THEORETICAL STABILITY ANALYSIS OF BASIS PURSUIT DENOISING

[1]SUNG-HSIEN HSIEH, [2]CHUN-SHIEN LU

[1,2]Inst. of Info. Science, Academia Sinica
Email: parvaty316@iis.sinica.edu.tw, Email: lcs@iis.sinica.edu.tw

Abstract - Basis Pursuit Denoising (BPDN) is a popular tool for sparse signal recovery with noisy interference by relaxing Basis Pursuit (BP) with error tolerance. In general, the error tolerance set to be the energy of noise is expected to outperform that set to be 0. However, in fact, under some conditions, the error between BPDN solution and ground truth probably becomes larger along with the increase of error tolerance. In this paper, we want to explore under what conditions, the error tolerance set to be the energy of noise results in the better performance than that set to be 0 under Gaussian random noise. We start this issue from discussing objective values of BPDN in both cases and show that objective values are directly related to performance especially when signal dimension is 2. We leave the case with larger signal dimension for future work. Our analyses are based on an innovative methodology, called conic geometry. Simulations validate our theoretical analyses.

I. INTRODUCTION

A. Background and Related Work

Compressive sensing (CS) [1] has received considerable attention due to its broad applications. At the encoder side, given an observed signal $\boldsymbol{x} \in \mathbb{R}^n$, the corresponding measurement vector can be obtained via sensing matrix $\boldsymbol{\Phi} \in \mathbb{R}^{m\times n}$ $(m < n)$ as: $\boldsymbol{y} = \Phi\boldsymbol{x}$. Nevertheless, the encoding phase is inevitably contaminated by noise, namely $\boldsymbol{y} = \boldsymbol{A}(\boldsymbol{x}+\boldsymbol{\eta}) = A\boldsymbol{x} + \boldsymbol{e}$, with $\boldsymbol{e} = \boldsymbol{A}\boldsymbol{\eta}$ being Gaussian random noise. In [2], [3], the authors show that if $\boldsymbol{x}$ is a k-sparse vector, then Basis Pursuit Denoising (BPDN) [4] is an efficient tool to reconstruct $\boldsymbol{x}$ at the decoder side as:

$$(BPDN) \qquad \min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|_{\ell_1} \quad s.t. \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\theta}\|_{\ell_2} \le \epsilon, \qquad \text{(I.1)}$$

where ϵ is often set to $\|\boldsymbol{e}\|_{\ell_2}$.

Let $\boldsymbol{x}^{P_0}$ and $\boldsymbol{x}^{P_\epsilon}$ be the solutions of problem $(BPDN)$ with $\epsilon = 0$ and $\epsilon = \|\boldsymbol{e}\|_{\ell_2}$, respectively, and let $\boldsymbol{x}^0$ be the ground truth. The existing works focus on deriving the upper bound of reconstruction error, $\|\boldsymbol{x}^0 - \boldsymbol{x}^{P_\epsilon}\|_{\ell_2}$. In CS, error bound is analyzed based on two classes of approaches: Restricted Isometric Property (RIP)-based [2], [5], [6] and Coherence-based [7], [8], [9].

B. Our Motivation

Instead of deriving the error bound of BPDN, we have insight into an interesting but unexplored problem of when there exists an $\epsilon = \|\boldsymbol{e}\|_{\ell_2}$ (see Eq. (I.1)) leading to a better recovery result. Fig. 1 illustrates our findings about the relationship between x^{P_0} and x^{P_ϵ}. In Fig. 1(a), it shows an example that, with a relaxed constraint $\epsilon = \|\boldsymbol{e}\|_{\ell_2}$, BPDN leads to worse results with $\|\boldsymbol{x}^{P_\epsilon}\|_{\ell_1} < \|\boldsymbol{x}^{P_0}\|_{\ell_1} < \|\boldsymbol{x}^0\|_{\ell_1}$. On the other hand, Fig. 1(b) shows an example, where $\boldsymbol{x}^{P_\epsilon}$ outperforms $\boldsymbol{x}^{P_0}$ with $\|\boldsymbol{x}^0\|_{\ell_1} < \|\boldsymbol{x}^{P_\epsilon}\|_{\ell_1} < \|\boldsymbol{x}^{P_0}\|_{\ell_1}$. In sum, if $\|\boldsymbol{x}^{P_0}\|_{\ell_1} \le \|\boldsymbol{x}^0\|_{\ell_1}$, then $\epsilon = \|\boldsymbol{e}\|_{\ell_2}$ will lead to additional error between $\|\boldsymbol{x}^0\|_{\ell_1}$ and $\|\boldsymbol{x}^{P_\epsilon}\|_{\ell_1}$, leading to unstable BPDN since $\boldsymbol{x}^{P_\epsilon}$ never touches $\boldsymbol{x}^0$ whatever ϵ is.

