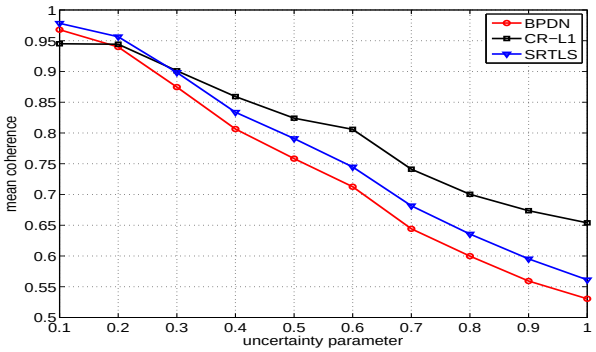
Fig. 3: Experiments of Group A: The normalized mean L1 and L2 errors and mean coherence versus the number of measurements when the uncertainty parameter $\tau = 0.3$.



(a) normalized mean L2 error vs uncertainty parameter



(b) normalized mean L1 error vs uncertainty parameter



(c) mean coherence vs uncertainty parameter

Fig. 4: Experiments of Group A: The normalized mean L1 and L2 errors and mean coherence versus the uncertainty parameter when the number of measurements $M = 100$.

distribution with mean zero, and covariance matrix being a random positive definite matrix. In addition, the standard deviation of the AWGN $\mathbf{n}$ is $\sigma = 0.30$.

Fig. 5 shows part of an ECG signal and its estimates from sub-samples by OMP, OMMP and ROMMP when the number of the compressive measurements is $M = 410$ and $\tau = 0.1$. We can see that the signal reconstructed by ROMMP is less noisy than the ones reconstructed by the OMP and OMMP.

In Fig. 6, the normalized mean L1 and L2 errors 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 normalized mean L1 and L2 errors 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 the accuracy much better 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 number of compressive measurements. We can see that OMP needs the largest number of iterations. The proposed ROMMP needs less iterations

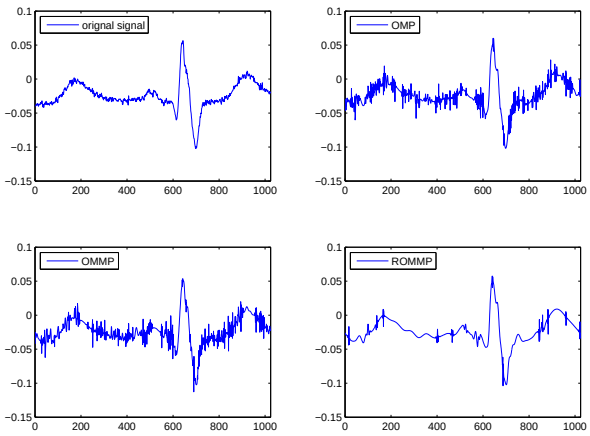
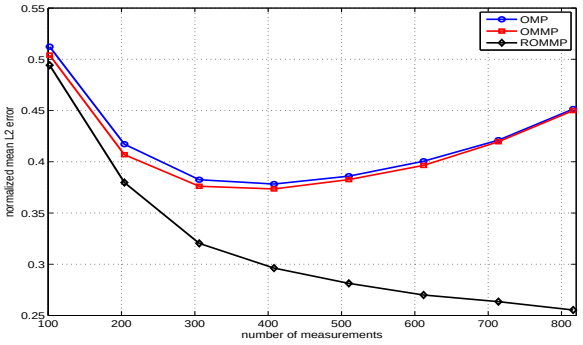


Fig. 5: ECG signal estimates from the sub-samples.

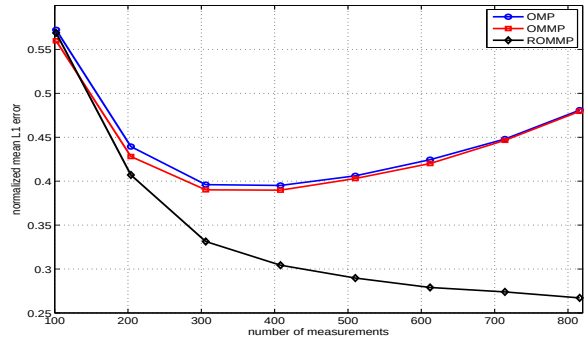
than OMMP. Considering that the computational complexity of each iteration of OMP, OMMP, and ROMMP is almost the same, we can conclude that the ROMMP has the least computational complexity.

