

Robustly Stable Signal Recovery in Compressed Sensing with Structured Matrix Perturbation

Zai Yang, Cishen Zhang, and Lihua Xie*, *Fellow, IEEE*

Abstract—The sparse signal recovery in standard compressed sensing (CS) requires that the sensing matrix be known *a priori*. The CS problem subject to perturbation in the sensing matrix is often encountered in practice and results in the literature have shown that the signal recovery error grows linearly with the perturbation level. This paper assumes a structure for the perturbation. Under mild conditions on the perturbed sensing matrix, it is shown that a sparse signal can be recovered by ℓ_1 minimization with the recovery error being at most proportional to the measurement noise level, similar to that in the standard CS. The recovery is exact in the special noise free case provided that the signal is sufficiently sparse with respect to the perturbation level. A similar result holds for compressible signals under an additional assumption of small perturbation. Algorithms are proposed for implementing the ℓ_1 minimization problem and numerical simulations are carried out that verify our analysis.

Index Terms—Compressed sensing, structured matrix perturbation, robustly stable signal recovery, alternating algorithm.

I. INTRODUCTION

Compressed sensing (CS) has been a very active research area since the pioneering works of Candès *et al.* [1], [2] and Donoho [3]. In CS, a signal $\mathbf{x}^o \in \mathbb{R}^n$ of length n is called k -sparse if it has at most k nonzero entries, and it is called compressible if its entries obey a power law

$$|x^o|_{(i)} \leq C_q i^{-q}, \quad (1)$$

where $|x^o|_{(i)}$ is the i th largest entry (in absolute value) of \mathbf{x}^o ($|x^o|_{(1)} \geq |x^o|_{(2)} \geq \dots \geq |x^o|_{(n)}$), $q > 1$ and C_q is a constant that depends only on q . Let \mathbf{x}^k be the truncated vector corresponding to the k largest entries (in absolute value) of \mathbf{x}^o . If \mathbf{x}^o is compressible, then it can be well approximated by the sparse signal \mathbf{x}^k in the sense that

$$\|\mathbf{x}^o - \mathbf{x}^k\|_2 \leq C'_q k^{-q+1/2} \quad (2)$$

where C'_q is a constant as C_q . To obtain the knowledge of \mathbf{x}^o , CS acquires linear measurements of \mathbf{x}^o as

$$\mathbf{y} = \Phi \mathbf{x}^o + \mathbf{e}, \quad (3)$$

where $\Phi \in \mathbb{R}^{m \times n}$ is the sensing matrix (or linear operator) with typically $k < m \ll n$, $\mathbf{y} \in \mathbb{R}^m$ is the vector of measurements, and $\mathbf{e} \in \mathbb{R}^m$ denotes the vector of measurement noises with bounded energy, i.e., $\|\mathbf{e}\|_2 \leq \epsilon$ for $\epsilon > 0$. Given

Φ and ϵ , the task of CS is to recover \mathbf{x}^o from a significantly reduced number of measurements \mathbf{y} . Candès *et al.* [1], [4] show that if \mathbf{x}^o is sparse, then it can be stably recovered under mild conditions on Φ with the recovery error being at most proportional to the measurement noise level ϵ by solving an ℓ_1 minimization problem. Similarly, the largest entries (in absolute value) of a compressible signal can be stably recovered. More details are presented in Subsection II-B. In addition to the ℓ_1 minimization, other approaches that provide similar guarantees are also reported thereafter, such as IHT [5] and greedy pursuit methods including OMP [6], StOMP [7] and CoSaMP [8].

The sensing matrix Φ is assumed known *a priori* in the standard CS, which is, however, not always the case in practical situations. For example, a matrix perturbation can be caused by quantization during implementation. In source separation [9], [10] the sensing matrix (or mixing system) is usually unknown and needs to be estimated, and thus estimation errors exist. In source localization such as direction of arrival (DOA) estimation [11], [12] and radar imaging [13], [14], the sensing matrix (overcomplete dictionary) is constructed via discretizing one or more continuous parameters, and errors exist typically in the sensing matrix since the true source locations may not be exactly on a discretized sampling grid. As a result, there have been recent active studies on CS where the sensing matrix is subject to an unknown perturbation. Herman and Strohmer [15] analyze the effect of a general matrix perturbation and show that the signal recovery is robust to the perturbation in the sense that the recovery error grows linearly with the perturbation level. Similar robust recovery results are also reported in [16], [17]. It is demonstrated in [17], [18] that the signal recovery may suffer from a large error under a large perturbation. In addition, the existence of recovery error caused by the perturbed sensing matrix is independent of the sparsity of the original signal. Algorithms have also been proposed to deal with sensing matrix perturbations. Zhu *et al.* [19] propose a sparse total least-squares approach to alleviating the effect of perturbation where they explore the structure of the perturbation to improve recovery performance. Yang *et al.* [12] formulate the off-grid DOA estimation problem from a sparse Bayesian inference perspective and iteratively recover the source signal and the matrix perturbation. It should be noted that existing algorithmic results provide no guarantees on signal recovery performance when there exist perturbations in the sensing matrix.

This paper is on the perturbed CS problem. A structured matrix perturbation is studied with each column of the perturbation matrix being a (unknown) constant times a (known)

* Author for correspondence.

Z. Yang and L. Xie are with EXQUISITUS, Centre for E-City, School of Electrical and Electronic Engineering, Nanyang Technological University, 639798, Singapore (e-mail: yang0248@e.ntu.edu.sg; elhxie@ntu.edu.sg).

C. Zhang is with the Faculty of Engineering and Industrial Sciences, Swinburne University of Technology, Hawthorn VIC 3122, Australia (e-mail: cishenzhang@swin.edu.au).

vector which defines the direction of perturbation. For certain structured matrix perturbation, we provide conditions for guaranteed signal recovery performance. Our analysis shows that robust stability (see definition in Subsection II-A) can be achieved for a sparse signal under similar mild conditions as those for the standard CS problem by solving an ℓ_1 minimization problem incorporated with the perturbation structure. In the special noise free case, the recovery is exact for a sufficiently sparse signal with respect to the perturbation level. A similar result holds for a compressible signal under an additional assumption of small perturbation (depending on the number of largest entries to be recovered). To verify our analysis, two algorithms for positive-valued and general signals respectively are proposed to solve the resulting nonconvex ℓ_1 minimization problem. Numerical simulations confirm our robustly stable signal recovery results.

A common approach in CS to signal recovery is solving an optimization problem, e.g., ℓ_1 minimization. Another contribution of this paper is to characterize a set of solutions to the optimization problem that can be good estimates of the signal to be recovered, which indicates that it is not necessary to obtain the optimal solution to the optimization problem. This is helpful to assess the “effectiveness” of an algorithm (see definition in Subsection III-E), for example, the ℓ_p ($p < 1$) minimization [20], [21] in the standard CS, in solving the optimization problem as in nonconvex optimization the output of an algorithm cannot be guaranteed to be the optimal solution.

Notations used in this paper are as follows. Bold-case letters are reserved for vectors and matrices. $\|\mathbf{x}\|_0$ denotes the pseudo ℓ_0 norm that counts the number of nonzero entries of a vector \mathbf{x} . $\|\mathbf{x}\|_1$ and $\|\mathbf{x}\|_2$ denote the ℓ_1 and ℓ_2 norms of a vector \mathbf{x} respectively. $\|\mathbf{A}\|_2$ and $\|\mathbf{A}\|_F$ are the spectral and Frobenius norms of a matrix \mathbf{A} respectively. \mathbf{x}^T is the transpose of a vector \mathbf{x} and \mathbf{A}^T is for a matrix \mathbf{A} . x_i is the i th entry of a vector \mathbf{x} . T^c is the complementary set of a set T . Unless otherwise stated, \mathbf{x}_T has entries of a vector \mathbf{x} on an index set T and zero entries on T^c . $\text{diag}(\mathbf{x})$ is a diagonal matrix with its diagonal entries being entries of a vector \mathbf{x} . \odot is the Hadamard (elementwise) product.

The rest of the paper is organized as follows. Section II first defines formally some terminologies used in this paper and then introduces existing results on the standard CS and perturbed CS. Section III presents the main results of the paper with some discussions. Section IV introduces algorithms for the ℓ_1 minimization problem in our considered perturbed CS and their analysis. Section V presents a series of numerical simulations to verify our main results. Conclusions are drawn in Section VI. Finally, proofs of some results in Section III and Section IV are provided in Appendices.

II. PRELIMINARY RESULTS

A. Definitions

For the purpose of clarification of expression, we define formally some terminologies for signal recovery used in this paper, including stability in the standard CS, robustness and robust stability in the perturbed CS.

Definition 1 ([1]): In the standard CS where Φ is known *a priori*, consider a recovered signal $\hat{\mathbf{x}}$ of \mathbf{x}^o from measurements $\mathbf{y} = \Phi\mathbf{x}^o + \mathbf{e}$ with $\|\mathbf{e}\|_2 \leq \epsilon$. We call that $\hat{\mathbf{x}}$ achieves stable signal recovery if

$$\|\hat{\mathbf{x}} - \mathbf{x}^o\|_2 \leq C_1^{stb} k^{-q+1/2} + C_2^{stb} \epsilon$$

holds for compressible signal \mathbf{x}^o obeying (1) and an integer k , or if

$$\|\hat{\mathbf{x}} - \mathbf{x}^o\|_2 \leq C_2^{stb} \epsilon$$

holds for k -sparse signal \mathbf{x}^o , with nonnegative constants C_1^{stb} , C_2^{stb} .

Definition 2: In the perturbed CS where $\Phi = \mathbf{A} + \mathbf{E}$ with \mathbf{A} known *a priori* and \mathbf{E} unknown with $\|\mathbf{E}\|_F \leq \eta$, consider a recovered signal $\hat{\mathbf{x}}$ of \mathbf{x}^o from measurements $\mathbf{y} = \Phi\mathbf{x}^o + \mathbf{e}$ with $\|\mathbf{e}\|_2 \leq \epsilon$. We call that $\hat{\mathbf{x}}$ achieves robust signal recovery if

$$\|\hat{\mathbf{x}} - \mathbf{x}^o\|_2 \leq C_1^{rbt} k^{-q+1/2} + C_2^{rbt} \epsilon + C_3^{rbt} \eta$$

holds for compressible signal \mathbf{x}^o obeying (1) and an integer k , or if

$$\|\hat{\mathbf{x}} - \mathbf{x}^o\|_2 \leq C_2^{rbt} \epsilon + C_3^{rbt} \eta$$

holds for k -sparse signal \mathbf{x}^o , with nonnegative constants C_1^{rbt} , C_2^{rbt} and C_3^{rbt} .

Definition 3: In the perturbed CS where $\Phi = \mathbf{A} + \mathbf{E}$ with \mathbf{A} known *a priori* and \mathbf{E} unknown with $\|\mathbf{E}\|_F \leq \eta$, consider a recovered signal $\hat{\mathbf{x}}$ of \mathbf{x}^o from measurements $\mathbf{y} = \Phi\mathbf{x}^o + \mathbf{e}$ with $\|\mathbf{e}\|_2 \leq \epsilon$. We call that $\hat{\mathbf{x}}$ achieves robustly stable signal recovery if

$$\|\hat{\mathbf{x}} - \mathbf{x}^o\|_2 \leq C_1^{rs}(\eta) k^{-q+1/2} + C_2^{rs}(\eta) \epsilon$$

holds for compressible signal \mathbf{x}^o obeying (1) and an integer k , or if

$$\|\hat{\mathbf{x}} - \mathbf{x}^o\|_2 \leq C_2^{rs}(\eta) \epsilon$$

holds for k -sparse signal \mathbf{x}^o , with nonnegative constants C_1^{rs} , C_2^{rs} depending on η .

