Received 14 November 2022; revised 14 February 2023; accepted 16 February 2023. Date of publication 30 March 2023; date of current version 11 May 2023. The review of this article was arranged by Associate Editor H. Vicky Zhao.

Digital Object Identifier 10.1109/OJSP.2023.3247301

Dynamic Orthogonal Matching Pursuit for Sparse Data Reconstruction

YUN-BIN ZHAO (Member, IEEE), AND ZHI-QUAN LUO (Fellow, IEEE)

Shenzhen Research Institute of Big Data, Chinese University of Hong Kong, Shenzhen 518172, China

CORRESPONDING AUTHOR: YUN-BIN ZHAO (e-mail yunbinzhao@cuhk.edu.cn).

This work was supported by the National Natural Science Foundation of China (NSFC) under Grants 12071307, 61571384, and 61731018.

ABSTRACT The orthogonal matching pursuit (OMP) is one of the mainstream algorithms for sparse data reconstruction or approximation. It acts as a driving force for the development of several other greedy methods for sparse data reconstruction, and it also plays a vital role in the development of compressed sensing theory for sparse signal and image reconstruction. In this article, we propose the so-called dynamic orthogonal matching pursuit (DOMP) and enhanced dynamic orthogonal matching pursuit (EDOMP) algorithms which are more efficient than OMP for sparse data reconstruction from a numerical point of view. We carry out a rigorous analysis to establish the reconstruction error bound for DOMP under the restricted isometry property of the measurement matrix. The main result claims that the reconstruction error via DOMP can be controlled and measured in terms of the number of iterations, sparsity level of data, and the noise level of measurements. Moveover, the finite convergence of DOMP for a class of large-scale compressed sensing problems is also shown.

INDEX TERMS Data reconstruction, dynamic orthogonal matching pursuit, error bound, orthogonal matching pursuit, restricted isometry property, sparse optimization.

I. INTRODUCTION

Data reconstruction/recovery is a common request in such areas as signal denoising, imaging reconstruction, statistical model selection, pattern recognition, streaming data tracking and wireless channel estimation (see, e.g., [1], [2], [3], [4], [5]). For instance, recovering a sparse signal from limited measurements (observations) is a central task in compressed-sensing-based signal processing [1], [2], [6], [7], [8]. Thus developing fast and robust algorithms for data reconstruction is significantly important in these scenarios. Without loss of generality, let $x \in \mathbb{R}^n$ denote the data (e.g., signal or image) which admits some sparsity structure in the sense that x is either k-sparse (i.e., x has at most $k \ll n$ nonzero components) or k-compressible (i.e., x can be approximated by a k-sparse vector). In this paper, the unknown data $x \in \mathbb{R}^n$ to reconstruct is called the target data. To reconstruct the data $x \in \mathbb{R}^n$, one may first collect a few measurements

$$y_i := (a^i)^T x + \nu_i,\ i = 1, \ldots, m, \tag{1}$$

where $m \ll n$, a^i's are given measurement vectors, and ν_i's denote the measurement errors. Let A be the $m \times n$ matrix whose row vectors consist of $(a^i)^T, i = 1, \ldots, m$, and let $y = (y_1, \ldots, y_m)^T$ and $\nu = (\nu_1, \ldots, \nu_m)^T$ be the vectors of measurements and errors, respectively. Then the system (1) can be written as $y := Ax + \nu$. Under the assumption that the target data x is k-sparse or k-compressible, the reconstruction of $x \in \mathbb{R}^n$ from the measurements y can be formulated as the sparse optimization problem

$$\min_{z \in \mathbb{R}^n} \{\|y - Az\|_2^2 :\ \|z\|_0 \le k\}, \tag{2}$$

where $\|z\|_0$ denotes the number of nonzero entries of z. Essentially, the model (2) is a combinatorial optimization problem. The vectors satisfying the constraint of (2) are k-sparse. In this paper, we assume that $k \ll n$ which is a typical assumption in the scenario of compressed-sensing-based signal and image reconstruction [1], [2], [8], [9]. The algorithms for (2) and similar problems have experienced a significant development over the past decades. The most widely used algorithms are based on convex/nonconvex optimization (e.g., [10], [11], [12], [13], [14], [15]), hard thresholding and greedy methods (e.g., [1], [2], [8], [16]), and the recent technique based on the concept of optimal k-thresholding [17], [18], [19]. The

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

OMP Algorithm. Perform the following steps until a certain stopping criterion is satisfied:
S1 Initialization: $x^{(0)} = 0$ and $S^{(0)} = \emptyset$.
S2 Given $x^{(p)}$ and $S^{(p)}$, let

$$S^{(p+1)} = S^{(p)} \cup L_1(A^T(y - Ax^{(p)})),$$
$$x^{(p+1)} = \arg\min_{z\in\mathbb{R}^n}\{\|y - Az\|_2 : \operatorname{supp}(z) \subseteq S^{(p+1)}\}.$$

orthogonal matching pursuit (OMP) is one of the popular greedy methods for the problem (2). Before recalling such a method, let us first introduce some notations used in the paper.

A. NOTATION

Throughout the paper, all vectors are column vectors unless otherwise specified. The transpose of the vector z and matrix A are denoted by z^T and A^T, respectively. The support of the vector z is denoted by $\operatorname{supp}(z) = \{i : z_i \neq 0\}$. We use $\mathcal{H}_k(\cdot)$ to denote the hard thresholding operator which retains the k largest entries in magnitude and zeros out the other entries of a vector, and we use $L_q(\cdot)$ to denote the index set of the q largest entries in magnitude of a vector. When two entries of z are equal in absolute value, $\mathcal{H}_k(z)$ and $L_q(z)$ might not be uniquely determined, in which case we select the entry with the smallest index. The complement set of $S \subseteq \{1, 2, \ldots, n\}$ with respect to $\{1, \ldots, n\}$ is denoted by $\overline{S} = \{1, 2, \ldots, n\}\backslash S$, and the cardinality of S is denoted by $|S|$. Given $x \in \mathbb{R}^n$ and $S \subseteq \{1, 2, \ldots, n\}$, the vector $x_S \in \mathbb{R}^n$ denotes the vector obtained from x by retaining the entries of x supported on S and setting other entries of x to be zeros. $\|x\|_2 = \sqrt{x^T x}$ and $\|x\|_\infty = \max_{1\le i\le n} |x_i|$ are the ℓ_2-norm and ℓ_∞-norm, respectively.

B. TRADITIONAL OMP ALGORITHM

Referred to as stagewise regression, the OMP algorithm first appeared in the area of statistics several decades ago. It was introduced to signal processing community around 1993 (see, e.g., [20], [21], [22]) and later to approximation theory (e.g., [23], [24], [25]). The OMP algorithm is stated as follows, which iteratively selects a vector basis that is best correlated to the residual at the current iterate.

The advantage of OMP lies in its simple structure allowing a fast implementation at a low computational cost. OMP has stimulated the development of several other compressed sensing algorithms, including the subspace pursuit (SP) [26], compressive sampling matching pursuit (CoSaMP) [27], multipath matching pursuit [28], stagewise orthogonal matching pursuit (StOMP) [29], regularized orthogonal matching pursuit (ROMP) [30], [31], general orthogonal matching pursuit (gOMP) [32], weak orthogonal matching pursuit (WOMP) [33], stagewise weak orthogonal matching pursuit (SWOMP) [34] and constrained matching pursuit (CMP) [35]. In each iteration, the OMP algorithm only select one index corresponding to the largest entry in magnitude of $A^T(y - Az)$, the gradient of the error metric $\|y - Az\|_2^2/2$. Thus the OMP algorithm does not utilize the gradient information efficiently when the gradient vector possesses several large entries in magnitude whose absolute values are close to each other. The gOMP algorithm is a direct generalization of OMP, allowing N indices to be selected simultaneously in every iteration, where $1 \le N < k$ is a prescribed integer number. However, such a selection rule remains very rigid, and it might significantly increase the chance for a wrong index being selected. For instance, if the gradient vector of error matric is τ-compressible, where $\tau < N$, then the N indices selected by gOMP would include an index corresponding to a vanished or insignificant entry of the gradient vector. The SP and CoSaMP in their nature are not close family members of the OMP-type methods. At each step, they iteratively and simultaneously search k indices as the approximation to the support of signal. The StOMP, ROMP, WOMP and SWOMP adopt other different index selection rules to possibly enhance the performance of the standard OMP procedure. A common feature of these methods is that they allow the algorithm to select more than one indices in each iteration. WOMP is remarkably different from OMP in the sense that the index for the largest absolute entry of $A^T(y - Az)$ might be excluded during its iterations. StOMP and SWOMP might select too many indices in a single iteration which admits a high risk for wrong indices to be selected. Recently, the CMP method, as a modification of OMP, was proposed for solving more complicated recovery problems with certain constraints. In this paper, we propose further modifications of OMP, which dynamically select a few indices among the k largest components in magnitude of $A^T(y - Az)$. We also allow the algorithms to adjust the selected index sets at every iteration by performing a shrinkage step when the total number of selected indices goes beyond the sparsity level of the signal (see the description of the proposed algorithm in Section II).

C. EXISTING RESULTS FOR OMP

The theoretical performance of OMP-type methods can be analyzed in terms of coherence tools (e.g., [1], [36], [37], [38], [39]). However, the restricted isometry property (RIP) of the measurement matrix becomes relatively more popular than coherence for the analysis of various compressed sensing algorithms including OMP. The RIP is defined as follows.