VI. Conclusion

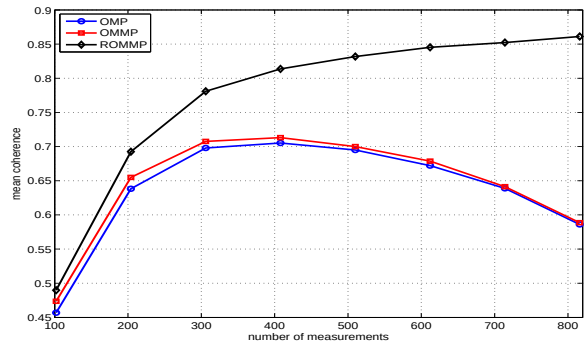
In this paper, we discuss the sampling and representation uncertainties in CS. A generalized sparse signal model considering both multiplicative noise and additive noise is given. Based on this model, a new optimization model for robust recovery of the generalized sparse signal is deduced by a stochastic analysis. Both convex relaxation and a greedy algorithm are used to solve the optimization model. Sufficient conditions for successful recovery are analyzed. Numerical experiments show that the proposed RL0 optimization based algorithms are in general superior to the previous ones.



(a) normalized mean L2 error vs number of measurements



(b) normalized mean L1 error vs number of measurements



(c) mean coherence vs number of measurements

Fig. 6: Experiments of Group B: The normalized mean L1 and L2 errors and mean coherence values with various number of measurements for a fixed uncertainty parameter.

References

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

[2] 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.

[3] D. L. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.

[4] A. V. Oppenheim, R. W. Schafer, J. R. Buck *et al.*, *Discrete-time signal processing*. New Jersey, USA: Prentice Hall, 1999, vol. 5.

[5] 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.

[6] 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.

[7] M. A. Herman and D. Needell, "Mixed operators in compressed sensing," in *Information Sciences and Systems (CISS), 2010 44th Annual Conference on*. IEEE, 2010, pp. 1–6.

[8] 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.

[9] 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.

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

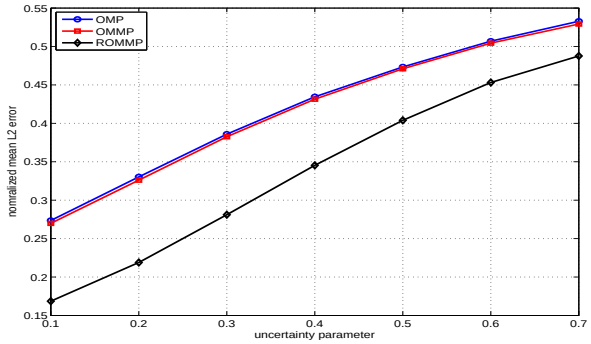
[11] H. Zhu, G. Leus, and G. B. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *Signal Processing, IEEE Transactions on*, vol. 59, no. 5, pp. 2002–2016, 2011.

[12] M. Rosenbaum and A. B. Tsybakov, "Sparse recovery under matrix uncertainty," *The Annals of Statistics*, vol. 38, no. 5, pp. 2620–2651, 2010.

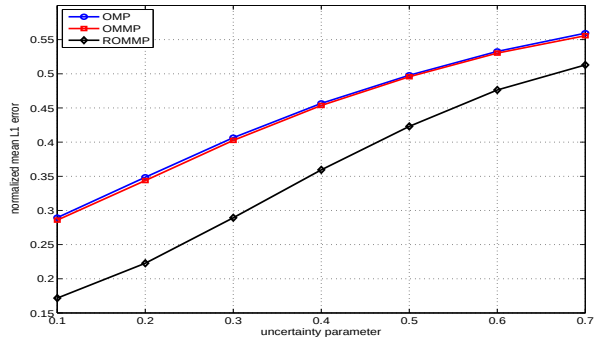
[13] J. T. Parker, V. Cevher, and P. Schniter, "Compressive sensing under matrix uncertainties: An approximate message passing approach," in *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on*. IEEE, 2011, pp. 804–808.

[14] F. Krzakala, M. Mézard, and L. Zdeborová, "Compressed sensing under matrix uncertainty: Optimum thresholds and robust approximate message passing," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 5519–5523.