Remark 1:

- (1) In the case where \mathbf{x}^o is compressible, the defined stable, robust, or robustly stable signal recovery is in fact for its k largest entries (in absolute value). The first term $O(k^{-q+1/2})$ in the error bounds above represents, by (2), the best approximation error (up to a scale) that can be achieved when we know everything about \mathbf{x}^o and select its k largest entries.
- (2) The Frobenius norm of \mathbf{E} , $\|\mathbf{E}\|_F$, can be replaced by any other norm in Definitions 2 and 3 since the norms are equivalent.

It should be noted that the stable recovery in the standard CS and the robustly stable recovery in the perturbed CS are exact in the noise free, sparse signal case while there is no such a guarantee for the robust recovery in the perturbed CS.

B. Stable Signal Recovery of Standard CS

The task of the standard CS is to recover the original signal \mathbf{x}^o via an efficient approach given the sensing matrix Φ , acquired sample \mathbf{y} and upper bound ϵ for the measurement noise. This paper focuses on the ℓ_1 norm minimization approach. The restricted isometry property (RIP) [22] has become a dominant tool to such analysis, which is defined as follows.

Definition 4: Define the k -restricted isometry constant (RIC) of a matrix Φ , denoted by $\delta_k(\Phi)$, as the smallest number such that

$$(1 - \delta_k(\Phi)) \|\mathbf{v}\|_2^2 \leq \|\Phi\mathbf{v}\|_2^2 \leq (1 + \delta_k(\Phi)) \|\mathbf{v}\|_2^2$$

holds for all k -sparse vectors \mathbf{v} . Φ is said to satisfy the k -RIP with constant $\delta_k(\Phi)$ if $\delta_k(\Phi) < 1$.

Based on the RIP, the following theorem holds.

Theorem 1 ([4]): Assume that $\delta_{2k}(\Phi) < \sqrt{2} - 1$ and $\|\mathbf{e}\|_2 \leq \epsilon$. Then an optimal solution \mathbf{x}^* to the basis pursuit denoising (BPDN) problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1, \text{ subject to } \|\mathbf{y} - \Phi\mathbf{x}\|_2 \leq \epsilon \quad (4)$$

satisfies

$$\|\mathbf{x}^* - \mathbf{x}^o\|_2 \leq C_0^{std} k^{-1/2} \|\mathbf{x}^o - \mathbf{x}^k\|_1 + C_1^{std} \epsilon \quad (5)$$

where $C_0^{std} = \frac{2[1+(\sqrt{2}-1)\delta_{2k}(\Phi)]}{1-(\sqrt{2}+1)\delta_{2k}(\Phi)}$, $C_1^{std} = \frac{4\sqrt{1+\delta_{2k}(\Phi)}}{1-(\sqrt{2}+1)\delta_{2k}(\Phi)}$.

Theorem 1 states that a k -sparse signal \mathbf{x}^o ($\mathbf{x}^k = \mathbf{x}^o$) can be stably recovered by solving a computationally efficient convex optimization problem provided $\delta_{2k}(\Phi) < \sqrt{2} - 1$. The same conclusion holds in the case of compressible signal \mathbf{x}^o since

$$k^{-1/2} \|\mathbf{x}^o - \mathbf{x}^k\|_1 \leq C_q'' k^{-q+1/2}$$

according to (1) and (2) with C_q'' being a constant as C_q in (1). In the special noise free, k -sparse signal case, such a recovery is exact. The RIP condition in Theorem 1 can be satisfied provided $m \geq O(k \log(n/k))$ with a large probability if the sensing matrix Φ is i.i.d. subgaussian distributed [23]. Note that the RIP condition for the stable signal recovery in the standard CS has been relaxed in [24], [25] but it is beyond the scope of this paper.

C. Robust Signal Recovery in Perturbed CS

In the standard CS, the sensing matrix Φ is assumed to be exactly known. Such an ideal assumption is not always the case in practice. Consider that the true sensing matrix is $\Phi = \mathbf{A} + \mathbf{E}$ where $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the known nominal sensing matrix and $\mathbf{E} \in \mathbb{R}^{m \times n}$ represents the unknown matrix perturbation. Unlike the additive noise term \mathbf{e} in the observation model in (3), a multiplicative “noise” $\mathbf{E}\mathbf{x}^o$ is introduced in the perturbed CS and is more difficult to analyze since it is correlated with the signal of interest. Denote $\|\mathbf{E}\|_2^{(k)}$ the largest spectral norm taken over all k -column submatrices of \mathbf{E} , and similarly define $\|\Phi\|_2^{(k)}$. The following theorem is stated in [15].

Theorem 2 ([15]): Assume that there exist constants $\varepsilon_{\mathbf{E},\Phi}^{(k)}$, ϵ and $\epsilon_{\mathbf{E},\mathbf{x}^o}$ such that $\frac{\|\mathbf{E}\|_2^{(k)}}{\|\Phi\|_2^{(k)}} \leq \varepsilon_{\mathbf{E},\Phi}^{(k)}$, $\|\mathbf{e}\|_2 \leq \epsilon$ and $\|\mathbf{E}\mathbf{x}^o\|_2 \leq \epsilon_{\mathbf{E},\mathbf{x}^o}$. Assume that $\delta_{2k}(\Phi) < \frac{\sqrt{2}}{(1+\varepsilon_{\mathbf{E},\Phi}^{(2k)})^2} - 1$ and $\|\mathbf{x}^o\|_0 \leq k$.

Then an optimal solution \mathbf{x}^* to the BPDN problem with the nominal sensing matrix \mathbf{A} , denoted by N-BPDN,

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1, \text{ subject to } \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \epsilon + \epsilon_{\mathbf{E},\mathbf{x}^o} \quad (6)$$

achieves robust signal recovery with

$$\|\mathbf{x}^* - \mathbf{x}^o\|_2 \leq C^{ptb} \epsilon + C^{ptb} \epsilon_{\mathbf{E},\mathbf{x}^o} \quad (7)$$

where $C^{ptb} = \frac{4\sqrt{1+\delta_{2k}(\Phi)}(1+\varepsilon_{\mathbf{E},\Phi}^{(2k)})}{1-(\sqrt{2}+1)[(1+\delta_{2k}(\Phi))(1+\varepsilon_{\mathbf{E},\Phi}^{(2k)})^2-1]}$.

Remark 2:

- (1) The relaxation of the inequality constraint in (6) from ϵ to $\epsilon + \epsilon_{\mathbf{E},\mathbf{x}^o}$ is to ensure that the original signal \mathbf{x}^o is a feasible solution to N-BPDN. Theorem 2 is a little different from that in [15], where the multiplicative “noise” $\mathbf{E}\mathbf{x}^o$ is bounded using $\varepsilon_{\mathbf{E},\Phi}^{(k)}$, $\delta_k(\Phi)$ and $\|\Phi\mathbf{x}^o\|_2$ rather than a constant $\epsilon_{\mathbf{E},\mathbf{x}^o}$.
- (2) Theorem 2 is applicable only to the small perturbation case where $\varepsilon_{\mathbf{E},\Phi}^{(2k)} < \sqrt[4]{2} - 1$ since $\delta_{2k}(\Phi) \geq 0$.
- (3) Theorem 2 generalizes Theorem 1 for the k -sparse signal case. As the perturbation $\mathbf{E} \rightarrow 0$, Theorem 2 coincides with Theorem 1 for the k -sparse signal case.

Theorem 2 states that, for a small matrix perturbation \mathbf{E} , the signal recovery of N-BPDN that is based on the nominal sensing matrix \mathbf{A} is robust to the perturbation with the recovery error growing at most linearly with the perturbation level. Note that, in general, the signal recovery in Theorem 2 is unstable according to the definition of stability in this paper since the recovery error cannot be bounded within a constant (independent of the noise) times the noise level as some perturbation occurs. A result on general signals in [15] is omitted that shows the robust recovery of a compressible signal. The same problem is studied and similar results are reported in [16] based on the greedy algorithm CoSaMP [8].

III. SP-CS: CS SUBJECT TO STRUCTURED PERTURBATION

A. Problem Description

In this paper we consider a structured perturbation in the form $\mathbf{E} = \mathbf{B}\Delta^o$ where $\mathbf{B} \in \mathbb{R}^{m \times n}$ is known *a priori*, $\Delta^o = \text{diag}(\beta^o)$ is a bounded uncertain term with $\beta^o \in [-r, r]^n$ and $r > 0$, i.e., each column of the perturbation is on a known direction. In addition, we assume that each column of \mathbf{B} has unit norm to avoid the scaling problem between \mathbf{B} and Δ^o .¹ As a result, the observation model in (3) becomes

$$\mathbf{y} = \Phi\mathbf{x}^o + \mathbf{e}, \quad \Phi = \mathbf{A} + \mathbf{B}\Delta^o \quad (8)$$

with $\Delta^o = \text{diag}(\beta^o)$, $\beta^o \in [-r, r]^n$ and $\|\mathbf{e}\|_2 \leq \epsilon$. Given \mathbf{y} , \mathbf{A} , \mathbf{B} , r and ϵ , the task of SP-CS is to recover \mathbf{x}^o , possibly, as well as β^o .

Remark 3: If $x_i^o = 0$ for some $i \in \{1, \dots, n\}$, then β_i^o has no contributions to the observation \mathbf{y} and hence it is impossible to recover β_i^o . As a result, the recovery of β^o in this paper refers only to the recovery on the support of \mathbf{x}^o .

¹In fact, the D-RIP condition on matrix $[\mathbf{A}, \mathbf{B}]$ in Subsection III-B implies that columns of both \mathbf{A} and \mathbf{B} have approximately unit norms.

B. Main Results of This Paper

In this paper, a vector \mathbf{v} is called $2k$ -duplicately (D-) sparse if $\mathbf{v} = [\mathbf{v}_1^T, \mathbf{v}_2^T]^T$ with \mathbf{v}_1 and \mathbf{v}_2 being of the same dimension and jointly k -sparse (each being k -sparse and sharing the same support). The concept of duplicate (D-) RIP is defined as follows.

Definition 5: Define the $2k$ -duplicate (D-) RIC of a matrix Φ , denoted by $\bar{\delta}_{2k}(\Phi)$, as the smallest number such that

$$(1 - \bar{\delta}_{2k}(\Phi)) \|\mathbf{v}\|_2^2 \leq \|\Phi\mathbf{v}\|_2^2 \leq (1 + \bar{\delta}_{2k}(\Phi)) \|\mathbf{v}\|_2^2$$

holds for all $2k$ -D-sparse vectors \mathbf{v} . Φ is said to satisfy the $2k$ -D-RIP with constant $\bar{\delta}_{2k}(\Phi)$ if $\bar{\delta}_{2k}(\Phi) < 1$.

With respect to the perturbed observation model in (8), let $\Psi = [\mathbf{A}, \mathbf{B}]$. The main results of this paper are stated in the following theorems. The proof of Theorem 3 is provided in Appendix A and proofs of Theorems 4 and 5 are in Appendix B.

Theorem 3: In the noise free case where $\mathbf{e} = \mathbf{0}$, assume that $\|\mathbf{x}^o\|_0 \leq k$ and $\bar{\delta}_{4k}(\Psi) < 1$. Then an optimal solution (\mathbf{x}^*, β^*) to the perturbed combinatorial optimization problem

$$\min_{\mathbf{x} \in \mathbb{R}^n, \beta \in [-r, r]^n} \|\mathbf{x}\|_0, \text{ subject to } \mathbf{y} = (\mathbf{A} + \mathbf{B}\Delta)\mathbf{x} \quad (9)$$

with $\Delta = \text{diag}(\beta)$ recovers \mathbf{x}^o and β^o .