Definition 1.1 ([11]): Let A be an $m \times n$ matrix. The restricted isometry constant (RIC) denoted by δ_q is the smallest number $\delta \ge 0$ such that

$$(1 - \delta)\|z\|_2^2 \le \|Az\|_2^2 \le (1 + \delta)\|z\|_2^2$$

for any q-sparse vector$z \in \mathbb{R}^n$. A is said to satisfy the restricted isometry property (RIP) of order q if $\delta_q < 1$.

It was shown in [40] that OMP can identify the support of a k-sparse signal in exact k iterations if $\delta_{k+1} < 1/(3\sqrt{k})$. This condition was relaxed to $\delta_{k+1} < 1/(1 + \sqrt{2k})$ in [41] and $\delta_{k+1} < 1/(\sqrt{k} + 1)$ in [42] and [43]. This condition was further improved to $\delta_{k+1} < (\sqrt{4k + 1} - 1)/(2k)$ in [44], which

is situated between $1/(\sqrt{k}+1)$ and $1/\sqrt{k}$. Examples are given in [42], [43] to show that the OMP algorithm might fail to achieve the correct support identification in exact k iterations when $\delta_{k+1} = 1/\sqrt{k}$. In noiseless situations (i.e., the measurements are accurate), it was shown that the condition $\delta_{k+1} < 1/\sqrt{k+1}$ is sharp to guarantee the success of OMP for k-sparse data reconstruction in exact k iterations [45]. In noisy settings (the measurements are inaccurate), the guaranteed performance of OMP has been shown in [46] and [47] under the condition $\delta_{k+1} < 1/(\sqrt{k}+1)$ and a certain assumption on the target signal. Another improvement was achieved recently in [48], and it was shown that the condition $\delta_{k+1} < 1/\sqrt{k+1}$ together with a condition on the signal is sufficient for OMP to correctly identify the support of a k-sparse vector in exact k iterations. Moreover, the stability of OMP was also studied under the RIP in [49].

While CoSaMP and SP are also motivated by OMP, the two algorithms equipped with a hard thresholding operator are very different from OMP. The interested readers can find the latest development of CoSaMP and SP algorithms in such reference as [8], [50], [51], [52], and can also find the results for other modifications of OMP such as StOMP, MMP, ROMP, gOMP and SWOMP. Here we only mention a few results for gOMP since it includes OMP as a special case and it is the closest kin to OMP. The gOMP algorithm was introduced first in [32] (see also [53], [54]). It was shown in [32] that the condition $\delta_{Nk} < \sqrt{N}/(\sqrt{k}+3\sqrt{N})$ is sufficient for the gOMP algorithm to reconstruct the k-sparse signal in k iterations. This condition was relaxed in [53] and [55] and was further relaxed to $\delta_{Nk} < \sqrt{N}/(\sqrt{k}+1.27\sqrt{N})$ in [56]. The reconstruction error via gOMP in noisy scenarios has been investigated in [48], [57], [58].

D. CONTRIBUTION OF THE PAPER

The main purpose of this paper is to propose an enhanced modification of OMP for sparse data reconstruction called *dynamic orthogonal matching pursuit* (DOMP) and its enhanced counterpart called *enhanced dynamic orthogonal matching pursuit* (EDOMP). The proposed algorithms dynamically select vector bases to reconstruct the target data according to the major gradient information of the error metric at every iteration. As a result, the gradient information is sufficiently and efficiently exploited to identify the support of the target data. The reconstruction error bound via DOMP is established under the RIP assumption. This bound also implies the finite convergence of DOMP in large-scale compressed sensing scenarios. Moreover, the numerical performance of the proposed algorithms is evaluated through random problem instances. The results indicate that due to the efficient usage of significant gradient information, the proposed algorithms are fast and robust for sparse data reconstruction and are very comparable to several state-of-art algorithms in this field.

The paper is organized as follows. The proposed algorithms are described in Section II. An approximation counterpart of the projection problem in the proposed algorithm is give in Section III, which is used to facilitate the theoretical analysis of the algorithm. Section IV is devoted to such a theoretical analysis which leads to the reconstruction error bound for the DOMP algorithm. The finite convergence of DOMP for large-scale compressed sensing problems is also discussed is Section IV. Numerical results are demonstrated in Section V.

DOMP Algorithm. Perform the following steps until a certain stopping criterion is satisfied:

S1 Initialization: Set $x^{(0)} = 0$, $S^{(0)} = \emptyset$ and $r^{(0)} = A^T(y - Ax^{(0)}) = A^T y$. Choose parameter $\gamma \in (0, 1]$.

S2 Given $x^{(p)}$, $S^{(p)}$ and $r^{(p)} = A^T(y - Ax^{(p)})$, let

$$\Theta^{(p)} = \{i : i \in L_k(r^{(p)}),\ |(r^{(p)})_i| \geq \gamma \|r^{(p)}\|_\infty\},$$

and then set

$$S^{(p+1)} = S^{(p)} \cup \Theta^{(p)}, \tag{DOMP1}$$

$$x^{(p+1)} = \arg\min_{z\in\mathbb{R}^n}\{\|y - Az\|_2 : \operatorname{supp}(z) \subseteq S^{(p+1)}\}. \tag{DOMP2}$$

II. DYNAMIC ORTHOGONAL MATCHING PURSUIT

The OMP algorithm only select one index corresponding to the largest magnitude of the gradient of error metric at every iteration. To increase the chance for the correct support of the target data being identified, it would be helpful to select indices at every iteration in a dynamic and adaptive manner. It is worth pointing out that this idea has been exploited in StOMP and SWOMP, but the total number of indices being selected at every iteration of these methods might increase too fast and exceed the sparsity level of the signal. This renders the algorithm either requires extra effort to reconstruct the signal or fails to reconstruct the signal. Thus we propose the following *dynamic orthogonal matching pursuit* (DOMP) method, which may efficiently utilize the underlying gradient information and control the number of selected indices at each step so that it does not exceed the sparsity level of signal. When the cardinality of selected indices does exceed the sparsity level, we may further enhance the algorithm by including a shrinkage step to bring the cardinality down to the sparsity level.

Two typical stopping criteria can be used in DOMP depending on the application environment: The algorithm may stop in prescribed maximum number of iterations, and may also stop when $\|y - Ax^{(p)}\|_2 \leq \varepsilon^*$ or $\|r^{(p)}\|_2 \leq \varepsilon^*$, where $\varepsilon^* > 0$ is a prespecified tolerance.

Since $x^{(p)}$ is an optimal solution to the convex optimization problem

$$\min\{\|y - Az\|_2 : \operatorname{supp}(z) \subseteq S^{(p)}\},$$

EDOMP Algorithm. Perform the following steps until a certain stopping criterion is satisfied:

S1 Initialization: Set $x^{(0)} = 0$, $S^{(0)} = \emptyset$, $r^{(0)} = A^T y$ and a parameter $\gamma \in (0, 1]$.

S2 Given $x^{(p)}$, $S^{(p)}$ and $r^{(p)} = A^T(y - Ax^{(p)})$, let

$$\Theta^{(p)} = \{i : i \in L_k(r^{(p)}),\ |(r^{(p)})_i| \geq \gamma \|r^{(p)}\|_\infty\}.$$

Set $S^{(p+1)} = S^{(p)} \cup \Theta^{(p)}$ and

$$\widetilde{x}^{(p+1)} = \arg\min_{z \in \mathbb{R}^n} \{\|y - Az\|_2 : \operatorname{supp}(z) \subseteq S^{(p+1)}\}.$$

If $|S^{(p+1)}| \leq k$, set $x^{(p+1)} = \widetilde{x}^{(p+1)}$; otherwise, set $Q = L_k(\widetilde{x}^{(p+1)})$ and

$$x^{(p+1)} = \arg\min_{z \in \mathbb{R}^n} \{\|y - Az\|_2 : \operatorname{supp}(z) \subseteq Q\}.$$

by the first-order optimality condition, the iterates generated by DOMP satisfy that

$$(r^{(p)})_{S^{(p)}} = 0 \text{ for any } p \geq 1. \tag{3}$$

This property will be frequently used in later analysis of the algorithm.

The DOMP algorithm dynamically select the vector basis according to the criterion $|(r^{(p)})_i| \geq \gamma \|r^{(p)}\|_\infty$ for $i \in L_k(r^{(p)})$. This criterion guarantees that only the indices corresponding to the significant entries of $r^{(p)}$ can be used to identify the support of target data. In general, $\Theta^{(p)}$ contains more than one elements corresponding to the indices of a few largest magnitudes of $r^{(p)}$. Clearly, even when $\gamma = 1$, the DOMP algorithm may not necessarily be the same as OMP since the set $\Theta^{(p)}$ may still contain more than one indices if $r^{(p)}$ possesses multiple entries whose absolute values are equal to $\|r^{(p)}\|_\infty$. The cardinality of $\Theta^{(p)}$ can be any value between 1 and k, depending on the prescribed value of γ and the number of significant entries of $r^{(p)}$. Such a dynamic selection of indices efficiently utilizes the gradient information of the error metric, and thus it might speed up the algorithm by increasing the chance for the correct support of target data being identified at every iteration. After a few iterations are performed, the cardinality of $S^{(p+1)}$ in DOMP might be larger than k. In this case, it is useful to perform certain shrinkage on the vector generated by (DOMP2) in order to maintain the k-sparsity of iterates so that they are feasible to the problem (2). This consideration leads to the following *enhanced dynamic orthogonal matching pursuit* (EDOMP), which is different from existing modifications of OMP.