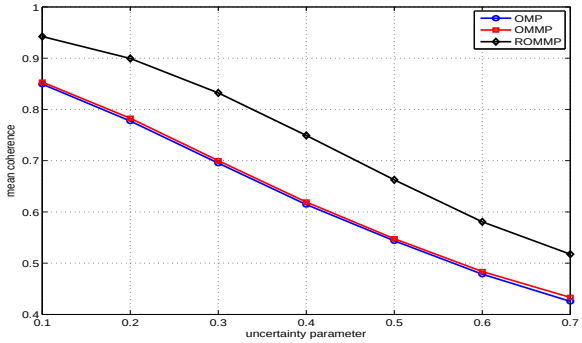
[15] Z. Yang, C. Zhang, and L. Xie, "Robustly stable signal recovery



(a) normalized mean L2 error vs vs uncertainty parameter

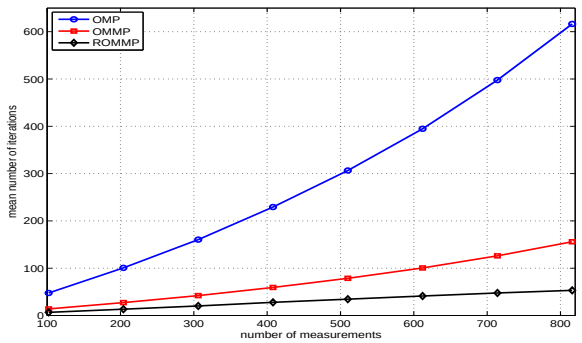


(b) normalized mean L1 error vs uncertainty parameter

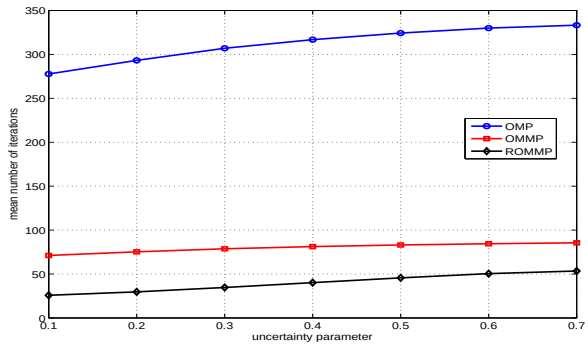


(c) mean coherence vs uncertainty parameter

Fig. 7: Experiments of Group B: The normalized mean L1 and L2 errors and mean coherence with various uncertainty parameter for a fixed number of measurements.



(a) mean number of iterations vs number of measurements



(b) mean number of iterations vs uncertainty parameter

Fig. 8: Experiments of Group B: The mean number of iterations at various number of measurements and various uncertainty parameters of additive and multiplicative noise.

in compressed sensing with structured matrix perturbation," *Signal Processing, IEEE Transactions on*, vol. 60, no. 9, pp. 4658–4671, 2012.

[16] 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.

[17] 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.

[18] A. Tabibiazar and O. Basir, "Energy-efficient compressive state recovery from sparsely noisy measurements," *Instrumentation and Measurement, IEEE Transactions on*, vol. 61, no. 9, pp. 2392–2400, 2012.

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

[20] J. Li and P. Stoica, *Robust adaptive beamforming*. New Jersey, USA: John Wiley & Sons, 2006.

[21] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, *Robust optimization*. New Jersey, USA: Princeton University Press, 2009.

[22] W. A. Fuller, *Measurement error models*. New Jersey, USA: John Wiley & Sons, 1987.

[23] C.-L. Cheng and J. W. Van Ness, *Statistical regression with measurement error*. New Jersey, USA: John Wiley & Sons, 1999.

[24] S. Van Huffel and J. Vandewalle, *The total least squares problem: computational aspects and analysis*. Philadelphia, USA: SIAM, 1991.

[25] P. P. Gallo, "Consistency of regression estimates when some variables are subject to error," *Communications in Statistics-Theory and Methods*, vol. 11, no. 9, pp. 973–983, 1982.

[26] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.

[27] 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.

[28] 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.

[29] 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.

[30] Y. Nesterov, "Smooth minimization of non-smooth functions," *Mathematical programming*, vol. 103, no. 1, pp. 127–152, 2005.

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

[32] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 21–30, 2008.

[33] J. D. Blanchard, C. Cartis, and J. Tanner, "Compressed sensing: How sharp is the restricted isometry property?" *SIAM review*, vol. 53, no. 1, pp. 105–125, 2011.

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

[35] B. Shim, J. Wang, and S. Kwon, "Generalized orthogonal matching

pursuit," *Signal Processing, IEEE Transactions on*, vol. 60, no. 12, pp. 6202–6216, 2012.

[36] 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.

[37] 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.

[38] 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.

[39] 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.