Theorem 4: Assume that $\bar{\delta}_{4k}(\Psi) < \left(\sqrt{2(1+r^2)} + 1\right)^{-1}$, $\|\mathbf{x}^o\|_0 \leq k$ and $\|\mathbf{e}\|_2 \leq \epsilon$. Then an optimal solution (\mathbf{x}^*, β^*) to the perturbed (P-) BPDN problem

$$\min_{\mathbf{x} \in \mathbb{R}^n, \beta \in [-r, r]^n} \|\mathbf{x}\|_1, \text{ subject to } \|\mathbf{y} - (\mathbf{A} + \mathbf{B}\Delta)\mathbf{x}\|_2 \leq \epsilon \quad (10)$$

achieves robustly stable signal recovery with

$$\|\mathbf{x}^* - \mathbf{x}^o\|_2 \leq C\epsilon, \quad (11)$$

$$\|(\beta^* - \beta^o) \odot \mathbf{x}^o\|_2 \leq \mathcal{C}\epsilon \quad (12)$$

where

$$C = \frac{4\sqrt{1 + \bar{\delta}_{4k}(\Psi)}}{1 - \left(\sqrt{2(1+r^2)} + 1\right) \bar{\delta}_{4k}(\Psi)},$$

$$\mathcal{C} = \frac{[2 + \sqrt{1+r^2} \|\Psi\|_2 C]}{\sqrt{1 - \bar{\delta}_{4k}(\Psi)}}.$$

Theorem 5: Assume that $\bar{\delta}_{4k}(\Psi) < \left(\sqrt{2(1+r^2)} + 1\right)^{-1}$ and $\|\mathbf{e}\|_2 \leq \epsilon$. Then an optimal solution (\mathbf{x}^*, β^*) to the P-BPDN problem in (10) satisfies that

$$\|\mathbf{x}^* - \mathbf{x}^o\|_2 \leq \left(C_0 k^{-1/2} + C_1\right) \|\mathbf{x}^o - \mathbf{x}^k\|_1 + C_2 \epsilon, \quad (13)$$

$$\begin{aligned} \|(\beta^* - \beta^o) \odot \mathbf{x}^k\|_2 \\ \leq \left(C_0 k^{-1/2} + C_1\right) \|\mathbf{x}^o - \mathbf{x}^k\|_1 + C_2 \epsilon \end{aligned} \quad (14)$$

where

$$C_0 = 2 \left[1 + \left(\sqrt{2(1+r^2)} - 1\right) \bar{\delta}_{4k}(\Psi)\right] / a,$$

$$C_1 = 2\sqrt{2}r\bar{\delta}_{4k}(\Psi) / a,$$

$$C_0 = \sqrt{1+r^2} \|\Psi\|_2 C_0 / b,$$

$$C_1 = \left[\sqrt{1+r^2} C_1 + 2r\right] \|\Psi\|_2 / b$$

with $a = 1 - \left(\sqrt{2(1+r^2)} + 1\right) \bar{\delta}_{4k}(\Psi)$, $b = \sqrt{1 - \bar{\delta}_{4k}(\Psi)}$ and $C_2 = C$, $\mathcal{C}_2 = \mathcal{C}$ with C, \mathcal{C} as defined in Theorem 4.

Remark 4: In general, the robustly stable signal recovery cannot be concluded for compressible signals since the error bound in (13) may be very large in the case of large perturbation by $C_1 = O(r)$. If the perturbation is small with $r = O(k^{-1/2})$, then the robust stability can be achieved for compressible signals provided that the D-RIP condition in Theorem 5 is satisfied.

C. Interpretation of the Main Results

Theorem 3 states that for a k -sparse signal \mathbf{x}^o , it can be recovered by solving a combinatorial optimization problem provided $\bar{\delta}_{4k}(\Psi) < 1$ when the measurements are exact. Meanwhile, β^o can be recovered. Since the combinatorial optimization problem is NP-hard and that its solution is sensitive to measurement noise [26], a more reliable approach, ℓ_1 minimization, is explored in Theorems 4 and 5.

Theorem 4 states the robustly stable recovery of a k -sparse signal \mathbf{x}^o in SP-CS with the recovery error being at most proportional to the noise level. Such robust stability is obtained by solving an ℓ_1 minimization problem incorporated with the perturbation structure provided that the D-RIC is sufficiently small with respect to the perturbation level in terms of r . Meanwhile, the perturbation parameter β^o can be stably recovered on the support of \mathbf{x}^o . As the D-RIP condition is satisfied in Theorem 4, the signal recovery error of the perturbed CS is constrained by the noise level ϵ , and the influence of the perturbation is limited to the coefficient before ϵ . For example, if $\bar{\delta}_{4k}(\Psi) = 0.2$, then $\|\mathbf{x}^* - \mathbf{x}^o\|_2 \leq 8.48\epsilon, 8.50\epsilon, 11.0\epsilon$ corresponding to $r = 0.01, 0.1, 1$, respectively. In the special noise free case, the recovery is exact. This is similar to that in the standard CS but in contrast to the existing robust signal recovery result in Subsection II-C where the recovery error exists once a matrix perturbation appears. Another interpretation of the D-RIP condition in Theorem 4 is that the robustly stable signal recovery requires that $r < \sqrt{\frac{1}{2} \left(\bar{\delta}_{4k}(\Psi)^{-1} - 1\right)^2} - 1$ for a fixed matrix Ψ . Using the aforementioned example where $\bar{\delta}_{4k}(\Psi) = 0.2$, the perturbation is required to satisfy $r < \sqrt{7}$. As a result, our robustly stable signal recovery result of SP-CS applies to the case of large perturbation if the D-RIC of Ψ is sufficiently small while the existing result does not as demonstrated in Remark 2.

Theorem 5 considers general signals and is a generalized form of Theorem 4. In comparison with Theorem 1 in the standard CS, one more term $C_1 \|\mathbf{x}^o - \mathbf{x}^k\|_1$ appears in the upper bound of the recovery error. The robust stability does not hold generally for compressible signals as illustrated in Remark 4 while it is true under an additional assumption $r = O(k^{-1/2})$.

The results in this paper generalize that in the standard CS. Without accounting for the symbolic difference between $\bar{\delta}_{2k}(\Phi)$ and $\bar{\delta}_{4k}(\Psi)$, the conditions in Theorems 1 and 5 coincide, as well as the upper bounds in (5) and (13) for the recovery errors, as the perturbation vanishes or equivalently $r \rightarrow 0$. As mentioned before, the RIP condition for guaranteed stable recovery in the standard CS has been relaxed.

Similar techniques may be adopted to possibly relax the D-RIP condition in SP-CS. While this paper is focused on the ℓ_1 minimization approach, it is also possible to modify other algorithms in the standard CS and apply them to SP-CS to provide similar recovery guarantees.

D. When is the D-RIP satisfied?

Existing works studying the RIP mainly focus on random matrices. In the standard CS, Φ has the k -RIP with constant δ with a large probability provided that $m \geq C_\delta k \log(n/k)$ and Φ has properly scaled i.i.d. subgaussian distributed entries with constant C_δ depending on δ and the distribution [23]. The D-RIP can be considered as a model-based RIP introduced in [27]. Suppose that \mathbf{A} , \mathbf{B} are mutually independent and both are i.i.d. subgaussian distributed (the true sensing matrix $\Phi = \mathbf{A} + \mathbf{B}\Delta^\circ$ is also i.i.d. subgaussian distributed if β° is independent of \mathbf{A} and \mathbf{B}). The model-based RIP is determined by the number of subspaces of the structured sparse signals that are referred to as the D-sparse ones in the present paper. For $\Psi = [\mathbf{A}, \mathbf{B}]$, the number of $2k$ -dimensional subspaces for $2k$ -D-sparse signals is $\binom{n}{k}$. Consequently, Ψ has the $2k$ -D-RIP with constant δ with a large probability also provided that $m \geq C_\delta k \log(n/k)$ by [27, Theorem 1] or [28, Theorem 3.3]. So, in the case of high dimensional system, the D-RIP condition on Ψ , as $r \rightarrow 0$, in Theorem 4 or 5 for guaranteed robustly stable signal recovery of SP-CS can be satisfied once the RIP condition on Φ (after proper scaling of columns) in the standard CS is met. It means that the perturbation in SP-CS gradually strengthens the D-RIP condition for robustly stable signal recovery but there exists no gap between SP-CS and the standard CS in the case of high dimensional systems.

It is noted that there is another way to stably recover the original signal \mathbf{x}° in SP-CS. Given the sparse signal case as an example where \mathbf{x}° is k -sparse. Let $\mathbf{z}^\circ = \begin{bmatrix} \mathbf{x}^\circ \\ \beta^\circ \odot \mathbf{x}^\circ \end{bmatrix}$, and it is $2k$ -sparse. The observation model can be written as $\mathbf{y} = \Psi \mathbf{z}^\circ + \mathbf{e}$. Then \mathbf{z}° and hence, \mathbf{x}° , can be stably recovered from the problem²

$$\min_{\mathbf{z}} \|\mathbf{z}\|_1, \text{ subject to } \|\mathbf{y} - \Psi \mathbf{z}\|_2 \leq \epsilon \quad (15)$$

provided that $\delta_{4k}(\Psi) < \sqrt{2} - 1$ by Theorem 1. It looks like that we transformed the perturbation into a signal of interest. Denote TPS-BPDN the problem in (15). In a high dimensional system, the condition $\delta_{4k}(\Psi) < \sqrt{2} - 1$ requires about twice as many as the measurements that makes the D-RIP condition $\bar{\delta}_{4k}(\Psi) < \sqrt{2} - 1$ hold by [27, Theorem 1] corresponding to the D-RIP condition in Theorem 4 or 5 as $r \rightarrow 0$. As a result, for a considerable range of perturbation level, the D-RIP condition in Theorem 4 or 5 for P-BPDN is weaker than that for TPS-BPDN since it varies slowly for a moderate perturbation (as an example, $\bar{\delta}_{4k}(\Psi) < 0.414, 0.413, 0.409$ corresponds to $r = 0, 0.1, 0.2$ respectively). Numerical simulations in Subsection V can verify our conclusion.

²It is hard to incorporate the knowledge $\beta^\circ \in [-r, r]^n$ into the problem in (15).

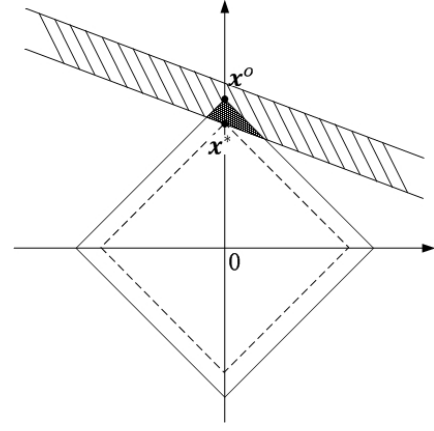


Fig. 1. Illustration of Corollary 2. The shaded band area refers to the feasible domain of BPDN. The triangular area, intersection of the feasible domain and the ℓ_1 ball $\{\mathbf{x} : \|\mathbf{x}\|_1 \leq \|\mathbf{x}^\circ\|_1\}$, is the set of all good recoveries.

E. Relaxation of the Optimal Solution

In Theorem 5 (Theorem 4 is a special case), (\mathbf{x}^*, β^*) is required to be an optimal solution to P-BPDN. Naturally, we would like to know if the requirement of the optimality is necessary for a “good” recovery in the sense that a good recovery validates the error bounds in (13) and (14) under the conditions in Theorem 5. Generally speaking, the answer is negative since, regarding the optimality of (\mathbf{x}^*, β^*) , only $\|\mathbf{x}^*\|_1 \leq \|\mathbf{x}^\circ\|_1$ and the feasibility of (\mathbf{x}^*, β^*) are used in the proof of Theorem 5 in Appendix B. Denote \mathcal{D} the feasible domain of P-BPDN, i.e.,

$$\mathcal{D} = \{(\mathbf{x}, \beta) : \beta \in [-r, r]^n, \|\mathbf{y} - (\mathbf{A} + \mathbf{B}\Delta)\mathbf{x}\|_2 \leq \epsilon \text{ with } \Delta = \text{diag}(\beta)\}. \quad (16)$$

We have the following corollary.