Numerical results in Section V indicate that the capability of DOMP and EDOMP in sparse data reconstruction is generally stronger than OMP (see Section V for details). In the next section, we perform a theoretical analysis to establish a reconstruction error bound for the DOMP algorithm. Such an error bound measures the distance between the target data and the iterates generated by the algorithm. To facilitate this analysis, let us first define an auxiliary optimization model.

Approximation counterpart of (DOMP2): Let $x^{(p)}$, $S^{(p)}$, $\Theta^{(p)}$ and $S^{(p+1)}$ be given in DOMP, and let $\Lambda^{(p)}$ be defined by (4). Let

$$\widehat{S}^{(p+1)} = S^{(p+1)} \cup \Lambda^{(p)}, \tag{AC1}$$

$$\widehat{x}^{(p+1)} = \arg\min_{z \in \mathbb{R}^n} \left\{\|y - Az\|_2^2 + \sigma \|z_{\Lambda^{(p)}}\|_2^2 : \operatorname{supp}(z) \subseteq \widehat{S}^{p+1}\right\}, \tag{AC2}$$

where $\|z_{\Lambda^{(p)}}\|_2^2 = \sum_{i \in \Lambda^{(p)}} z_i^2$ and σ is a given large positive parameter.

III. APPROXIMATION COUNTERPART

Denote by $S = L_k(x)$, and let $x^{(p)}$, $S^{(p)}$, $r^{(p)}$, $\Theta^{(p)}$ and $S^{(p+1)}$ be defined in DOMP. Define

$$\Lambda^{(p)} := L_k(r^{(p)}) \setminus (S^{(p)} \cup \Theta^{(p)} \cup S). \tag{4}$$

We now introduce an approximation counterpart (AC) to the projection problem (DOMP2), which generates the index set $\widehat{S}^{(p+1)}$ and the vector $\widehat{x}^{(p+1)}$.

When $r^{(p)} = 0$, $x^{(p)}$ is already a global solution to the problem (2). Thus, to analyze the DOMP algorithm, we assume without loss of generality that $r^{(p)} \neq 0$, in which case we can see from the definition of $\Theta^{(p)}$ that $(r^{(p)})_i \neq 0$ for any $i \in \Theta^{(p)}$. This together with (3) implies that

$$S^{(p)} \cap \Theta^{(p)} = \emptyset. \tag{5}$$

Since $S^{(p+1)} = S^{(p)} \cup \Theta^{(p)}$, we also see from (4) that

$$\Lambda^{(p)} \cap S^{(p+1)} = \emptyset. \tag{6}$$

Thus

$$|S^{(p+1)}| = |S^{(p)}| + |\Theta^{(p)}|,\ |\widehat{S}^{(p+1)}| = |S^{(p+1)}| + |\Lambda^{(p)}|. \tag{7}$$

The solution of the optimization problem in (AC2) depends on σ, and thus $\widehat{x}^{(p+1)}$ should be written as $\widehat{x}^{(p+1)}(\sigma)$. For notational convenience, however, we simply use $\widehat{x}^{(p+1)}$ to denote the solution of (AC2). The above-described approximation counterpart is only used as an auxiliary tool to support the analysis of DOMP in the next section. From (AC2), one can see that for a sufficiently large σ, the components of $\widehat{x}^{(p+1)}$ supported on $\Lambda^{(p)}$ would vanish completely or be sufficiently small. Thus $\widehat{x}^{(p+1)}$ can be made arbitrarily close to $x^{(p+1)}$ (the iterate generated by DOMP) provided that σ is sufficiently large. The following lemma makes this rigorous.

Lemma 3.1: At the iterate $x^{(p)}$, let $x^{(p+1)}$ and $S^{(p+1)}$ be generated by DOMP, and let $\widehat{S}^{(p+1)}$ and $\widehat{x}^{(p+1)}$be defined in (AC1) and (AC2) accordingly. Then one has

$$\|A(\widehat{x}^{(p+1)} - x^{(p+1)})\|_2 \leq C\vartheta_{(y,A)}, \tag{8}$$

where $C := \sqrt{\frac{2\|A\|_2}{\sqrt{\sigma}} + \frac{\|A\|_2^2}{\sigma}} + \frac{\|A\|_2}{\sqrt{\sigma}}$ and $\vartheta_{(y,A)}$ is the constant defined as

$$\vartheta_{(y,A)} := \max_{U\subseteq\{1,\dots,n\}, U\neq\emptyset} (\min_{z\in\mathbb{R}^n}\{\|y-Az\|_2 : \operatorname{supp}(z)\subseteq U\}). \tag{9}$$

Proof: Suppose that $x^{(p)}$, $\Theta^{(p)}$ and $x^{(p+1)}$ are generated by DOMP. Let $\Lambda^{(p)}$ be defined by (4) and $\widehat{x}^{(p+1)}$ be a solution to the minimization problem (AC2). Note that $x^{(p+1)}$ is a feasible solution to the minimization problem (AC2), by the optimality of $\widehat{x}^{(p+1)}$, we have

$$\begin{aligned}\|y-A\widehat{x}^{(p+1)}\|_2^2 &+ \sigma\|(\widehat{x}^{(p+1)})_{\Lambda^{(p)}}\|_2^2\\ &\le \|y-Ax^{(p+1)}\|_2^2 + \sigma\|(x^{(p+1)})_{\Lambda^{(p)}}\|_2^2\\ &= \|y-Ax^{(p+1)}\|_2^2,\end{aligned} \tag{10}$$

where the equality follows from (6) which implies that $(x^{(p+1)})_{\Lambda^{(p)}} = 0$. Thus

$$\begin{aligned}\sigma\|(\widehat{x}^{(p+1)})_{\Lambda^{(p)}}\|_2^2 &\le \|y-Ax^{(p+1)}\|_2^2\\ &= \min_{z\in\mathbb{R}^n}\{\|y-Az\|_2^2 : \operatorname{supp}(z)\subseteq S^{(p+1)}\}\\ &\le \vartheta_{(y,A)}^2,\end{aligned}$$

where $\vartheta_{(y,A)}$ is the constant defined in (9). The inequalities above imply that

$$\|(\widehat{x}^{(p+1)})_{\Lambda^{(p)}}\|_2 \le \frac{\vartheta_{(y,A)}}{\sqrt{\sigma}},\ \|y-Ax^{(p+1)}\|_2 \le \vartheta_{(y,A)}. \tag{11}$$

We now bound $\|A(x^{(p+1)} - \widehat{x}^{(p+1)})\|_2$. From (10), we see that

$$\|y-A\widehat{x}^{(p+1)}\|_2 \le \|y-Ax^{(p+1)}\|_2. \tag{12}$$

By (AC1) and (6), we see that $\widehat{x}^{(p+1)} = (\widehat{x}^{(p+1)})_{\widehat{S}^{(p+1)}} = (\widehat{x}^{(p+1)})_{S^{(p+1)}} + (\widehat{x}^{(p+1)})_{\Lambda^{(p)}}$. Thus by the triangular inequality, we have

$$\begin{aligned}&\|y-A\widehat{x}^{(p+1)}\|_2\\ &\quad= \|y-A[(\widehat{x}^{(p+1)})_{S^{(p+1)}} + (\widehat{x}^{(p+1)})_{\Lambda^{(p)}}]\|_2\\ &\quad\ge \|y-A(\widehat{x}^{(p+1)})_{S^{(p+1)}}\|_2 - \|A(\widehat{x}^{(p+1)})_{\Lambda^{(p)}}\|_2.\end{aligned} \tag{13}$$

Merging (12) and (13) leads to

$$\|y-A(\widehat{x}^{(p+1)})_{S^{(p+1)}}\|_2 \le \|y-Ax^{(p+1)}\|_2 + \|A(\widehat{x}^{p+1})_{\Lambda^{(p)}}\|_2,$$

which together with (11) implies that

$$\begin{aligned}&\|y-A(\widehat{x}^{(p+1)})_{S^{(p+1)}}\|_2^2\\ &\le (\|y-Ax^{(p+1)}\|_2 + \|A(\widehat{x}^{(p+1)})_{\Lambda^{(p)}}\|_2)^2\\ &= \|y-Ax^{(p+1)}\|_2^2\\ &\quad + 2\|y-Ax^{(p+1)}\|_2\|A(\widehat{x}^{(p+1)})_{\Lambda^{(p)}}\|_2 + \|A(\widehat{x}^{(p+1)})_{\Lambda^{(p)}}\|_2^2\\ &\le \|y-Ax^{(p+1)}\|_2^2\\ &\quad + 2\vartheta_{(y,A)}\|A\|_2\|(\widehat{x}^{(p+1)})_{\Lambda^{(p)}}\|_2 + \|A\|_2^2\|(\widehat{x}^{(p+1)})_{\Lambda^{(p)}}\|_2^2\\ &\le \|y-Ax^{(p+1)}\|_2^2\\ &\quad + 2\vartheta_{(y,A)}\|A\|_2\left(\frac{\vartheta_{(y,A)}}{\sqrt{\sigma}}\right) + \|A\|_2^2\left(\frac{\vartheta_{(y,A)}}{\sqrt{\sigma}}\right)^2\\ &= \|y-Ax^{(p+1)}\|_2^2 + \left(\frac{2\|A\|_2}{\sqrt{\sigma}} + \frac{\|A\|_2^2}{\sigma}\right)(\vartheta_{(y,A)})^2.\end{aligned} \tag{14}$$