II. STABILITY OF BASIS PURSUIT DENOISING

We describe our theoretical analyses of stability of BPDN with all proofs being omitted here due to limited space.

A. Main Results

As illustrated in Figs. 1(a) and (b), they reveal an imagine that if $\boldsymbol{x}^{P_0}$ (with $\epsilon = 0$) has already fallen within the ℓ_1-ball of $\boldsymbol{x}^0$, BPDN (with $\epsilon > 0$) may yield worse recovery results. To provide rigorous analysis, we relax the ℓ_1-ball of $\boldsymbol{x}^0$ as the descent cone $\mathcal{C}$ for the subsequent discussions, where descent cone is defined in [10]. In addition, we call $\mathcal{N}_A$ is the null space of A, $\bigoplus$ is Minkowski sum.

First, we show the sufficient condition such that BPDN acts as Fig. 1(a) and Fig. 1(b) respectively:

Theorem 2.1: Given an error tolerance $\epsilon > 0$, *if* $\boldsymbol{\eta} \in \mathcal{C}\bigoplus\mathcal{N}_A$, then we have $\|\boldsymbol{x}^{P_\epsilon}\|_{\ell_1} \le \|\boldsymbol{x}^{P_0}\|_{\ell_1} \le \|\boldsymbol{x}^0\|_{\ell_1}$.

Theorem 2.2: If $\boldsymbol{\eta} \notin \mathcal{C}\bigoplus\mathcal{N}_A$, there exists an error tolerance $\epsilon > 0$ such that $\|\boldsymbol{x}^{P_0}\|_{\ell_1} > \|\boldsymbol{x}^{P_\epsilon}\|_{\ell_1} > \|\boldsymbol{x}^0\|_{\ell_1}$.

Definition 2.1: (Stability of BPDN) Given $\boldsymbol{\eta} \sim \mathcal{N}(0, \boldsymbol{I}^n)$ and $\epsilon = \|\boldsymbol{e}\|_{\ell_2}$, the algorithm BPDN is stable if $\|x^0 - x^{P_\epsilon}\|_2 \le \|x^0 - x^{P_0}\|_2$.

Second, the sufficient condition of stable BPDN is equivalent to $\boldsymbol{\eta} \notin \mathcal{C}\bigoplus\mathcal{N}_A$ when $n = 2$. To further calculate the probability of $\mathbb{P}\{\boldsymbol{\eta} \notin \mathcal{C}\bigoplus\mathcal{N}_A\}$, we obtain

Theorem 2.3: Suppose both $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\tilde{\boldsymbol{A}} \in \mathbb{R}^{(m-1)\times n}$ are Gaussian random matrices, and $\boldsymbol{\eta} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}^n)$. Then

$$\begin{aligned}\mathbb{P}\{\mathcal{C}\bigcap\mathcal{N}_{\tilde{A}} = \{\boldsymbol{0}\}\} &\le \mathbb{P}\{\boldsymbol{\eta} \notin \mathcal{C}\bigoplus\mathcal{N}_A\} \\ &\le \mathbb{P}\{\mathcal{C}\bigcap\mathcal{N}_A = \{\boldsymbol{0}\}\}.\end{aligned}$$

Theorem 2.3 is relevant to $\mathcal{C}\bigcap\mathcal{N}_A = \{\boldsymbol{0}\}$, which is, in fact, a necessary and sufficient condition that BP succeeds, defined as follows.

Definition 2.2: [10] (Success of BP) Given n, m, and k, the problem BP in noiseless environments is called success if BP solution is the unique optimal solution.