Corollary 1: Under the assumptions in Theorem 5, any $(\mathbf{x}, \beta) \in \mathcal{D}$ that meets $\|\mathbf{x}\|_1 \leq \|\mathbf{x}^\circ\|_1$ satisfies that

$$\begin{aligned} \|\mathbf{x} - \mathbf{x}^\circ\|_2 &\leq (C_0 k^{-1/2} + C_1) \|\mathbf{x}^\circ - \mathbf{x}^k\|_1 + C_2 \epsilon, \\ \|(\beta - \beta^\circ) \odot \mathbf{x}^k\|_2 &\leq (C_0 k^{-1/2} + C_1) \|\mathbf{x}^\circ - \mathbf{x}^k\|_1 + C_2 \epsilon \end{aligned}$$

with $C_i, C_i, i = 0, 1, 2$, as defined in Theorem 5.

Corollary 1 generalizes Theorem 5 and its proof follows directly that of Theorem 5. It shows that a good recovery in SP-CS is not necessarily an optimal solution to P-BPDN. A similar result holds in the standard CS that generalizes Theorem 1, and the proof of Theorem 1 in [4] applies directly to such case.

Corollary 2: Under the assumptions in Theorem 1, any \mathbf{x} that meets $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \epsilon$ and $\|\mathbf{x}\|_1 \leq \|\mathbf{x}^\circ\|_1$ satisfies that

$$\|\mathbf{x} - \mathbf{x}^\circ\|_2 \leq C_0^{std} k^{-1/2} \|\mathbf{x}^\circ - \mathbf{x}^k\|_1 + C_1^{std} \epsilon \quad (17)$$

with C_0^{std}, C_1^{std} as defined in Theorem 1.

An illustration of Corollary 2 is presented in Fig. 1, where the shaded band area refers to the feasible domain of BPDN in (4) and all points in the triangular area, intersection of the feasible domain of BPDN and the ℓ_1 ball $\{\mathbf{x} : \|\mathbf{x}\|_1 \leq \|\mathbf{x}^\circ\|_1\}$, are good candidates for recovery of \mathbf{x}° . The reason why

one seeks for the optimal solution \mathbf{x}^* is to guarantee that the inequality $\|\mathbf{x}\|_1 \leq \|\mathbf{x}^o\|_1$ holds while $\|\mathbf{x}^o\|_1$ is generally unavailable *a priori*. Corollary 2 can explain why a satisfactory recovery can be obtained in practice using some algorithm that may not produce an optimal solution to BPDN, e.g., rONE-L1 [29]. Corollaries 1 and 2 are useful for checking the effectiveness of an algorithm in the case when the output cannot be guaranteed to be optimal.³ Namely, an algorithm is called effective in solving some ℓ_1 minimization problem if it can produce a feasible solution with its ℓ_1 norm no larger than that of the original signal. Similar ideas have been adopted in [31], [32].

IV. ALGORITHMS FOR P-BPDN

A. Special Case: Positive Signals

In general, the P-BPDN problem in (10) for SP-CS is nonconvex and an optimal solution cannot be easily achieved. This subsection studies a special case where the original signal \mathbf{x}^o is positive-valued (except zero entries). Such a case has been studied in the standard CS [34], [35]. By incorporating the positiveness of \mathbf{x}^o , P-BPDN is modified into the positive P-BPDN (PP-BPDN) problem

$$\min_{\mathbf{x}, \beta} \|\mathbf{x}\|_1, \text{ subject to } \begin{cases} \|\mathbf{y} - (\mathbf{A} + \mathbf{B}\Delta)\mathbf{x}\|_2 \leq \epsilon, \\ \mathbf{x} \succcurlyeq \mathbf{0}, \\ r\mathbf{I} \succcurlyeq \beta \succcurlyeq -r\mathbf{I}, \end{cases}$$

where \succcurlyeq is \geq with an elementwise operation and $\mathbf{0}$, \mathbf{I} are column vectors composed of 0, 1 respectively with proper dimensions. It is noted that the robustly stable signal recovery results in the present paper apply directly to the solution to PP-BPDN in such case. This subsection shows that the nonconvex PP-BPDN problem can be transformed into a convex one and hence its optimal solution can be efficiently obtained. Denote $\mathbf{p} = \beta \odot \mathbf{x}$. A new, convex problem (P_1) is introduced as follows.

$$(P_1) \quad \min_{\mathbf{x}, \mathbf{p}} \|\mathbf{x}\|_1, \text{ subject to } \begin{cases} \left\| \mathbf{y} - \Psi \begin{bmatrix} \mathbf{x} \\ \mathbf{p} \end{bmatrix} \right\|_2 \leq \epsilon, \\ \mathbf{x} \succcurlyeq \mathbf{0}, \\ r\mathbf{x} \succcurlyeq \mathbf{p} \succcurlyeq -r\mathbf{x}. \end{cases}$$

Theorem 6: Problems PP-BPDN and (P_1) are equivalent in the sense that, if (\mathbf{x}^*, β^*) is an optimal solution to PP-BPDN, then there exists $\mathbf{p}^* = \beta^* \odot \mathbf{x}^*$ such that $(\mathbf{x}^*, \mathbf{p}^*)$ is an optimal solution to (P_1), and that, if $(\mathbf{x}^*, \mathbf{p}^*)$ is an optimal solution to (P_1), then there exists β^* with $\beta_i^* = \begin{cases} p_i^*/x_i^*, & \text{if } x_i^* > 0; \\ 0, & \text{otherwise} \end{cases}$ such that (\mathbf{x}^*, β^*) is an optimal solution to PP-BPDN.

Proof: We only prove the first part of Theorem 6 using contradiction. The second part follows similarly. Suppose that $(\mathbf{x}^*, \mathbf{p}^*)$ with $\mathbf{p}^* = \beta^* \odot \mathbf{x}^*$ is not an optimal solution to (P_1). Then there exists $(\mathbf{x}', \mathbf{p}')$ in the feasible domain of (P_1) such that $\|\mathbf{x}'\|_1 < \|\mathbf{x}^*\|_1$. Define β' as $\beta'_i = \begin{cases} p'_i/x'_i, & \text{if } x'_i > 0; \\ 0, & \text{otherwise} \end{cases}$. It is easy to show that (\mathbf{x}', β') is a

feasible solution to PP-BPDN. By $\|\mathbf{x}'\|_1 < \|\mathbf{x}^*\|_1$ we conclude that (\mathbf{x}^*, β^*) is not an optimal solution to PP-BPDN, which leads to contradiction. ■

Theorem 6 states that an optimal solution to PP-BPDN can be efficiently obtained by solving the convex problem (P_1).

B. AA-P-BPDN: Alternating Algorithm for P-BPDN

This subsection studies P-BPDN in (10) for general signals where P-BPDN is nonconvex and the global minimum cannot be easily obtained. A simple method is to solve a series of BPDN problems with

$$\mathbf{x}^{(j+1)} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1, \text{ subject to } \left\| \mathbf{y} - (\mathbf{A} + \mathbf{B}\Delta^{(j)})\mathbf{x} \right\|_2 \leq \epsilon, \quad (18)$$

$$\beta^{(j+1)} = \arg \min_{\beta \in [-r, r]^n} \left\| \mathbf{y} - (\mathbf{A} + \mathbf{B}\Delta)\mathbf{x}^{(j+1)} \right\|_2 \quad (19)$$

starting from $\beta^{(0)} = \mathbf{0}$, where the superscript (j) indicates the j th iteration and $\Delta^{(j)} = \text{diag}(\beta^{(j)})$. Denote AA-P-BPDN the alternating algorithm defined by (18) and (19). To analyze AA-P-BPDN, we first present the following two lemmas.

Lemma 1: For a matrix sequence $\{\Phi^{(j)}\}_{j=1}^{\infty}$ composed of fat matrices, let $\mathcal{D}^j = \{\mathbf{v} : \|\mathbf{y} - \Phi^{(j)}\mathbf{v}\|_2 \leq \epsilon\}$, $j = 1, 2, \dots$, and $\mathcal{D}^* = \{\mathbf{v} : \|\mathbf{y} - \Phi^*\mathbf{v}\|_2 \leq \epsilon\}$ with $\epsilon > 0$. If $\Phi^{(j)} \rightarrow \Phi^*$, as $j \rightarrow +\infty$, then for any $\mathbf{v} \in \mathcal{D}^*$ there exists a sequence $\{\mathbf{v}^{(j)}\}_{j=1}^{\infty}$ with $\mathbf{v}^{(j)} \in \mathcal{D}^{(j)}$, $j = 1, 2, \dots$, such that $\mathbf{v}^{(j)} \rightarrow \mathbf{v}$, as $j \rightarrow +\infty$.

Lemma 1 studies the variation of feasible domains \mathcal{D}^j , $j = 1, 2, \dots$, of a series of BPDN problems whose sensing matrices $\Phi^{(j)}$, $j = 1, 2, \dots$, converge to Φ^* . It states that the sequence of the feasible domains also converges to \mathcal{D}^* in the sense that for any point in \mathcal{D}^* , there exists a sequence of points, each of which belongs to one \mathcal{D}^j , that converges to the point. To prove Lemma 1, we first show that it holds for any interior point of \mathcal{D}^* by constructing such a sequence. Then we show that it also holds for a boundary point of \mathcal{D}^* by that for any boundary point there exists a sequence of interior points of \mathcal{D}^* that converges to it. The detailed proof is given in Appendix C.

Lemma 2: An optimal solution \mathbf{x}^* to the BPDN problem in (4) satisfies that $\mathbf{x}^* = \mathbf{0}$, if $\|\mathbf{y}\|_2 \leq \epsilon$, or $\|\mathbf{y} - \Phi\mathbf{x}^*\|_2 = \epsilon$, otherwise.

Proof: It is trivial for the case where $\|\mathbf{y}\|_2 \leq \epsilon$. Consider the other case where $\|\mathbf{y}\|_2 > \epsilon$. Note first that $\mathbf{x}^* \neq \mathbf{0}$. We use contradiction to show that the equality $\|\mathbf{y} - \Phi\mathbf{x}^*\|_2 = \epsilon$ holds. Suppose that $\|\mathbf{y} - \Phi\mathbf{x}^*\|_2 < \epsilon$. Introduce $f(\theta) = \|\mathbf{y} - \theta\Phi\mathbf{x}^*\|_2$. Then $f(0) > \epsilon$, and $f(1) < \epsilon$. There exists θ_0 , $0 < \theta_0 < 1$, such that $f(\theta_0) = \epsilon$ since $f(\theta)$ is continuous on the interval $[0, 1]$. Hence, $\mathbf{x}' = \theta_0\mathbf{x}^*$ is a feasible solution to BPDN in (4). We conclude that \mathbf{x}^* is not optimal by $\|\mathbf{x}'\|_1 = \theta_0\|\mathbf{x}^*\|_1 < \|\mathbf{x}^*\|_1$, which leads to contradiction. ■

Lemma 2 studies the location of an optimal solution to the BPDN problem. It states that the optimal solution locates at the origin if the origin is a feasible solution, or at the boundary of the feasible domain otherwise. This can be easily observed from Fig. 1. Based on Lemmas 1 and 2, we have the following results for AA-P-BPDN.