By Taylor's expansion, we also have that

$$\begin{aligned}&\|y-A(\widehat{x}^{(p+1)})_{S^{(p+1)}}\|_2^2 = \|y-Ax^{(p+1)}\|_2^2\\ &\quad + 2[-A^T(y-Ax^{(p+1)})]^T[(\widehat{x}^{(p+1)})_{S^{(p+1)}} - x^{(p+1)}]\\ &\quad + [(\widehat{x}^{(p+1)})_{S^{(p+1)}} - x^{(p+1)}]^T A^T A[(\widehat{x}^{(p+1)})_{S^{(p+1)}} - x^{(p+1)}]\\ &= \|y-Ax^{(p+1)}\|_2^2 + \|A[(\widehat{x}^{(p+1)})_{S^{(p+1)}} - x^{(p+1)}]\|_2^2,\end{aligned} \tag{15}$$

where the last equality follows from (3) and the fact that $\operatorname{supp}[(\widehat{x}^{(p+1)})_{S^{(p+1)}} - x^{(p+1)}] \subseteq S^{(p+1)}$. Combining (14) and (15) yields

$$\|A[(\widehat{x}^{(p+1)})_{S^{(p+1)}} - x^{(p+1)}]\|_2^2 \le \left[\frac{2\|A\|_2}{\sqrt{\sigma}} + \frac{\|A\|_2^2}{\sigma}\right](\vartheta_{(y,A)})^2. \tag{16}$$

By (AC1) and (6), i.e., $\widehat{S}^{(p+1)} = S^{(p+1)} \cup \Lambda^{(p)}$ and $S^{(p+1)} \cap \Lambda^{(p)} = \emptyset$, we immediately see that $(\overline{\widehat{S}^{(p+1)}}) \cap \Lambda^{(p)} = \emptyset$ and $\overline{S^{(p+1)}} = \overline{\widehat{S}^{(p+1)}} \setminus \Lambda^{(p)} = \overline{\widehat{S}^{(p+1)}} \cup \Lambda^{(p)}$, and hence

$$(\widehat{x}^{(p+1)})_{\overline{S^{(p+1)}}} = (\widehat{x}^{(p+1)})_{\overline{\widehat{S}^{(p+1)}}} + (\widehat{x}^{(p+1)})_{\Lambda^{(p)}} = (\widehat{x}^{(p+1)})_{\Lambda^{(p)}},$$

where the last equality follows from the fact that $(\widehat{x}^{(p+1)})_{\overline{\widehat{S}^{(p+1)}}} = 0$. This together with (11), (16) and $(x^{(p+1)})_{\overline{S^{(p+1)}}} = 0$ implies that

$$\begin{aligned}&\|A(\widehat{x}^{(p+1)} - x^{(p+1)})\|_2\\ &= \|A[(\widehat{x}^{(p+1)} - x^{(p+1)})_{S^{(p+1)}} + (\widehat{x}^{(p+1)} - x^{(p+1)})_{\overline{S^{(p+1)}}}]\|_2\\ &= \|A[(\widehat{x}^{(p+1)})_{S^{(p+1)}} - x^{(p+1)}] + A(\widehat{x}^{(p+1)})_{\Lambda^{(p)}}\|_2\\ &\le \|A[(\widehat{x}^{(p+1)})_{S^{(p+1)}} - x^{(p+1)}]\|_2 + \|A(\widehat{x}^{(p+1)})_{\Lambda^{(p)}}\|_2\\ &\le \vartheta_{(y,A)}\sqrt{\frac{2\|A\|_2}{\sqrt{\sigma}} + \frac{\|A\|_2^2}{\sigma}} + \|A\|_2\frac{\vartheta_{(y,A)}}{\sqrt{\sigma}}\end{aligned}$$

which is the inequality (8). ■

A basic property of the optimization problem (AC2) is given in the next lemma which follows directly from the optimality condition of convex optimization. The proof of the lemma is omitted.

Lemma 3.2: Let $\widehat{x}^{(p+1)}$ be a minimizer of (AC2), i.e.,

$$\widehat{x}^{(p+1)} = \arg\min_{z\in\mathbb{R}^n}\{\|y-Az\|_2^2 + \sigma\|z_{\Lambda^{(p)}}\|_2^2 : \operatorname{supp}(z)\subseteq \widehat{S}^{(p+1)}\},$$

where $\widehat{S}^{(p+1)}$ is defined by (AC1). Then

$$[-A^T(y-A\widehat{x}^{(p+1)}) + \sigma(\widehat{x}^{(p+1)})_{\Lambda^{(p)}}]_{\widehat{S}^{(p+1)}} = 0. \tag{17}$$

IV. RECONSTRUCTION ERROR BOUNDS

In this section, we establish the reconstruction error bound for DOMP. The efficiency of the algorithm for sparse data reconstruction can be interpreted via such an error bound

which measures the distance of the generated solution and target data. The error bound may also imply the stability and convergence of the algorithm under certain conditions. Let us first recall some properties of δ_t that will be used to establish the desired error bound.

Lemma 4.1 ([8]): Let $u \in \mathbb{R}^n$ and $\Lambda \subseteq \{1, 2, \ldots, n\}$. Then the following three statements are true:

i) If $|\Lambda \cup \text{supp}(u)| \leq t$, then $\|[(I - A^T A)u]_\Lambda\|_2 \leq \delta_t \|u\|_2$.

ii) If $\Lambda \cap \text{supp}(u) = \emptyset$ and $|\Lambda \cup \text{supp}(u)| \leq t$, then $\|(A^T Au)_\Lambda\|_2 \leq \delta_t \|u\|_2$.

iii) If $|\Lambda| \leq t$, one has $\|(A^T u)_\Lambda\|_2 \leq \sqrt{1+\delta_t}\|u\|_2$.

It should be pointed out that δ_t is not the only tool that can be used to analyze algorithms. Other tools such as the spark [1], mutual coherence [1], null space property [8] and range space property of A^T [15], [59] might be also used to achieve this goal. However, we only use δ_t to establish the main result in this paper.

A. MAIN RESULT FOR DOMP

The purpose of this section is to answer the question: How does the reconstruction error decay in the course of iterations? In other words, we would like to measure the quality of the iterate, generated by the algorithm, as an approximation to the target data. The result for DOMP is summarized in Theorem 4.2 below, whose proof is postponed to the end of the next subsection after we establish some helpful technical results.

Theorem 4.2: Suppose that $y := Ax + \nu$ are the measurements of the target data $x \in \mathbb{R}^n$ where ν are the measurement errors. Let $\gamma \in (0, 1]$ be a given constant. If the RIC of the measurement matrix A satisfies

$$\delta_{ck} < \frac{1}{\sqrt{1 + \left(1 + \sqrt{1 + k\gamma^2}\right)^2}}, \tag{18}$$

where $c > 2$ is an integer number, then at the pth iteration, provided that $|S^{p+1}| \leq (c-2)k$, the iterate $x^{(p+1)}$ generated by DOMP approximates x_S with error

$$\|x^{(p+1)} - x_S\|_2 \leq \beta^p \left((\varrho/\beta)^k \|x_S - x^{(0)}\|_2\right) + \tau\|\nu'\|_2, \tag{19}$$

where $\nu' = Ax_{\overline{S}} + \nu$ and $S = L_k(x)$. The constants β and ϱ are given respectively as

$$\beta = \frac{\left(1 + \sqrt{1 + k\gamma^2}\right)\delta_{ck}}{\sqrt{1 - \delta_{ck}^2}} < 1, \ \varrho = \sqrt{\frac{1 + \delta_{ck}}{1 - \delta_{ck}}},$$

and τ is a constant dependent only of δ_{ck}, k and γ.

Remark 4.3: If x is k-compressible with very small tail $\|x_{\overline{S}}\|_2$ and the measurements are accurate enough, then $\|\nu'\|_2$ would be very small. In particular, if x is k-sparse and the measurements are accurate, then $\|\nu'\|_2 = 0$. From (19), the reconstruction quality in these cases would completely depend on the iterations being executed. We also note that $|S^{(p+1)}|$ is the total number of indices being selected after p iterations. At the pth iteration, the indices in $\Theta^{(p)}$ are added to $S^{(p)}$ to form the set $S^{(p+1)}$. By (5), $|S^{(p+1)}| = \sum_{i=1}^{p} |\Theta^{(p)}| \leq pk$, thus the largest integer number p^* satisfying $|S^{(p^*+1)}| \leq (c-2)k$ would be at least $(c-2)$. This means when $c \gg 2$ the error bound (19) is satisfied for a large number $p = p^*$ and thus $\|x^{(p)} - x_S\|_2 \approx 0$. Thus $x^{(p)}$ is a quality reconstruction of x_S after the algorithm is performed enough iterations.

The lemma below follows immediately from Theorem 4.2.