Fact 2.1: [10] (Optimality condition for problem BP) Given n, m, and k, and consider the problem BP in noiseless environment. The problem BP succeeds if and only if $\mathcal{C}\bigcap\mathcal{N}_A = \{\boldsymbol{0}\}$.

Finally, by combing all the previous results, we can obtain

Theorem 2.4: Notations follow the definition in Theorem 2.3. With $n = 2$, Then

$$\begin{aligned}\mathbb{P}\{\mathcal{C}\bigcap\mathcal{N}_{\tilde{A}} = \{\boldsymbol{0}\}\} &\le \mathbb{P}\{\text{BPDN is stable}\} \\ &\le \mathbb{P}\{\mathcal{C}\bigcap\mathcal{N}_A = \{\boldsymbol{0}\}\}.\end{aligned}$$

For $n > 2$, the sufficient condition is more complex shown in Fig. 3. The cylinder due to the constraint in BPDN may lead to the case that BPDN is not stable even though $\boldsymbol{\eta} \notin \mathcal{C}\bigoplus\mathcal{N}_A$. Taking such a cylinder into account in the theoretical analysis will be a challenge.

III. SIMULATION RESULTS

We validate the results in Theorem 2.3 via simulations by checking whether the probability is bounded between $\mathbb{P}\{\mathcal{C} \bigcap \mathcal{N}_{\tilde{\mathcal{A}}} = \{\mathbf{0}\}\}$ and $\mathbb{P}\{\mathcal{C} \bigcap \mathcal{N}_{\mathcal{A}} = \{\mathbf{0}\}\}$. The result is shown in Fig. 2. We can observe that minor estimation error occurs in the region of phase transition of BP. Consequently, the proposed theoretical bound accurately describes the empirical results.

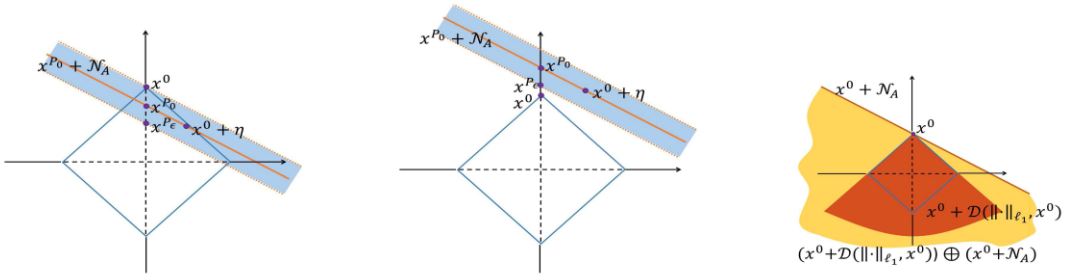


Fig. 1. An illustration of how BPDN works from conic geometry. (a) Due to noise, the set of solutions satisfying $\boldsymbol{y} = \boldsymbol{A}(\boldsymbol{x}+\boldsymbol{\eta})$ belongs to $\boldsymbol{x}^0+\boldsymbol{\eta}+\mathcal{N}_{\boldsymbol{A}}$, where $\boldsymbol{x}^{P_0}$ is the minimal ℓ_1-norm solution. In this case, relaxed constraints in BPDN lead to a larger set of solutions, expressed as the cylinder around $\boldsymbol{x}^0+\boldsymbol{\eta}+\mathcal{N}_{\boldsymbol{A}}$ (in blue). As a result, $\boldsymbol{x}^{P_\epsilon}$ is far from $\boldsymbol{x}^0$ with $\|\boldsymbol{x}^{P_\epsilon}\|_{\ell_1} < \|\boldsymbol{x}^0\|_{\ell_1}$. (b) Another example with $\|\boldsymbol{x}^0\|_{\ell_1} < \|\boldsymbol{x}^{P_\epsilon}\|_{\ell_1}$. (c) The space of Minkowski sum of $\mathcal{D}\left(\|\cdot\|_{\ell_1}, \boldsymbol{x}^0\right)$ and $\mathcal{N}_{\boldsymbol{A}}$, where the ℓ_1-ball of $\boldsymbol{x}^0 \subseteq \mathcal{D}(\|\cdot\|_{\ell_1}, \boldsymbol{x}^0)$ (in orange) and $(\boldsymbol{x}^0+\mathcal{C}) \bigoplus (\boldsymbol{x}^0+\mathcal{N}_{\mathcal{A}})$ (in yellow).