³It is common when the problem to be solved is nonconvex, such as P-BPDN as discussed in Section IV and ℓ_p ($0 \leq p < 1$) minimization approaches [20], [21], [30] in the standard CS. In addition, Corollaries 1 and 2 can be readily extended to the ℓ_p ($0 \leq p < 1$) minimization approaches.

Theorem 7: Any accumulation point (\mathbf{x}^*, β^*) of the sequence $\{(\mathbf{x}^{(j)}, \beta^{(j)})\}_{j=1}^{\infty}$ is a stationary point of AA-P-BPDN in the sense that

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1, \text{ subject to} \\ \|\mathbf{y} - (\mathbf{A} + \mathbf{B}\Delta^*)\mathbf{x}\|_2 \leq \epsilon, \quad (20)$$

$$\beta^* = \arg \min_{\beta \in [-r, r]^n} \|\mathbf{y} - (\mathbf{A} + \mathbf{B}\Delta)\mathbf{x}^*\|_2 \quad (21)$$

with $\Delta^* = \text{diag}(\beta^*)$.

Theorem 8: An optimal solution (\mathbf{x}^*, β^*) to P-BPDN in (10) is a stationary point of AA-P-BPDN.

Theorem 7 studies the property of the solution $(\mathbf{x}^{(j)}, \beta^{(j)})$ produced by AA-P-BPDN. It shows that $(\mathbf{x}^{(j)}, \beta^{(j)})$ is arbitrarily close to a stationary point of AA-P-BPDN as the iteration index j is large enough.⁴ Hence, the output of AA-P-BPDN can be considered as a stationary point provided that an appropriate termination criterion is set. Theorem 8 tells that an optimal solution to P-BPDN is a stationary point of AA-P-BPDN. So, it is possible for AA-P-BPDN to produce an optimal solution to P-BPDN. The proofs of Theorems 7 and 8 are provided in Appendix D and Appendix E respectively.

C. Effectiveness of AA-P-BPDN

As reported in the last subsection, it is possible for AA-P-BPDN to produce an optimal solution to P-BPDN. But it is not easy to check the optimality of the output of AA-P-BPDN because of the nonconvexity of P-BPDN. Instead, we study the effectiveness of AA-P-BPDN in solving P-BPDN in this subsection with the concept of effectiveness as defined in Subsection III-E. By Corollary 1, a good signal recovery $\hat{\mathbf{x}}$ of \mathbf{x}^o is not necessarily an optimal solution. It requires only that $(\hat{\mathbf{x}}, \hat{\beta})$, where $\hat{\beta}$ denotes the recovery of β^o , be a feasible solution to P-BPDN and that $\|\hat{\mathbf{x}}\|_1 \leq \|\mathbf{x}^o\|_1$ holds. As shown in the proof of Theorem 7 in Appendix D, that $(\mathbf{x}^{(j)}, \beta^{(j)})$ for any $j \geq 1$ is a feasible solution to P-BPDN and that the sequence $\{\|\mathbf{x}^{(j)}\|_1\}_{j=1}^{\infty}$ is monotone decreasing and converges. So, the effectiveness of AA-P-BPDN in solving P-BPDN can be assessed via numerical simulations by checking whether $\|\mathbf{x}^{AA}\|_1 \leq \|\mathbf{x}^o\|_1$ holds with \mathbf{x}^{AA} denoting the output of AA-P-BPDN. The effectiveness of AA-P-BPDN is verified in Subsection V via numerical simulations, where we observe that the inequality $\|\mathbf{x}^{AA}\|_1 \leq \|\mathbf{x}^o\|_1$ holds in all experiments (over 3700 trials).

V. NUMERICAL SIMULATIONS

This section demonstrates the robustly stable signal recovery results of SP-CS in the present paper, as well as the effectiveness of AA-P-BPDN in solving P-BPDN in (10), via numerical simulations. AA-P-BPDN is implemented in Matlab with problems in (18) and (19) being solved using CVX [33]. AA-P-BPDN is terminated as $\frac{\|\|\mathbf{x}^{(j)}\|_1 - \|\mathbf{x}^{(j-1)}\|_1\|}{\|\mathbf{x}^{(j-1)}\|_1} \leq 1 \times 10^{-6}$

⁴It is shown in the proof of Theorem 7 in Appendix D that the sequence $\{(\mathbf{x}^{(j)}, \beta^{(j)})\}_{j=1}^{\infty}$ is bounded. And it can be shown, for example, using contradiction, that for a bounded sequence $\{a_j\}_{j=1}^{\infty}$, there exists an accumulation point of $\{a_j\}_{j=1}^{\infty}$ such that a_j is arbitrarily close to it as j is large enough.

or the maximum number of iterations, set to 200, is reached. PP-BPDN is also implemented in Matlab and solved by CVX.

We first consider general signals. The sparse signal case is mainly studied. The variation of the signal recovery error is studied with respect to the noise level, perturbation level and number of measurements respectively. Besides AA-P-BPDN for P-BPDN in SP-CS, performances of three other approaches are also studied. The first one assumes that the perturbation is known *a priori* and recovers the original signal \mathbf{x}^o by solving, namely, the oracle (O-) BPDN problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1, \text{ subject to } \|\mathbf{y} - (\mathbf{A} + \mathbf{B}\Delta^o)\mathbf{x}\|_2 \leq \epsilon.$$

The O-BPDN approach produces the best recovery result of SP-CS within the scope of ℓ_1 minimization of CS since it exploits the exact perturbation (oracle information). The second one corresponds to the robust signal recovery of the perturbed CS as described in Subsection II-C and solves N-BPDN in (6) where $\epsilon_{E, \mathbf{x}^o} = \|\mathbf{B}\Delta^o \mathbf{x}^o\|_2$ is used though it is not available in practice. The last one refers to the other approach to SP-CS that seeks for the signal recovery by solving TPS-BPDN in (15) as discussed in Subsection III-E.

The first experiment studies the signal recovery error with respect to the noise level. We set the signal length $n = 200$, sample size $m = 80$, sparsity level $k = 10$ and perturbation parameter $r = 0.1$. The noise level ϵ varies from 0.05 to 2 with interval 0.05. For each combination of (n, m, k, r, ϵ) , the signal recovery error, as well as β^o recovery error (on the support of \mathbf{x}^o), is averaged over $R = 50$ trials. In each trial, matrices \mathbf{A} and \mathbf{B} are generated from Gaussian distribution and each column of them has zero mean and unit norm after proper scaling. The sparse signal \mathbf{x}^o is composed of unit spikes with random signs and locations. Entries of β^o are uniformly distributed in $[-r, r]$. The noise \mathbf{e} is zero mean Gaussian distributed and then scaled such that $\|\mathbf{e}\|_2 = \epsilon$. Using the same data, the four approaches, including O-BPDN, N-BPDN, TPS-BPDN and AA-P-BPDN for P-BPDN, are used to recover \mathbf{x}^o respectively in each trial. The simulation results are shown in Fig. 2. It can be seen that both signal and β^o recovery errors of AA-P-BPDN for P-BPDN in SP-CS are proportional to the noise, which is consistent with our robustly stable signal recovery result in the present paper. The error of N-BPDN grows linearly with the noise but a large error still exhibits in the noise free case. Except the ideal case of O-BPDN, our proposed P-BPDN has the smallest error.

The second experiment studies the effect of the structured perturbation. Experiment settings are the same as those in the first experiment except that we set $(n, m, k, \epsilon) = (200, 80, 10, 0.5)$ and vary $r \in \{0.05, 0.1, \dots, 1\}$. Fig. 3 presents our simulation results. A nearly constant error is obtained using O-BPDN in the standard CS since the perturbation is assumed to be known in O-BPDN. The error of AA-P-BPDN for P-BPDN in SP-CS slowly increases with the perturbation level and is quite close to that of O-BPDN for a moderate perturbation. Such a behavior is consistent with our analysis. Besides, it can be observed that the error of N-BPDN grows linearly with the perturbation level. Again, our proposed P-BPDN has the smallest error except O-BPDN.

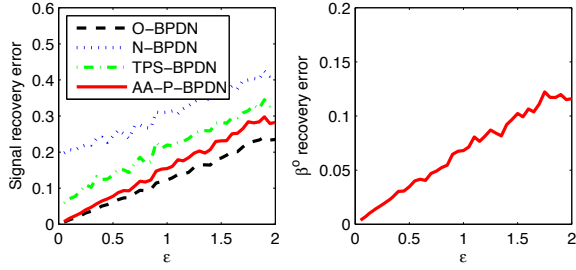


Fig. 2. Signal and perturbation recovery errors with respect to the noise level ϵ with parameter settings $(n, m, k, r) = (200, 80, 10, 0.1)$. Both signal and β^o recovery errors of AA-P-BPDN for P-BPDN in SP-CS are proportional to ϵ .

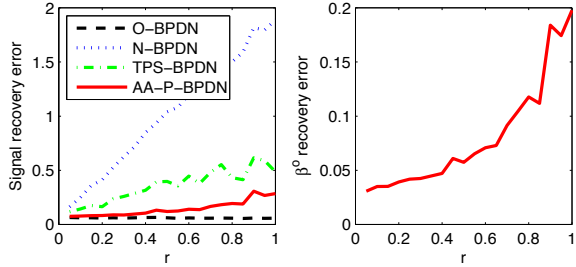


Fig. 3. Signal and perturbation recovery errors with respect to the perturbation level in terms of r with parameter settings $(n, m, k, \epsilon) = (200, 80, 10, 0.5)$. The error of AA-P-BPDN for P-BPDN in SP-CS slowly increases with the perturbation level and is quite close to that of the ideal case of O-BPDN for a moderate perturbation.

The third experiment studies the variation of the recovery error with the number of measurements. We set $(n, k, r, \epsilon) = (200, 10, 0.1, 0.2)$ and vary $m \in \{30, 35, \dots, 100\}$. Simulation results are presented in Fig. 4. Signal recovery errors of all four approaches decrease as the number of measurements increases. Again, it is observed that O-BPDN of the ideal case achieves the best result followed by our proposed P-BPDN. For example, to obtain the signal recovery error of 0.05, about 55 measurements are needed for O-BPDN while the numbers are, respectively, 65 for AA-P-BPDN and 95 for TPS-BPDN. It is impossible for N-BPDN to achieve such a small error in our observation because of the existence of the perturbation.

Another experiment result is shown in Fig. 5 for a compressible signal that is generated by taking a fixed sequence

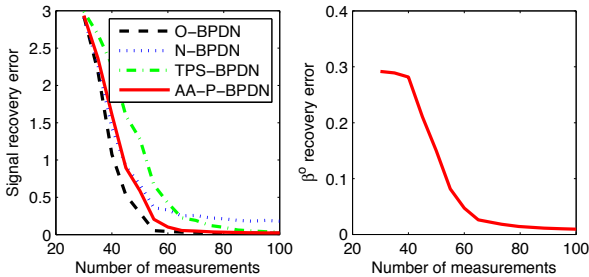


Fig. 4. Signal and perturbation recovery errors with respect to the number of measurements with parameter settings $(n, k, r, \epsilon) = (200, 10, 0.1, 0.2)$. AA-P-BPDN for P-BPDN in SP-CS has the best performance except the ideal case of O-BPDN.