Corollary 4.4: Let $y := Ax + \nu$ be the measurements of the target data $x \in \mathbb{R}^n$ where ν are the measurement errors. Let $\gamma \in (0, 1]$ be a given constant. If the matrix A satisfies the RIP of order $t^* > 2k$ where

$$t^* := \max\left\{t : \delta_t < \frac{1}{\sqrt{1 + \left(1 + \sqrt{1 + k\gamma^2}\right)^2}}\right\}, \tag{20}$$

then at the pth iteration, provided $|S^{p+1}| \leq t^* - 2k$, the iterate $x^{(p+1)}$ generated by DOMP approximates x_S with error

$$\|x^{(p+1)} - x_S\|_2 \leq \hat{\beta}^p \left((\hat{\varrho}/\hat{\beta})^k \|x_S - x^{(0)}\|_2\right) + \hat{\tau}\|\nu'\|_2,$$

where $\nu' = Ax_{\overline{S}} + \nu$ and $S = L_k(x)$. The constants $\hat{\beta}$ and $\hat{\varrho}$ are given respectively as

$$\hat{\beta} = \frac{\left(1 + \sqrt{1 + k\gamma^2}\right)\delta_{t^*}}{\sqrt{1 - \delta_{t^*}^2}} < 1, \ \hat{\varrho} = \sqrt{\frac{1 + \delta_{t^*}}{1 - \delta_{t^*}}},$$

and $\hat{\tau}$ is a constant dependent only of δ_{t^*}, k and γ.

The results above indicate that the quality of the iterate as an approximation to x_S can be measured provided that the total number of selected indices is lower than $t^* - 2k$. The RIC bounds (18) and (20) depends on the parameter γ. In theory, a small parameter γ can alleviate the requirement on measurement matrices since the smaller the parameter γ, the larger the right-hand sides of (18) and (20).

B. PROOF OF THE MAIN RESULT FOR DOMP

We first establish several useful technical results. The first one below is partially shown in [52].

Lemma 4.5: Given constants $\alpha_1, \alpha_2, \alpha_3 \geq 0$ where $\alpha_1 < 1$, if t satisfies that

$$t \leq \alpha_1\sqrt{t^2 + \alpha_2^2} + \alpha_3, \tag{21}$$

then

$$t \leq \frac{\alpha_1}{\sqrt{1 - \alpha_1^2}}\alpha_2 + \frac{\alpha_3}{1 - \alpha_1}, \ \sqrt{t^2 + \alpha_2^2} \leq \frac{\alpha_2}{\sqrt{1 - \alpha_1^2}} + \frac{\alpha_3}{1 - \alpha_1}. \tag{22}$$

Proof: The first inequality in (22) was shown in [52] (see Lemma 4.1 therein). Thus we only need to show the second inequality in (22), which follows from the first inequality and the fact that $\sqrt{(a+c)^2 + b^2} \leq \sqrt{a^2 + b^2} + c$ for any $a, b, c \geq 0$. ■

Some fundamental properties of DOMP and the optimization problem (AC2) are given in the next two lemmas.

Lemma 4.6: Let $y := Ax + \nu$, $S^{(p+1)}$ be generated by DOMP, $\widehat{S}^{(p+1)}$ be defined in (AC1), and let $\widehat{x}^{(p+1)}$ be a minimizer of (AC2). If A satisfies that $\delta_{k+|\widehat{S}^{(p+1)}|} < 1$, then

$$\|x_S - \widehat{x}^{(p+1)}\|_2 \le \frac{\|(x_S - \widehat{x}^{(p+1)})_{\overline{\widehat{S}^{(p+1)}}}\|_2}{\sqrt{1-\delta^2_{k+|\widehat{S}^{(p+1)}|}}} + \frac{\sqrt{1+\delta_{|\widehat{S}^{(p+1)}|}}}{1-\delta_{k+|\widehat{S}^{(p+1)}|}}\|\nu'\|_2, \quad (23)$$

where $\nu' = Ax_{\overline{S}} + \nu$ and $S = L_k(x)$.

Proof: By Lemma 3.2, we have that $[-A^T(y - A\widehat{x}^{(p+1)}) + \sigma(\widehat{x}^{(p+1)})_{\Lambda^{(p)}}]_{\widehat{S}^{(p+1)}} = 0$. As $y = Ax + \nu = Ax_S + \nu'$ where $\nu' = Ax_{\overline{S}} + \nu$, the equality above can be written as

$$\left[-A^T A(x_S - \widehat{x}^{(p+1)}) - A^T\nu' + \sigma(\widehat{x}^{(p+1)})_{\Lambda^{(p)}}\right]_{\widehat{S}^{(p+1)}} = 0.$$

Note that $[(\widehat{x}^{(p+1)})_{\Lambda^{(p)}}]_{\widehat{S}^{(p+1)}} = (\widehat{x}^{(p+1)})_{\Lambda^{(p)}}$ since $\Lambda^{(p)} \subseteq \widehat{S}^{(p+1)}$. The equality above is equivalent to

$$(x_S - \widehat{x}^{(p+1)})_{\widehat{S}^{(p+1)}} = [(I - A^T A)(x_S - \widehat{x}^{(p+1)})]_{\widehat{S}^{(p+1)}} - (A^T\nu')_{\widehat{S}^{(p+1)}} + \sigma(\widehat{x}^{(p+1)})_{\Lambda^{(p)}}. \quad (24)$$

By the definition of $\Lambda^{(p)}$ in (4), we see that $S \cap \Lambda^{(p)} = \emptyset$. Thus $(x_S)_{\Lambda^{(p)}} = 0$ and

$$\begin{aligned}[(x_S - \widehat{x}^{(p+1)})_{\widehat{S}^{(p+1)}}]^T(\widehat{x}^{(p+1)})_{\Lambda^{(p)}} &= [(x_S - \widehat{x}^{(p+1)})_{\Lambda^{(p)}}]^T(\widehat{x}^{(p+1)})_{\Lambda^{(p)}} \\ &= -\|(\widehat{x}^{(p+1)})_{\Lambda^{(p)}}\|_2^2 \le 0. \end{aligned} \quad (25)$$

Multiplying (24) by $(x_S - \widehat{x}^{(p+1)})_{\widehat{S}^{(p+1)}}$, using (25) and Lemma 4.1 (i) and (iii), we obtain

$$\begin{aligned}
&\|(x_S - \widehat{x}^{(p+1)})_{\widehat{S}^{(p+1)}}\|_2^2 \\
&= [(x_S - \widehat{x}^{(p+1)})_{\widehat{S}^{(p+1)}}]^T[(I - A^T A)(x_S - \widehat{x}^{(p+1)})]_{\widehat{S}^{(p+1)}} \\
&\quad - [(x_S - \widehat{x}^{(p+1)})_{\widehat{S}^{(p+1)}}]^T(A^T\nu')_{\widehat{S}^{(p+1)}} \\
&\quad + \sigma[(x_S - \widehat{x}^{(p+1)})_{\widehat{S}^{(p+1)}}]^T(\widehat{x}^{(p+1)})_{\Lambda^{(p)}} \\
&\le [(x_S - \widehat{x}^{(p+1)})_{\widehat{S}^{(p+1)}}]^T[(I - A^T A)(x_S - \widehat{x}^{(p+1)})]_{\widehat{S}^{(p+1)}} \\
&\quad - [(x_S - \widehat{x}^{(p+1)})_{\widehat{S}^{(p+1)}}]^T(A^T\nu')_{\widehat{S}^{(p+1)}} \\
&\le \delta_{k+|\widehat{S}^{(p+1)}|}\|x_S - \widehat{x}^{(p+1)}\|_2\|(x_S - \widehat{x}^{(p+1)})_{\widehat{S}^{(p+1)}}\|_2 \\
&\quad + \sqrt{1+\delta_{|\widehat{S}^{(p+1)}|}}\|\nu'\|_2\|(x_S - \widehat{x}^{(p+1)})_{\widehat{S}^{(p+1)}}\|_2 \\
&= \|(x_S - \widehat{x}^{(p+1)})_{\widehat{S}^{(p+1)}}\|_2(\delta_{k+|\widehat{S}^{(p+1)}|}\|x_S - \widehat{x}^{(p+1)}\|_2 \\
&\quad + \sqrt{1+\delta_{|\widehat{S}^{(p+1)}|}}\|\nu'\|_2). \end{aligned} \quad (26)$$

If $\|(x_S - \widehat{x}^{(p+1)})_{\widehat{S}^{(p+1)}}\|_2 = 0$, then $\|x_S - \widehat{x}^{(p+1)}\|_2 = \|(x_S - \widehat{x}^{(p+1)})_{\overline{\widehat{S}^{(p+1)}}}\|_2$, and hence the inequality (23) is satisfied trivially. Thus we assume without loss of generality that $\|(x_S - \widehat{x}^{(p+1)})_{\widehat{S}^{(p+1)}}\|_2 \ne 0$. We cancel this term from (26) to obtain that

$$\begin{aligned}
&\|(x_S - \widehat{x}^{(p+1)})_{\widehat{S}^{(p+1)}}\|_2 \\
&\le \delta_{k+|\widehat{S}^{(p+1)}|}\|x_S - \widehat{x}^{(p+1)}\|_2 + \sqrt{1+\delta_{|\widehat{S}^{(p+1)}|}}\|\nu'\|_2 \\
&= \delta_{k+|\widehat{S}^{(p+1)}|}\sqrt{\|(x - \widehat{x}^{(p+1)})_{\widehat{S}^{(p+1)}}\|_2^2 + \|(x - \widehat{x}^{(p+1)})_{\overline{\widehat{S}^{(p+1)}}}\|_2^2} \\
&\quad + \sqrt{1+\delta_{|\widehat{S}^{(p+1)}|}}\|\nu'\|_2,
\end{aligned}$$

which is in the form of (21) with $t = \|(x_S - \widehat{x}^{p+1})_{\widehat{S}^{(p+1)}}\|_2$, $\alpha_1 = \delta_{k+|\widehat{S}^{(p+1)}|}$, $\alpha_2 = \|(x_S - \widehat{x}^{(p+1)})_{\overline{\widehat{S}^{(p+1)}}}\|_2$, $\alpha_3 = \sqrt{1+\delta_{|\widehat{S}^{(p+1)}|}}\|\nu'\|_2$, and

$$\begin{aligned}
\|x_S - \widehat{x}^{(p+1)}\|_2 &= \sqrt{\|(x_S - \widehat{x}^{(p+1)})_{\widehat{S}^{(p+1)}}\|_2^2 + \|(x_S - \widehat{x}^{(p+1)})_{\overline{\widehat{S}^{(p+1)}}}\|_2^2} \\
&= \sqrt{t^2 + \alpha_2^2}.
\end{aligned}$$