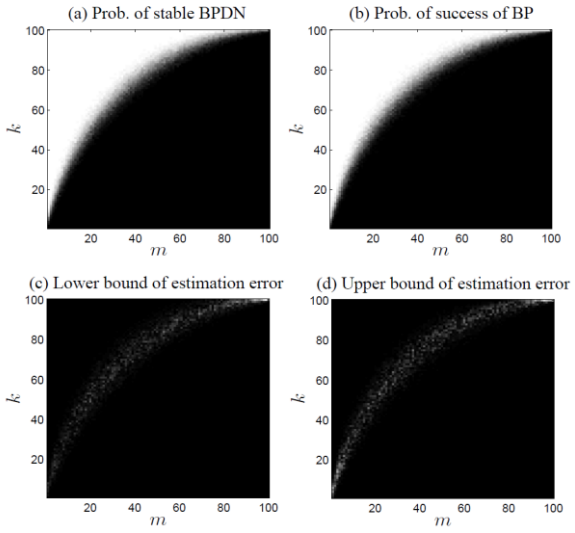


Fig. 2. Experimental results for validating the error between theoretical results and practical results with $n = 100$. (a) Probability of $\boldsymbol{\eta} \notin \mathcal{C} \bigoplus \mathcal{N}_{\mathcal{A}}$ in Theorem 2.3 (white means prob. is 1 and black means prob. is 0). (b) Probability of BP success in Definition 2.2. (c) Lower bound of estimation error. (d) Upper bound of estimation error.

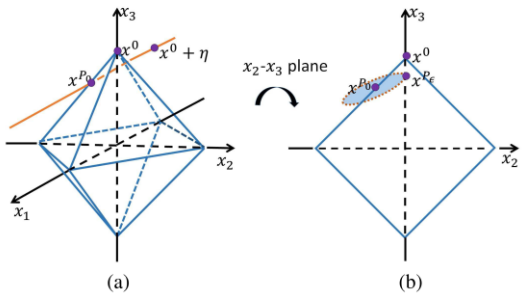


Fig. 3. An example wherein $\|\boldsymbol{x}^{P_0} - \boldsymbol{x}^0\|_{\ell_2} > \|\boldsymbol{x}^{P_\epsilon} - \boldsymbol{x}^0\|_{\ell_2}$. (a) Without error tolerance, $\|\boldsymbol{x}^{P_0}\|_{\ell_1} \leq \|\boldsymbol{x}^0\|_{\ell_1}$, (b) Given $\epsilon > 0$, compared with $\boldsymbol{x}^{P_0}$, $\boldsymbol{x}^{P_\epsilon}$ is close to the ground truth.

REFERENCES

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

[2] E. J. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," Comm. Pure Appl. Math., vol. 59, pp. 1207–1223, 2005.

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

[4] S.S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," SIAM Journal of Scientific Computing, vol. 20, no. 1, pp. 33–61, 1998.

[5] E. J. Candes and T. Tao, "The dantzig selector: Statistical estimation when p is much larger than n," Ann. Statist., vol. 35, no. 6, pp. 2313–2351, 2007.

[6] P. J. Bickel, Y. Ritov, and A. B. Tsybakov, "Simultaneous analysis oflasso and dantzig selector," Ann. Statist., vol. 37, no. 4, pp. 1705–1732,2009.

[7] Z. Ben-Haim, Y. C. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," IEEE Transactions on Signal Processing, vol. 58, pp. 5030–5043, 2010.

[8] J. A. Tropp, "Just relax: convex programming methods for identifying sparse signals in noise," IEEE Transactions on Information Theory, vol. 52, pp. 1030–1051, 2006.

[9] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," IEEE Transactions on Information Theory, vol. 52, pp. 6–18, 2013.

[10] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp, "Living on the edge: Phase transition in convex programs with random data," Information and Inference, vol. 3, pp. 224–294, 2013.

★★★