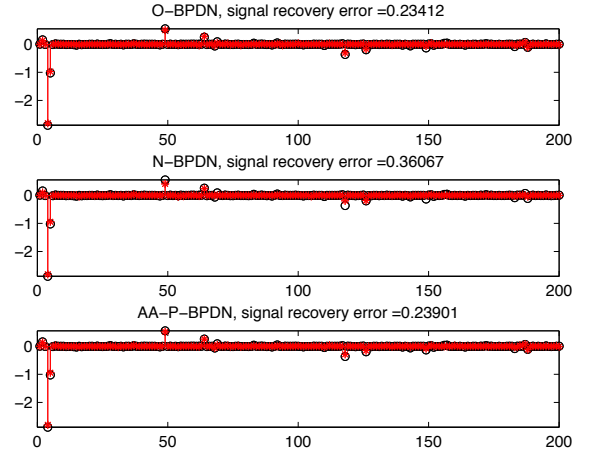


Fig. 5. Recovery of a compressible signal from noisy measurements with $(m, n, r, \epsilon) = (200, 70, 0.1, 0.2)$. The performance of AA-P-BPDN for P-BPDN in SP-CS is very close to that of the ideal O-BPDN. Black circles: original signal; red stars: recoveries.

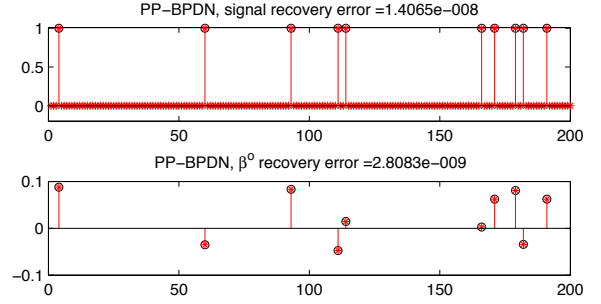


Fig. 6. Exact recovery of a positive sparse signal from noise-free measurements with $(m, n, k, r, \epsilon) = (200, 50, 10, 0.1, 0)$. PP-BPDN is solved by solving (P_1) . β^o and its recovery are shown only on the support of \mathbf{x}^o . Black circles: original signal and β^o ; red stars: recoveries.

$\{2.8843 \cdot i^{-1.5}\}_{i=1}^n$ with $n = 200$, randomly permuting it, and multiplying by a random sign sequence (the coefficient 2.8843 is chosen such that the compressible signal has the same ℓ_2 norm as the sparse signals in the previous experiments). It is sought to be recovered from $m = 70$ noisy measurements with $\epsilon = 0.2$ and $r = 0.1$. The signal recovery error of AA-P-BPDN for P-BPDN in SP-CS is about 0.239, while errors of O-BPDN, N-BPDN and TPS-BPDN are about 0.234, 0.361 and 0.314 respectively (the recovery of TPS-BPDN is omitted in Fig. 5).

For the special positive signal case, an optimal solution to PP-BPDN can be efficiently obtained. An experiment result is shown in Fig. 6, where a sparse signal of length $n = 200$, composed of $k = 10$ positive unit spikes, is exactly recovered from $m = 50$ noise free measurements with $r = 0.1$ by solving (P_1) .

VI. CONCLUSION

This paper studied the CS problem in the presence of measurement noise and a structured matrix perturbation. A concept named as robust stability for signal recovery was broached. It

was shown that the robust stability can be achieved for a sparse signal by solving an ℓ_1 minimization problem P-BPDN under mild conditions. In the presence of measurement noise, the recovery error is at most proportional to the noise level and the recovery is exact in the special noise free case. A general result for compressible signals was also reported. An alternating algorithm named as AA-P-BPDN was proposed to solve the nonconvex P-BPDN problem, and numerical simulations were carried out, verifying our theoretical analysis. A special positive signal case was also studied and a computationally efficient approach was proposed to find an optimal solution to the nonconvex P-BPDN problem.

In this paper it was shown that the D-RIP condition for guaranteed robust stability in SP-CS can be satisfied for mutually independent random perturbation and nominal sensing matrix. One future work is to study practical situations where the perturbation and the nominal sensing matrix may be correlated.

APPENDIX A PROOF OF THEOREM 3

Denote $\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \beta \odot \mathbf{x} \end{bmatrix}$ and similarly define \mathbf{z}^o and \mathbf{z}^* . Then the problem in (9) can be rewritten into

$$\min_{\mathbf{x} \in \mathbb{R}^n, \beta \in [-r, r]^n} \|\mathbf{x}\|_0, \quad \text{subject to } \mathbf{y} = \Psi \mathbf{z}. \quad (22)$$

Let $\bar{\delta}_k = \bar{\delta}_k(\Psi)$ hereafter for brevity.

First note that \mathbf{x}^o is k -sparse and \mathbf{z}^o is $2k$ -D-sparse. Since (\mathbf{x}^*, β^*) is a solution to the problem in (22), we have $\|\mathbf{x}^*\|_0 \leq \|\mathbf{x}^o\|_0 \leq k$ and, hence, \mathbf{z}^* is $2k$ -D-sparse. By $\mathbf{y} = \Psi \mathbf{z}^o = \Psi \mathbf{z}^*$ we obtain $\Psi(\mathbf{z}^o - \mathbf{z}^*) = \mathbf{0}$ and thus $\mathbf{z}^o - \mathbf{z}^* = \mathbf{0}$ by $\bar{\delta}_{4k} < 1$ and the fact that $\mathbf{z}^o - \mathbf{z}^*$ is $4k$ -D-sparse. We complete the proof by observing that $\mathbf{z}^o - \mathbf{z}^* = \begin{bmatrix} \mathbf{x}^o - \mathbf{x}^* \\ \beta^o \odot \mathbf{x}^o - \beta^* \odot \mathbf{x}^* \end{bmatrix} = \mathbf{0}$.

APPENDIX B PROOFS OF THEOREMS 4 AND 5

We only present the proof of Theorem 5 since Theorem 4 is a special case of Theorem 5. We first show the following lemma.

Lemma 3: We have

$$|\langle \Psi \mathbf{v}, \Psi \mathbf{v}' \rangle| \leq \bar{\delta}_{2(k+k')} \|\mathbf{v}\|_2 \|\mathbf{v}'\|_2$$

for all $2k$ -D-sparse \mathbf{v} and $2k'$ -D-sparse \mathbf{v}' supported on disjoint subsets.

Proof: Without loss of generality, assume that \mathbf{v} and \mathbf{v}' are unit vectors with disjoint supports as above. Then by the definition of D-RIP and $\|\mathbf{v} \pm \mathbf{v}'\|_2^2 = \|\mathbf{v}\|_2^2 + \|\mathbf{v}'\|_2^2 = 2$ we have

$$2(1 - \bar{\delta}_{2(k+k')}) \leq \|\Psi \mathbf{v} \pm \Psi \mathbf{v}'\|_2^2 \leq 2(1 + \bar{\delta}_{2(k+k')}).$$

And thus

$$|\langle \Psi \mathbf{v}, \Psi \mathbf{v}' \rangle| \leq \frac{1}{4} \left| \|\Psi \mathbf{v} + \Psi \mathbf{v}'\|_2^2 - \|\Psi \mathbf{v} - \Psi \mathbf{v}'\|_2^2 \right| \leq \bar{\delta}_{2(k+k')},$$

which completes the proof. \blacksquare

Using the notations \mathbf{z} , \mathbf{z}^o , \mathbf{z}^* and $\bar{\delta}_k$ in Appendix A, P-BPDN in (10) can be rewritten into

$$\min_{\mathbf{x} \in \mathbb{R}^n, \beta \in [-r, r]^n} \|\mathbf{x}\|_1, \quad \text{subject to } \|\mathbf{y} - \Psi \mathbf{z}\|_2 \leq \epsilon. \quad (23)$$

Let $\mathbf{h} = \mathbf{x}^* - \mathbf{x}^o$ and decompose \mathbf{h} into a sum of k -sparse vectors $\mathbf{h}_{T_0}, \mathbf{h}_{T_1}, \mathbf{h}_{T_2}, \dots$, where T_0 denotes the set of indices of the k largest entries (in absolute value) of \mathbf{x}^o , T_1 that of the k largest entries of $\mathbf{h}_{T_0^c}$ with T_0^c being the complementary set of T_0 , T_2 that of the next k largest entries of $\mathbf{h}_{T_0^c}$ and so on.

We misuse notations $\mathbf{z}_{T_j}^* = \begin{bmatrix} \mathbf{x}_{T_j}^* \\ \beta_{T_j}^* \odot \mathbf{x}_{T_j}^* \end{bmatrix}$, $j = 0, 1, 2, \dots$, and similarly define $\mathbf{z}_{T_j}^o$. Let $\mathbf{f} = \mathbf{z}^* - \mathbf{z}^o$ and $\mathbf{f}_{T_j} = \mathbf{z}_{T_j}^* - \mathbf{z}_{T_j}^o$ for $j = 0, 1, 2, \dots$. For brevity we write $T_{01} = T_0 \cup T_1$. To bound $\|\mathbf{h}\|_2$, in the first step we show that $\|\mathbf{h}_{T_{01}^c}\|_2$ is essentially bounded by $\|\mathbf{h}_{T_{01}}\|_2$, and then in the second step we show that $\|\mathbf{h}_{T_{01}}\|_2$ is sufficiently small.

The first step follows the proof of Theorem 1.3 in [4]. Note that

$$\|\mathbf{h}_{T_j}\|_2 \leq k^{1/2} \|\mathbf{h}_{T_j}\|_\infty \leq k^{-1/2} \|\mathbf{h}_{T_{j-1}}\|_1, \quad j \geq 2, \quad (24)$$

and thus

$$\begin{aligned} \|\mathbf{h}_{T_{01}^c}\|_2 &= \left\| \sum_{j \geq 2} \mathbf{h}_{T_j} \right\|_2 \leq \sum_{j \geq 2} \|\mathbf{h}_{T_j}\|_2 \\ &\leq k^{-1/2} \sum_{j \geq 1} \|\mathbf{h}_{T_j}\|_1 \leq k^{-1/2} \|\mathbf{h}_{T_0^c}\|_1. \end{aligned} \quad (25)$$

By $\mathbf{x}^* = \mathbf{x}^o + \mathbf{h}$ is an optimal solution, we have

$$\begin{aligned} \|\mathbf{x}^o\|_1 &\geq \|\mathbf{x}^o + \mathbf{h}\|_1 = \sum_{i \in T_0} |x_i^o + h_i| + \sum_{i \in T_0^c} |x_i^o + h_i| \\ &\geq \|\mathbf{x}_{T_0}^o\|_1 - \|\mathbf{h}_{T_0}\|_1 + \|\mathbf{h}_{T_0^c}\|_1 - \|\mathbf{x}_{T_0^c}^o\|_1 \end{aligned} \quad (26)$$

and thus

$$\|\mathbf{h}_{T_0^c}\|_1 \leq \|\mathbf{h}_{T_0}\|_1 + 2 \|\mathbf{x}_{T_0^c}^o\|_1. \quad (27)$$

By (25), (27) and the inequality $\|\mathbf{h}_{T_0}\|_1 \leq k^{1/2} \|\mathbf{h}_{T_0}\|_2$ we have

$$\|\mathbf{h}_{T_{01}^c}\|_2 \leq \sum_{j \geq 2} \|\mathbf{h}_{T_j}\|_2 \leq \|\mathbf{h}_{T_0}\|_2 + 2k^{-1/2} e_0 \quad (28)$$

with $e_0 \equiv \|\mathbf{x}^o - \mathbf{x}^k\|_1$.