Thus it immediately follows from Lemma 4.5 that

$$\|x_S - \widehat{x}^{(p+1)}\|_2 \le \frac{1}{\sqrt{1-\delta^2_{k+|\widehat{S}^{(p+1)}|}}}\|(x_S - \widehat{x}^{(p+1)})_{\overline{\widehat{S}^{(p+1)}}}\|_2 + \frac{\sqrt{1+\delta_{|\widehat{S}^{(p+1)}|}}}{1-\delta_{k+|\widehat{S}^{(p+1)}|}}\|\nu'\|_2,$$

which is exactly the relation (23). ■

Under the condition $\Theta^{(p)} \cap S = \emptyset$, we now bound the term $\|(x_S - \widehat{x}^{(p+1)})_{\overline{\widehat{S}^{(p+1)}}}\|_2$ which is equal to $\|(x_S - x^{(p)})_{\overline{\widehat{S}^{(p+1)}}}\|_2$ since $(\widehat{x}^{(p+1)})_{\overline{\widehat{S}^{(p+1)}}} = 0$ and $(x^{(p)})_{\overline{\widehat{S}^{(p+1)}}} = 0$.

Lemma 4.7: Let $y := Ax + \nu$, and let $x^{(p)}$, $S^{(p)}$, $\Theta^{(p)}$ and $S^{(p+1)}$ be generated by DOMP. Let $\widehat{S}^{(p+1)}$ be defined by (AC1), i.e., $\widehat{S}^{(p+1)} = S^{(p+1)} \cup \Lambda^{(p)}$ where $\Lambda^{(p)}$ is defined by (4). Suppose that $\Theta^{(p)} \cap S = \emptyset$ where $S = L_k(x)$. If the matrix A satisfies that $\delta_{k+|\widehat{S}^{(p+1)}|} < 1$, then one has

$$\|(x_S - x^{(p)})_{\overline{\widehat{S}^{(p+1)}}}\|_2 \le \rho^*(\delta_{2k+|S^{(p)}|}\|x_S - x^{(p)}\|_2 + \sqrt{1+\delta_k}\|\nu'\|_2), \quad (27)$$

where $\rho^* = 1 + \sqrt{1+\gamma^2 k}$, $\nu' = Ax_{\overline{S}} + \nu$ and $\gamma \in (0, 1]$ is a given parameter in DOMP.

Proof: Let $x^{(p)}$ be the current iterate and $r^{(p)}$, $S^{(p)}$, $\Theta^{(p)}$, $S^{(p+1)}$ be defined in DOMP. Suppose that $\Theta^{(p)}$ and S are disjoint, i.e.,

$$S \cap \Theta^{(p)} = \emptyset. \quad (28)$$

We denote by $\widetilde{L} := L_k(r^{(p)})$ for notational convenience. Note that $|\widetilde{L}| = k = |S|$. Thus

$$|S\backslash\widetilde{L}| = |S| - |S \cap \widetilde{L}| = |\widetilde{L}| - |S \cap \widetilde{L}| = |\widetilde{L}\backslash S|. \quad (29)$$

This means the number of elements in $S\backslash\widetilde{L}$ is equal to the number of elements in $\widetilde{L}\backslash S$. By the definition of $\widetilde{L}$, the entries of $r^{(p)} = A^T(y - Ax^{(p)})$ supported on $\widetilde{L}\backslash S$ are among the k

largest magnitudes of $r^{(p)}$. This together with (29) implies that

$$\|(r^{(p)})_{S\setminus\widetilde{L}}\|_2 \le \|(r^{(p)})_{\widetilde{L}\setminus S}\|_2. \tag{30}$$

By (4), $S \cap \Lambda^{(p)} = \emptyset$ which together with (28) implies that

$$S \cap \widehat{S}^{(p+1)} = S \cap (S^{(p)} \cup \Theta^{(p)} \cup \Lambda^{(p)}) = S \cap S^{(p)}. \tag{31}$$

Note that $S = (S\setminus T) \cup (S \cap T)$ for any set T, where $S \setminus T$ and $S \cap T$ are disjoint. By setting $T = \widehat{S}^{(p+1)}$ and $T = \widetilde{L}$ respectively, we immediately obtain the two relations:

$$\|(r^{(p)})_S\|_2^2 = \|(r^{(p)})_{S\setminus\widehat{S}^{(p+1)}}\|_2^2 + \|(r^{(p)})_{S\cap\widehat{S}^{(p+1)}}\|_2^2, \tag{32}$$

$$\|(r^{(p)})_S\|_2^2 = \|(r^{(p)})_{S\setminus\widetilde{L}}\|_2^2 + \|(r^{(p)})_{S\cap\widetilde{L}}\|_2^2. \tag{33}$$

Note that $x^{(p)}$ is a minimizer of the convex optimization problem $\min\{\|y - Az\|_2 : \mathrm{supp}(z) \subseteq S^{(p)}\}$. Thus $x^{(p)}$ and $S^{(p)}$ satisfy (3), i.e.,

$$(r^{(p)})_{S^{(p)}} = 0. \tag{34}$$

This together with (31) implies that $\|(r^{(p)})_{S\cap\widehat{S}^{(p+1)}}\|_2^2 = \|(r^{(p)})_{S\cap S^{(p)}}\|_2^2 = 0$. Thus, merging (32) and (33) yields

$$\|(r^{(p)})_{S\setminus\widehat{S}^{(p+1)}}\|_2^2 = \|(r^{(p)})_{S\setminus\widetilde{L}}\|_2^2 + \|(r^{(p)})_{S\cap\widetilde{L}}\|_2^2. \tag{35}$$

By (30), (34), and (35), one has

$$\begin{aligned}\Omega^2 &:= \|(r^{(p)})_{(S\cup S^{(p)})\setminus\widehat{S}^{(p+1)}}\|_2^2 = \|(r^{(p)})_{S\setminus\widehat{S}^{(p+1)}}\|_2^2\\ &= \|(r^{(p)})_{S\setminus\widetilde{L}}\|_2^2 + \|(r^{(p)})_{S\cap\widetilde{L}}\|_2^2\\ &\le \|(r^{(p)})_{\widetilde{L}\setminus S}\|_2^2 + \|(r^{(p)})_{S\cap\widetilde{L}}\|_2^2\\ &= \|(r^{(p)})_{\widetilde{L}\setminus(S\cup S^{(p)})}\|_2^2 + \|(r^{(p)})_{S\cap\widetilde{L}}\|_2^2,\end{aligned} \tag{36}$$

where the last equality follows from (34). In fact, by eliminating the components of $(r^{(p)})_{\widetilde{L}\setminus S}$ indexed by $S^{(p)}$, we immediately see that

$$\|(r^{(p)})_{\widetilde{L}\setminus S}\|_2 = \|(r^{(p)})_{\widetilde{L}\setminus(S\cup S^{(p)})}\|_2. \tag{37}$$

We now estimate the last term $\|(r^{(p)})_{S\cap\widetilde{L}}\|_2^2$ in (36). Note that the elements in $\Theta^{(p)}$ are added to $S^{(p)}$ to form the next set $S^{(p+1)}$. From the definition of $\Theta^{(p)}$, it contains the indices in $\widetilde{L} = L_k(r^{(p)})$ such that $|(r^{(p)})_i| \ge \gamma\|r^{(p)}\|_\infty$. Clearly, we have either $|\Theta^{(p)}| < k$ or $|\Theta^{(p)}| = k$. We now consider the two cases separately.

Case 1: $|\Theta^{(p)}| = k$. In this case, all indices of $\widetilde{L}$ are selected into Θ^k. So $\Theta^{(p)} = \widetilde{L}$. Under the assumption (28), it implies that $S \cap \widetilde{L} = S \cap \Theta^{(p)} = \emptyset$ and hence $\|(r^{(p)})_{S\cap\widetilde{L}}\|_2 = 0$. As a result, the relation (36) reduces to $\Omega \le \|(r^{(p)})_{\widetilde{L}\setminus(S\cup S^{(p)})}\|_2$.

Case 2: $|\Theta^{(p)}| < k$. In this case, by the structure of DOMP, not all indices in $\widetilde{L}$ are selected into $\Theta^{(p)}$. Without loss of generality, we assume that $S \cap \widetilde{L} \ne \emptyset$ since otherwise $\|(r^{(p)})_{S\cap\widetilde{L}}\|_2^2 = 0$. Under the condition (28), we see that $(S \cap \widetilde{L}) \cap \Theta^{(p)} = (S \cap \Theta^{(p)}) \cap \widetilde{L} = \emptyset$. So the indices in $S \cap \widetilde{L}$ are not in $\Theta^{(p)}$, and hence $|(r^{(p)})_i| < \gamma\|r^{(p)}\|_\infty$ for all $i \in S \cap \widetilde{L}$, which implies that

$$\|(r^{(p)})_{S\cap\widetilde{L}}\|_\infty < \gamma\|r^{(p)}\|_\infty. \tag{38}$$

Again, as $\Theta^{(p)} \subseteq \widetilde{L}$ and the elements in $S \cap \widetilde{L}$ are not in $\Theta^{(p)}$, we see that $\Theta^{(p)} \subseteq \widetilde{L}\setminus(\widetilde{L} \cap S) = \widetilde{L}\setminus S$. Since the index of the largest magnitude of $r^{(p)}$ is in $\Theta^{(p)}$, one has

$$\|r^{(p)}\|_\infty = \|(r^{(p)})_{\Theta^p}\|_\infty \le \|(r^{(p)})_{\Theta^p}\|_2 \le \|(r^{(p)})_{\widetilde{L}\setminus S}\|_2.$$

Merging this relation and (37) yields

$$\|r^{(p)}\|_\infty \le \|(r^{(p)})_{\widetilde{L}\setminus(S\cup S^{(p)})}\|_2. \tag{39}$$

By using (38) and (39), we deduce that

$$\begin{aligned}\|(r^{(p)})_{S\cap\widetilde{L}}\|_2 &\le \sqrt{|S\cap\widetilde{L}|}\,\|(r^{(p)})_{S\cap\widetilde{L}}\|_\infty\\ &< \gamma\sqrt{|S\cap\widetilde{L}|}\,\|r^{(p)}\|_\infty\\ &\le \gamma\sqrt{|S\cap\widetilde{L}|}\,\|(r^{(p)})_{\widetilde{L}\setminus(S\cup S^{(p)})}\|_2\\ &\le \gamma\sqrt{k}\,\|(r^{(p)})_{\widetilde{L}\setminus(S\cup S^{(p)})}\|_2,\end{aligned} \tag{40}$$

where the last inequality follows from $\gamma\sqrt{|S\cap\widetilde{L}|} \le \gamma\sqrt{k}$. Combining (36) and (40) leads to

$$\Omega \le \sqrt{1+\gamma^2 k}\,\left\|(r^{(p)})_{\widetilde{L}\setminus(S\cup S^{(p)})}\right\|_2. \tag{41}$$

So the bound (41) is valid in both Case 1 and Case 2. Since $[\widetilde{L}\setminus(S \cup S^{(p)})] \cap \mathrm{supp}(x_S - x^{(p)}) = \emptyset$, by Lemma 4.1 (ii) and (iii), we obtain

$$\begin{aligned}&\|(r^{(p)})_{\widetilde{L}\setminus(S\cup S^{(p)})}\|_2\\ &= \|(A^TA(x_S - x^{(p)}) + A^T\nu')_{\widetilde{L}\setminus(S\cup S^{(p)})}\|_2\\ &\le \|(A^TA(x_S - x^{(p)}))_{\widetilde{L}\setminus(S\cup S^{(p)})}\|_2 + \|(A^T\nu')_{\widetilde{L}\setminus(S\cup S^{(p)})}\|_2\\ &\le \delta_{k+|S^{(p)}|+|\widetilde{L}|}\|x_S - x^{(p)}\|_2 + \sqrt{1+\delta_k}\|\nu'\|_2\\ &= \delta_{2k+|S^{(p)}|}\|x_S - x^{(p)}\|_2 + \sqrt{1+\delta_k}\|\nu'\|_2.\end{aligned}$$

Merging (41) and the relation above leads to

$$\Omega \le \sqrt{1+k\gamma^2}\left(\delta_{2k+|S^{(p)}|}\|x_S - x^{(p)}\|_2 + \sqrt{1+\delta_k}\|\nu'\|_2\right). \tag{42}$$

Denote by

$$W := [(x_S - x^{(p)}) - A^T(y - Ax^{(p)})]_{(S\cup S^{(p)})\setminus\widehat{S}^{(p+1)}}.$$

Note that $|\mathrm{supp}(x_S - x^{(p)})| \le k + |S^{(p)}|$ and $|(S \cup S^{(p)}) \setminus \widehat{S}^{(p+1)}| \le |S| = k$ since $S^{(p)} \subseteq \widehat{S}^{(p+1)}$. By Lemma 4.1 (i) and (iii), one has

$$\begin{aligned}\|W\|_2 &= \|[(I - A^TA)(x_S - x^{(p)}) - A^T\nu']_{(S\cup S^{(p)})\setminus\widehat{S}^{(p+1)}}\|_2\\ &\le \delta_{k+|S^{(p)}|}\|x_S - x^{(p)}\|_2 + \sqrt{1+\delta_k}\|\nu'\|_2.\end{aligned} \tag{43}$$

Note that $x_S - x^{(p)} = (x_S - x^{(p)})_{(S\cup S^{(p)})}$ since $\mathrm{supp}(x_S - x^{(p)}) \subseteq S \cup S^{(p)}$. We have

$$\begin{aligned}&\|(x_S - x^{(p)})_{\overline{\widehat{S}^{(p+1)}}}\|_2\\ &= \|(x_S - x^{(p)})_{(S\cup S^{(p)})\setminus\widehat{S}^{(p+1)}}\|_2\\ &= \|[x_S - x^{(p)} - A^T(y - Ax^{(p)})]_{(S\cup S^{(p)})\setminus\widehat{S}^{(p+1)}}\end{aligned}$$

$$+ [A^T(y - Ax^{(p)})]_{(S\cup S^{(p)})\setminus \widehat{S}^{(p+1)}}\|_2$$
$$\leq \|W\|_2 + \Omega. \tag{44}$$

Denote by $\phi := 1 + \sqrt{1 + k\gamma^2}$. Combining (42), (43) and (44) yields

$$\begin{aligned}
&\|(x_S - x^{(p)})_{\widehat{S}^{(p+1)}}\|_2 \\
&= \left(\delta_{k+|S^{(p)}|} + \sqrt{1+k\gamma^2}\delta_{2k+|S^{(p)}|}\right)\|x_S - x^{(p)}\|_2 \\
&+ \phi\sqrt{1+\delta_k}\|\nu'\|_2 \\
&\leq \phi\left(\delta_{2k+|S^{(p)}|}\|x_S - x^{(p)}\|_2 + \sqrt{1+\delta_k}\|\nu'\|_2\right),
\end{aligned}$$

where the last inequality follows from the fact $\delta_{k+|S^{(p)}|} \leq \delta_{2k+|S^{(p)}|}$. Thus (27) is satisfied. ■

We have all ingredients needed to show the main result stated in Theorem 4.2.

The Proof of Theorem 4.2: Let $x^{(p)}$ be the current iterate and $S^{(p)}$, $\Theta^{(p)}$, $S^{(p+1)}$ and $x^{(p+1)}$ are defined by DOMP. Without loss of generality, we assume $r^{(p)} \neq 0$. Let $\widehat{S}^{(p+1)}$ be defined as (AC1) and $\widehat{x}^{(p+1)}$ be a solution to the auxiliary minimization problem in (AC2). Still, $S = L_k(x)$ is the index set for the k largest magnitudes of the target vector x. At the pth iteration, there are only two cases: either $S \cap \Theta^{(p)} \neq \emptyset$ or $S \cap \Theta^{(p)} = \emptyset$.

Case 1: $S \cap \Theta^{(p)} \neq \emptyset$. In this case, at least one of the indices in S is added to $S^{(p+1)} = S^{(p)} \cup \Theta^p$. By (5), $\Theta^{(p)} \cap S^{(p)} = \emptyset$. Thus set $S^{(p+1)}$ contains more elements of S than $S^{(p)}$ in this case. Note that $\mathrm{supp}(x^{(p)}) \subseteq S^{(p)} \subseteq S^{(p+1)}$. Thus $x^{(p)}$ is a feasible vector to the optimization problem (DOMP2), to which $x^{(p+1)}$ is a minimizer. Thus by optimality, one has $\|y - Ax^{(p+1)}\|_2 \leq \|y - Ax^{(p)}\|_2$. Note that $y = Ax_S + \nu'$ where $\nu' = Ax_{\overline{S}} + \nu$. By the triangular inequality, it follows from the inequality above that

$$\|A(x_S - x^{(p+1)})\|_2 - \|\nu'\|_2 \leq \|A(x_S - x^{(p)})\|_2 + \|\nu'\|_2,$$

By noting that $|\mathrm{supp}(x_S - x^{(p+1)})| \leq k + |S^{(p+1)}|$ and $|\mathrm{supp}(x_S - x^{(p)})| \leq k + |S^{(p)}|$ and by the definition of RIC of A, the inequality above implies that

$$\begin{aligned}
&\sqrt{1 - \delta_{k+|S^{(p+1)}|}}\|x^{(p+1)} - x_S\|_2 \\
&\leq \sqrt{1 + \delta_{k+|S^{(p)}|}}\|x^{(p)} - x_S\|_2 + 2\|\nu'\|_2.
\end{aligned}$$

Since $\delta_{k+|S^{(p)}|} \leq \delta_{k+|S^{(p+1)}|}$, one has

$$\begin{aligned}
\|x^{(p+1)} - x_S\|_2 \leq &\sqrt{\frac{1 + \delta_{k+|S^{(p+1)}|}}{1 - \delta_{k+|S^{(p+1)}|}}}\|x^{(p)} - x_S\|_2 \\
&+ \frac{2}{\sqrt{1 - \delta_{k+|S^{(p+1)}|}}}\|\nu'\|_2.
\end{aligned} \tag{45}$$

The first relation in (7) implies that $|S^{(p+1)}| = \sum_{i=1}^{p} |\Theta^{(i)}|$, which means $\{|S^{(p)}|\}$ is nondecreasing sequence. Let p^* denote the largest integer number p such that $|S^{(p+1)}| \leq (c-2)k$. Then for any $p \leq p^*$, the coefficients on the right-hand side of (45) are bounded by

$$\sqrt{\frac{1 + \delta_{k+|S^{(p+1)}|}}{1 - \delta_{k+|S^{(p+1)}|}}} \leq \varrho, \quad \frac{2}{\sqrt{1 - \delta_{k+|S^{(p+1)}|}}} \leq c_1, \tag{46}$$

where $\varrho := \sqrt{\frac{1+\delta_{ck}}{1-\delta_{ck}}}$ and $c_1 := \frac{2}{\sqrt{1-\delta_{ck}}}$. Thus it follows from (45) and (46) that

$$\|x^{(p+1)} - x_S\|_2 \leq \varrho\|x^{(p)} - x_S\|_2 + c_1\|\nu'\|_2 \tag{47}$$

for any $p \leq p^*$. We now analyze the second case.