In the second step, we bound $\|\mathbf{h}_{T_{01}}\|_2$ by utilizing its relationship with $\|\mathbf{f}_{T_{01}}\|_2$. Note that \mathbf{f}_{T_j} for each $j = 0, 1, \dots$ is $2k$ -D-sparse. By $\Psi \mathbf{f}_{T_{01}} = \Psi \mathbf{f} - \sum_{j \geq 2} \Psi \mathbf{f}_{T_j}$ we have

$$\begin{aligned} \|\Psi \mathbf{f}_{T_{01}}\|_2^2 &= \langle \Psi \mathbf{f}_{T_{01}}, \Psi \mathbf{f} \rangle - \sum_{j \geq 2} \langle \Psi \mathbf{f}_{T_{01}}, \Psi \mathbf{f}_{T_j} \rangle \\ &\leq |\langle \Psi \mathbf{f}_{T_{01}}, \Psi \mathbf{f} \rangle| + \sum_{j \geq 2} \left| \langle \Psi \mathbf{f}_{T_0}, \Psi \mathbf{f}_{T_j} \rangle \right| \\ &\quad + \sum_{j \geq 2} \left| \langle \Psi \mathbf{f}_{T_1}, \Psi \mathbf{f}_{T_j} \rangle \right| \\ &\leq \|\Psi \mathbf{f}_{T_{01}}\|_2 \cdot \|\Psi \mathbf{f}\|_2 + \bar{\delta}_{4k} \|\mathbf{f}_{T_0}\|_2 \sum_{j \geq 2} \|\mathbf{f}_{T_j}\|_2 \\ &\quad + \bar{\delta}_{4k} \|\mathbf{f}_{T_1}\|_2 \sum_{j \geq 2} \|\mathbf{f}_{T_j}\|_2 \end{aligned} \quad (29)$$

$$\leq \|\mathbf{f}_{T_{01}}\|_2 \left(2\epsilon \sqrt{1 + \bar{\delta}_{4k}} + \sqrt{2} \bar{\delta}_{4k} \sum_{j \geq 2} \|\mathbf{f}_{T_j}\|_2 \right). \quad (30)$$

We used Lemma 3 in (29). In (30), we used the D-RIP, and inequalities $\|\mathbf{f}_{T_0}\|_2 + \|\mathbf{f}_{T_1}\|_2 \leq \sqrt{2}\|\mathbf{f}_{T_{01}}\|_2$ and

$$\|\Psi\mathbf{f}\|_2 = \|\Psi(\mathbf{z}^* - \mathbf{z}^o)\|_2 \leq \|\mathbf{y} - \Psi\mathbf{z}^*\|_2 + \|\mathbf{y} - \Psi\mathbf{z}^o\|_2 \leq 2\epsilon. \quad (31)$$

By noting that $\beta^o, \beta^* \in [-r, r]^n$ and

$$\mathbf{f} = \begin{bmatrix} \mathbf{h} \\ \beta^* \odot \mathbf{h} + (\beta^* - \beta^o) \odot \mathbf{x}^o \end{bmatrix} \quad (32)$$

we have

$$\|\mathbf{f}_{T_j}\|_2 \leq \sqrt{1+r^2}\|\mathbf{h}_{T_j}\|_2 + 2r\|\mathbf{x}_{T_j}^o\|_2, \quad j = 0, 1, \dots \quad (33)$$

Meanwhile,

$$\sum_{j \geq 2} \|\mathbf{x}_{T_j}^o\|_2 \leq \sum_{j \geq 2} \|\mathbf{x}_{T_j}^o\|_1 = \|\mathbf{x}_{T_{01}^c}^o\|_1 \leq e_0. \quad (34)$$

Applying the D-RIP, (30), (33) and then (28) and (34) it gives

$$\begin{aligned} (1 - \bar{\delta}_{4k}) \|\mathbf{f}_{T_{01}}\|_2^2 &\leq \|\Psi\mathbf{f}_{T_{01}}\|_2^2 \\ &\leq \|\mathbf{f}_{T_{01}}\|_2 \left\{ 2\sqrt{1 + \bar{\delta}_{4k}}\epsilon + \sqrt{2(1+r^2)\bar{\delta}_{4k}}\|\mathbf{h}_{T_0}\|_2 \right. \\ &\quad \left. + 2\sqrt{2}\bar{\delta}_{4k} \left[\sqrt{1+r^2}k^{-1/2} + r \right] e_0 \right\}, \end{aligned}$$

and thus

$$\begin{aligned} \|\mathbf{h}_{T_{01}}\|_2 &\leq \|\mathbf{f}_{T_{01}}\|_2 \\ &\leq c_1\epsilon + c_0\|\mathbf{h}_{T_{01}}\|_2 + (c_2k^{-1/2} + c_3)e_0 \end{aligned}$$

with $c_0 \equiv \frac{\sqrt{2(1+r^2)\bar{\delta}_{4k}}}{1-\bar{\delta}_{4k}}$, $c_1 \equiv \frac{2\sqrt{1+\bar{\delta}_{4k}}}{1-\bar{\delta}_{4k}}$, $c_2 \equiv 2c_0$ and $c_3 \equiv \frac{2\sqrt{2}\bar{\delta}_{4k}r}{1-\bar{\delta}_{4k}}$. Hence, we get a bound

$$\|\mathbf{h}_{T_{01}}\|_2 \leq (1-c_0)^{-1} \left[c_1\epsilon + (c_2k^{-1/2} + c_3)e_0 \right],$$

which together with (28) gives

$$\begin{aligned} \|\mathbf{h}\|_2 &\leq \|\mathbf{h}_{T_{01}}\|_2 + \|\mathbf{h}_{T_{01}^c}\|_2 \\ &\leq 2\|\mathbf{h}_{T_{01}}\|_2 + 2k^{-1/2}e_0 \\ &\leq \frac{2c_1}{1-c_0}\epsilon + \left[\left(\frac{2c_2}{1-c_0} + 2 \right) k^{-1/2} + \frac{2c_3}{1-c_0} \right] e_0, \end{aligned} \quad (35)$$

which concludes (13).

By (31), (32), (35) and the RIP we have

$$\begin{aligned} &[1 - \delta_k(\mathbf{B})]^{1/2} \|(\beta_{T_0}^* - \beta_{T_0}^o) \odot \mathbf{x}_{T_0}^o\|_2 \\ &\leq \|\mathbf{B}[(\beta_{T_0}^* - \beta_{T_0}^o) \odot \mathbf{x}_{T_0}^o]\|_2 \\ &= \left\| \Psi \begin{bmatrix} \mathbf{0} \\ (\beta_{T_0}^* - \beta_{T_0}^o) \odot \mathbf{x}_{T_0}^o \end{bmatrix} \right\|_2 \\ &= \left\| \Psi \left(\mathbf{f} - \begin{bmatrix} \mathbf{h} \\ \beta^* \odot \mathbf{h} \end{bmatrix} - \begin{bmatrix} \mathbf{0} \\ (\beta_{T_0}^* - \beta_{T_0}^o) \odot \mathbf{x}_{T_0}^o \end{bmatrix} \right) \right\|_2 \\ &\leq \|\Psi\mathbf{f}\|_2 + \left\| \Psi \begin{bmatrix} \mathbf{h} \\ \beta^* \odot \mathbf{h} \end{bmatrix} \right\|_2 + \|\mathbf{B}[(\beta_{T_0}^* - \beta_{T_0}^o) \odot \mathbf{x}_{T_0}^o]\|_2 \\ &\leq 2\epsilon + \sqrt{1+r^2}\|\Psi\|_2\|\mathbf{h}\|_2 + 2r\|\mathbf{B}\|_2e_0 \\ &\leq c_4\epsilon + (c_5k^{-1/2} + c_6)e_0 \end{aligned}$$

with $\delta_k(\mathbf{B}) \leq \delta_{2k}(\mathbf{B}) \leq \bar{\delta}_{4k}$, $c_4 \equiv 2 + \frac{2\sqrt{1+r^2}\|\Psi\|_2c_1}{1-c_0}$, $c_5 \equiv \sqrt{1+r^2} \left(\frac{2c_2}{1-c_0} + 2 \right) \|\Psi\|_2$ and $c_6 \equiv \left(\frac{2\sqrt{1+r^2}c_3}{1-c_0} + 2r \right) \|\Psi\|_2$, and thus

$$\begin{aligned} &\|(\beta_{T_0}^* - \beta_{T_0}^o) \odot \mathbf{x}_{T_0}^o\|_2 \\ &\leq \frac{1}{\sqrt{1-\bar{\delta}_{4k}}} \left[c_4\epsilon + (c_5k^{-1/2} + c_6)e_0 \right], \end{aligned}$$

which concludes (14). We complete the proof by noting that the above results make sense if $c_0 < 1$, i.e.,

$$\bar{\delta}_{4k} < \frac{1}{\sqrt{2(1+r^2)} + 1}.$$

APPENDIX C

PROOF OF LEMMA 1

We first consider the case where \mathbf{v} is an interior point of \mathcal{D}^* , i.e., it holds that $\|\mathbf{y} - \Phi^*\mathbf{v}\|_2 = \epsilon_0 < \epsilon$. Let $\eta = \epsilon - \epsilon_0$. Construct a sequence $\{\mathbf{v}^{(j)}\}_{j=1}^\infty$ such that $\|\mathbf{v}^{(j)} - \mathbf{v}\|_2 \leq 1/j$. It is obvious that $\mathbf{v}^{(j)} \rightarrow \mathbf{v}$. We next show that $\mathbf{v}^{(j)} \in \mathcal{D}^j$ as j is large enough. By $\Phi^{(j)} \rightarrow \Phi^*$, $\mathbf{v}^{(j)} \rightarrow \mathbf{v}$ and that the sequence $\{\mathbf{v}^{(j)}\}_{j=1}^\infty$ is bounded, there exists a positive integer j_0 such that, as $j \geq j_0$,

$$\begin{aligned} \|\Phi^* - \Phi^{(j)}\|_2 \|\mathbf{v}^{(j)}\|_2 &\leq \eta/2 \\ \|\Phi^*\|_2 \|\mathbf{v} - \mathbf{v}^{(j)}\|_2 &\leq \eta/2. \end{aligned}$$

Hence, as $j \geq j_0$,

$$\begin{aligned} &\|\mathbf{y} - \Phi^{(j)}\mathbf{v}^{(j)}\|_2 \\ &= \left\| (\mathbf{y} - \Phi^*\mathbf{v}) + (\Phi^* - \Phi^{(j)})\mathbf{v}^{(j)} + \Phi^*(\mathbf{v} - \mathbf{v}^{(j)}) \right\|_2 \\ &\leq \|\mathbf{y} - \Phi^*\mathbf{v}\|_2 + \left\| (\Phi^* - \Phi^{(j)})\mathbf{v}^{(j)} \right\|_2 + \|\Phi^*(\mathbf{v} - \mathbf{v}^{(j)})\|_2 \\ &\leq \|\mathbf{y} - \Phi^*\mathbf{v}\|_2 + \|\Phi^* - \Phi^{(j)}\|_2 \|\mathbf{v}^{(j)}\|_2 + \|\Phi^*\|_2 \|\mathbf{v} - \mathbf{v}^{(j)}\|_2 \\ &\leq \epsilon_0 + \eta/2 + \eta/2 = \epsilon, \end{aligned}$$

from which we have $\mathbf{v}^{(j)} \in \mathcal{D}^j$ for $j \geq j_0$. By re-selecting arbitrary $\mathbf{v}^{(j)} \in \mathcal{D}^j$ for $j < j_0$ we obtain the conclusion.

For the other case where \mathbf{v} is a boundary point of \mathcal{D}^* , there exists a sequence $\{\mathbf{v}^{(l)}\}_{l=1}^\infty \subset \mathcal{D}^*$ with all $\mathbf{v}^{(l)}$ being interior points of \mathcal{D}^* such that $\mathbf{v}^{(l)} \rightarrow \mathbf{v}$, as $l \rightarrow +\infty$. According to the first part of the proof, for each $l = 1, 2, \dots$, there exists a sequence $\{\mathbf{v}^{(j)}\}_{j=1}^\infty$ with $\mathbf{v}^{(j)} \in \mathcal{D}^j$, $j = 1, 2, \dots$, such that $\mathbf{v}^{(j)} \rightarrow \mathbf{v}^{(l)}$, as $j \rightarrow +\infty$. The sequence $\{\mathbf{v}^{(j)}\}_{j=1}^\infty$ is what we expected since

$$\|\mathbf{v}^{(j)} - \mathbf{v}\|_2 \leq \|\mathbf{v}^{(j)} - \mathbf{v}^{(l)}\|_2 + \|\mathbf{v}^{(l)} - \mathbf{v}\|_2 \rightarrow 0,$$

as $j \rightarrow +\infty$.

APPENDIX D

PROOF OF THEOREM 7

We first show the existence of an accumulation point. It follows the inequality

$$\|\mathbf{y} - (\mathbf{A} + \mathbf{B}\Delta^{(j)})\mathbf{x}^{(j)}\|_2 \leq \|\mathbf{y} - (\mathbf{A} + \mathbf{B}\Delta^{(j-1)})\mathbf{x}^{(j)}\|_2 \leq \epsilon$$

that $\mathbf{x}^{(j)}$ is a feasible solution to the problem in (18), and thus $\|\mathbf{x}^{(j+1)}\|_1 \leq \|\mathbf{x}^{(j)}\|_1$ for $j = 1, 2, \dots$. Then we have $\|\mathbf{x}^{(j)}\|_1 \leq \|\mathbf{x}^{(1)}\|_1 \leq \|\mathbf{A}^\dagger \mathbf{y}\|_1$ for $j = 1, 2, \dots$, since $\mathbf{A}^\dagger \mathbf{y}$ is a feasible solution to the problem in (18) at the first iteration with the superscript \dagger denoting the pseudo-inverse operator. This together with $\beta^{(j)} \in [-r, r]^n$, $j = 1, 2, \dots$, leads that the sequence $\{(\mathbf{x}^{(j)}, \beta^{(j)})\}_{j=1}^\infty$ is bounded. Thus, there exists an accumulation point (\mathbf{x}^*, β^*) of $\{(\mathbf{x}^{(j)}, \beta^{(j)})\}_{j=1}^\infty$.

For the accumulation point (\mathbf{x}^*, β^*) there exists a subsequence $\{(\mathbf{x}^{(j_l)}, \beta^{(j_l)})\}_{l=1}^\infty$ of $\{(\mathbf{x}^{(j)}, \beta^{(j)})\}_{j=1}^\infty$ such that $(\mathbf{x}^{(j_l)}, \beta^{(j_l)}) \rightarrow (\mathbf{x}^*, \beta^*)$, as $l \rightarrow +\infty$. By (19), we have, for all $\beta \in [-r, r]^n$,

$$\left\| \mathbf{y} - (\mathbf{A} + \mathbf{B}\Delta^{(j_l)}) \mathbf{x}^{(j_l)} \right\|_2 \leq \left\| \mathbf{y} - (\mathbf{A} + \mathbf{B}\Delta) \mathbf{x}^{(j_l)} \right\|_2,$$

at both sides of which by taking $l \rightarrow +\infty$, we have, for all $\beta \in [-r, r]^n$,

$$\left\| \mathbf{y} - (\mathbf{A} + \mathbf{B}\Delta^*) \mathbf{x}^* \right\|_2 \leq \left\| \mathbf{y} - (\mathbf{A} + \mathbf{B}\Delta) \mathbf{x}^* \right\|_2,$$

which concludes (21).

For (20), we first point out that $\|\mathbf{x}^{(j)}\|_1 \rightarrow \|\mathbf{x}^*\|_1$, as $j \rightarrow +\infty$, since $\{\|\mathbf{x}^{(j)}\|_1\}_{j=1}^\infty$ is decreasing and \mathbf{x}^* is one of its accumulation points. As in Lemma 1, let $\mathcal{D}^j = \{\mathbf{x} : \|\mathbf{y} - (\mathbf{A} + \mathbf{B}\Delta^{(j)}) \mathbf{x}\|_2 \leq \epsilon\}$ and $\mathcal{D}^* = \{\mathbf{x} : \|\mathbf{y} - (\mathbf{A} + \mathbf{B}\Delta^*) \mathbf{x}\|_2 \leq \epsilon\}$. By $\mathbf{A} + \mathbf{B}\Delta^{(j)} \rightarrow \mathbf{A} + \mathbf{B}\Delta^*$, as $l \rightarrow +\infty$, and Lemma 1, for any $\mathbf{x} \in \mathcal{D}^*$ there exists a sequence $\{\mathbf{x}^{(l)}\}_{l=1}^\infty$ with $\mathbf{x}^{(l)} \in \mathcal{D}^{j_l}$, $l = 1, 2, \dots$, such that $\mathbf{x}^{(l)} \rightarrow \mathbf{x}$, as $l \rightarrow +\infty$. By (18), we have, for $l = 1, 2, \dots$,

$$\left\| \mathbf{x}^{(j_l+1)} \right\|_1 \leq \left\| \mathbf{x}^{(l)} \right\|_1,$$

at both sides of which by taking $l \rightarrow +\infty$, we have

$$\|\mathbf{x}^*\|_1 \leq \|\mathbf{x}\|_1 \quad (36)$$

since $\|\mathbf{x}^{(j)}\|_1 \rightarrow \|\mathbf{x}^*\|_1$, as $j \rightarrow +\infty$, and $\mathbf{x}^{(l)} \rightarrow \mathbf{x}$, as $l \rightarrow +\infty$. Finally, (20) is concluded as (36) holds for arbitrary $\mathbf{x} \in \mathcal{D}^*$.

APPENDIX E PROOF OF THEOREM 8

We need to show that an optimal solution (\mathbf{x}^*, β^*) satisfies (20) and (21). It is obvious for (20). For (21), we discuss two cases based on Lemma 2. If $\|\mathbf{y}\|_2 \leq \epsilon$, then $\mathbf{x}^* = \mathbf{0}$ and, hence, (21) holds for any $\beta^* \in [-r, r]^n$. If $\|\mathbf{y}\|_2 > \epsilon$, $\|\mathbf{y} - (\mathbf{A} + \mathbf{B}\Delta^*) \mathbf{x}^*\|_2 = \epsilon$ holds by (20) and Lemma 2. Next we use contradiction to show that (21) holds in such case.

Suppose that (21) does not hold as $\|\mathbf{y}\|_2 > \epsilon$. That is, there exists $\beta' \in [-r, r]^n$ such that

$$\|\mathbf{y} - (\mathbf{A} + \mathbf{B}\Delta') \mathbf{x}^*\|_2 < \|\mathbf{y} - (\mathbf{A} + \mathbf{B}\Delta^*) \mathbf{x}^*\|_2 = \epsilon$$

holds with $\Delta' = \text{diag}(\beta')$. Then by Lemma 2 we see that \mathbf{x}^* is a feasible but not optimal solution to the problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1, \text{ subject to } \|\mathbf{y} - (\mathbf{A} + \mathbf{B}\Delta') \mathbf{x}\|_2 \leq \epsilon. \quad (37)$$

Hence, $\|\mathbf{x}'\|_1 < \|\mathbf{x}^*\|_1$ holds for an optimal solution \mathbf{x}' to the problem in (37). Meanwhile, (\mathbf{x}', β') is a feasible solution to the P-BPDN problem in (10). Thus (\mathbf{x}^*, β^*) is not an optimal solution to the P-BPDN problem in (10) by $\|\mathbf{x}'\|_1 < \|\mathbf{x}^*\|_1$, which leads to contradiction.

REFERENCES

- [1] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [2] —, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [3] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [4] E. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathématique*, vol. 346, no. 9–10, pp. 589–592, 2008.
- [5] T. Blumensath and M. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, 2009.
- [6] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [7] D. Donoho, Y. Tsaig, I. Drori, and J. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," Available online at <http://www.cs.tau.ac.il/~idrori/StOMP.pdf>, 2006.
- [8] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.
- [9] T. Blumensath and M. Davies, "Compressed sensing and source separation," in *International Conference on Independent Component Analysis and Signal Separation*. Springer, 2007, pp. 341–348.
- [10] T. Xu and W. Wang, "A compressed sensing approach for underdetermined blind audio source separation with sparse representation," in *Statistical Signal Processing, 15th Workshop on*. IEEE, 2009, pp. 493–496.
- [11] J. Zheng and M. Kaveh, "Directions-of-arrival estimation using a sparse spatial spectrum model with uncertainty," in *Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on*. IEEE, 2011, pp. 2848–2551.
- [12] Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse bayesian inference," *Arxiv preprint, available online at <http://arxiv.org/abs/1108.5838>*, 2011.
- [13] M. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2275–2284, 2009.
- [14] L. Zhang, M. Xing, C. Qiu, J. Li, and Z. Bao, "Achieving higher resolution isar imaging with limited pulses via compressed sampling," *Geoscience and Remote Sensing Letters*, vol. 6, no. 3, pp. 567–571, 2009.
- [15] M. Herman and T. Strohmer, "General deviants: an analysis of perturbations in compressed sensing," *IEEE J. Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 342–349, 2010.
- [16] M. Herman and D. Needell, "Mixed operators in compressed sensing," in *Information Sciences and Systems (CISS), 44th Annual Conference on*. IEEE, 2010, pp. 1–6.
- [17] Y. Chi, L. Scharf, A. Pezeshki, and A. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2182–2195, 2011.
- [18] D. Chae, P. Sadeghi, and R. Kennedy, "Effects of basis-mismatch in compressive sampling of continuous sinusoidal signals," in *Future Computer and Communication (ICFCC), 2nd International Conference on*, vol. 2. IEEE, 2010, pp. 739–743.
- [19] H. Zhu, G. Leus, and G. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2002–2016, 2011.
- [20] R. Chartrand and V. Staneva, "Restricted isometry properties and non-convex compressive sensing," *Inverse Problems*, vol. 24, p. 035020, 2008.
- [21] R. Saab, R. Chartrand, and O. Yilmaz, "Stable sparse approximations via nonconvex optimization," in *Acoustics, Speech and Signal Processing, IEEE International Conference on*. IEEE, 2008, pp. 3885–3888.
- [22] E. Candès, "Compressive sampling," in *Proceedings of the International Congress of Mathematicians*, vol. 3. Citeseer, 2006, pp. 1433–1452.
- [23] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.
- [24] S. Foucart and M. Lai, "Sparsest solutions of underdetermined linear systems via lq-minimization for $0 < q_1 < 1$," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 395–407, 2009.

- [25] T. Cai, L. Wang, and G. Xu, "Shifting inequality and recovery of sparse signals," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1300–1308, 2010.
- [26] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 6–18, 2006.
- [27] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.
- [28] T. Blumensath and M. Davies, "Sampling theorems for signals from the union of finite-dimensional linear subspaces," *IEEE Transactions on Information Theory*, vol. 55, no. 4, pp. 1872–1882, 2009.
- [29] Z. Yang, C. Zhang, J. Deng, and W. Lu, "Orthonormal expansion ℓ_1 -minimization algorithms for compressed sensing," *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 6285–6290, 2011.
- [30] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *Signal Processing Letters*, vol. 14, no. 10, pp. 707–710, 2007.
- [31] Z. Yang and C. Zhang, "Sparsity-undersampling tradeoff of compressed sensing in the complex domain," in *Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on*. IEEE, 2011, pp. 3668–3671.
- [32] Z. Yang, C. Zhang, and L. Xie, "On phase transition of compressed sensing in the complex domain," *Signal Processing Letters*, in press, 2011.
- [33] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming," Available online at <http://cvxr.com/cvx>, 2008.
- [34] D. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, 2009.
- [35] M. Stojnic, "Various thresholds for ℓ_1 -optimization in compressed sensing," *Arxiv preprint*, available online at <http://arxiv.org/abs/0907.3666>, 2009.