Case 2: $S \cap \Theta^{(p)} = \emptyset$. Let p^* be defined as above, i.e., p^* is the largest number p such that $|S^{(p+1)}| \leq (c-2)k$. Denote by $\phi = 1 + \sqrt{1 + k\gamma^2}$. By Lemmas 4.6 and 4.7, we immediately obtain that

$$\begin{aligned}
&\|x_S - \widehat{x}^{(p+1)}\|_2 \\
&\leq \frac{\phi(\delta_{2k+|S^{(p)}|}\|x_S - x^{(p)}\|_2 + \sqrt{1+\delta_k}\|\nu'\|_2)}{\sqrt{1 - \delta^2_{k+|\widehat{S}^{(p+1)}|}}} \\
&\quad + \frac{\sqrt{1 + \delta_{|\widehat{S}^{(p+1)}|}}\|\nu'\|_2}{1 - \delta_{k+|\widehat{S}^{(p+1)}|}} \\
&= \frac{\phi\delta_{2k+|S^{(p)}|}\|x_S - x^{(p)}\|_2}{\sqrt{1 - \delta^2_{k+|\widehat{S}^{(p+1)}|}}} \\
&\quad + \left(\frac{\phi\sqrt{1+\delta_k}}{\sqrt{1 - \delta^2_{k+|\widehat{S}^{(p+1)}|}}} + \frac{\sqrt{1 + \delta_{|\widehat{S}^{(p+1)}|}}}{1 - \delta_{k+|\widehat{S}^{(p+1)}|}}\right)\|\nu'\|_2 \\
&\leq \frac{\phi\delta_{2k+|S^{(p+1)}|}\|x_S - x^{(p)}\|_2}{\sqrt{1 - \delta^2_{2k+|S^{(p+1)}|}}} \\
&\quad + \left(\frac{\phi\sqrt{1+\delta_k}}{\sqrt{1 - \delta^2_{2k+|S^{(p+1)}|}}} + \frac{\sqrt{1 + \delta_{2k+|S^{(p+1)}|}}}{1 - \delta_{2k+|S^{(p+1)}|}}\right)\|\nu'\|_2 \\
&\leq \frac{\phi\delta_{ck}\|x_S - x^{(p)}\|_2}{\sqrt{1-\delta^2_{ck}}} + \left(\frac{\phi\sqrt{1+\delta_k}}{\sqrt{1-\delta^2_{ck}}} + \frac{\sqrt{1+\delta_{ck}}}{1-\delta_{ck}}\right)\|\nu'\|_2,
\end{aligned} \tag{48}$$

where the second inequality follows from the fact that $\delta_{2k+|S^{(p)}|} \leq \delta_{2k+|S^{(p+1)}|}$ since $|S^{(p)}| \leq |S^{(p+1)}|$, and that $\delta_{k+|\widehat{S}^{(p+1)}|} \leq \delta_{2k+|S^{(p+1)}|}$ since $|\widehat{S}^{(p+1)}| = |S^{(p+1)}| + |\Lambda^{(p)}| \leq k + |S^{(p+1)}|$ which follows from the second relation in (7). The last inequality follows from the fact $\delta_{2k+|S^{(p+1)}|} \leq \delta_{ck}$ due to $p \leq p^*$. Denote by

$$\epsilon := \frac{\vartheta_{(y,A)}}{\sqrt{1-\delta_{ck}}}\left(\sqrt{\left(\frac{2}{\sqrt{\sigma}} + \frac{\|A\|_2}{\sigma}\right)\|A\|_2} + \frac{\|A\|_2}{\sqrt{\sigma}}\right), \tag{49}$$

where $\vartheta_{(y,A)}$ and σ are defined in Lemma 3.1. Since $\mathrm{supp}(\widehat{x}^{(p+1)} - x^{(p+1)}) \subseteq \widehat{S}^{(p+1)}$, by the definition of RIC

of A and Lemma 3.1, we have

$$\begin{aligned}&\|\widehat{x}^{(p+1)} - x^{(p+1)}\|_2\\ &\le \frac{\|A(\widehat{x}^{(p+1)} - x^{(p+1)})\|_2}{\sqrt{1-\delta_{|\widehat{S}^{(p+1)}|}}} \le \frac{\|A(\widehat{x}^{(p+1)} - x^{(p+1)})\|_2}{\sqrt{1-\delta_{ck}}}\\ &\le \frac{\vartheta_{(y,A)}}{\sqrt{1-\delta_{ck}}}\left(\sqrt{\left(\frac{2}{\sqrt{\sigma}} + \frac{\|A\|_2}{\sigma}\right)\|A\|_2} + \frac{\|A\|_2}{\sqrt{\sigma}}\right) = \epsilon.\end{aligned} \tag{50}$$

Then using the triangular inequality and (48) and (50) yields

$$\begin{aligned}\|x_S - x^{(p+1)}\|_2 &\le \|x_S - \widehat{x}^{(p+1)}\|_2 + \|\widehat{x}^{(p+1)} - x^{(p+1)}\|_2\\ &\le \frac{\phi\delta_{ck}}{\sqrt{1-\delta_{ck}^2}}\|x_S - x^{(p)}\|_2\\ &\quad + \left(\frac{\phi\sqrt{1+\delta_k}}{\sqrt{1-\delta_{ck}^2}} + \frac{\sqrt{1+\delta_{ck}}}{1-\delta_{ck}}\right)\|\nu'\|_2 + \epsilon\\ &= \beta\|x_S - x^{(p)}\|_2 + c_2\|\nu'\|_2 + \epsilon,\end{aligned} \tag{51}$$

where the constants β and c_2 are given by

$$\beta = \frac{\phi\delta_{ck}}{\sqrt{1-\delta_{ck}^2}},\quad c_2 = \left(\frac{\phi\sqrt{1+\delta_k}}{\sqrt{1-\delta_{ck}^2}} + \frac{\sqrt{1+\delta_{ck}}}{1-\delta_{ck}}\right). \tag{52}$$

It is easy to verify that $\beta < 1$ under the assumption of the theorem. In fact, $\beta = \frac{\phi\delta_{ck}}{\sqrt{1-\delta_{ck}^2}} < 1$ is guaranteed under the condition

$$\delta_{ck} < \frac{1}{\sqrt{1+\phi^2}} = \frac{1}{\sqrt{1+\left(1+\sqrt{1+k\gamma^2}\right)^2}}.$$

Suppose that the algorithm DOMP has performed p iterations where $p \le p^*$. Without loss of generality, suppose that within these p iterations the aforementioned *Case 1* appears τ^* times, and thus *Case 2* appears $p - \tau^*$ times. Clearly, *Case 1* appears at most k times since $|S| = k$, so $\tau^* \le k$. The error bound for this case is given by (47), and the error bound for *Case 2* is given by (51). Therefore, it is not difficult to verify that after p iterations, one has

$$\begin{aligned}\|x_S - x^{(p+1)}\|_2 &\le (\varrho^{\tau^*}\beta^{p-\tau^*})\|x_S - x^{(0)}\|_2\\ &\quad + c_1\beta^{p-\tau^*}\left(\frac{\varrho^{\tau^*}-1}{\varrho-1}\right)\|\nu'\|_2\\ &\quad + \frac{c_2\|\nu'\|_2 + \epsilon}{1-\beta}.\end{aligned} \tag{53}$$

where all constants $(\beta, \varrho, c_1, c_2)$ are determined only by δ_{ck} and/or k and γ. To see (53), without loss of generality, we assume that the first τ^* iterations correspond to *Case 1*, and the remaining $p - \tau^*$ iterations correspond to *Case 2*. Then using the fact $\sum_{i=0}^{p-\tau^*-1}\beta^i = (1-\beta^{p-\tau^*})/(1-\beta) < 1/(1-\beta)$ for $\beta < 1$, it follows from (51) that

$$\begin{aligned}\|x_S - x^{(p+1)}\|_2 &\le \beta^{p-\tau^*}\|x_S - x^{(\tau^*+1)}\|_2\\ &\quad + (1+\beta+\ldots+\beta^{p-\tau^*-1})(c_2\|\nu'\|_2 + \epsilon)\\ &\le \beta^{p-\tau^*}\|x_S - x^{(\tau^*+1)}\|_2 + \frac{c_2\|\nu'\|_2 + \epsilon}{1-\beta}.\end{aligned} \tag{54}$$

Also, by (47) and the fact $\sum_{i=0}^{\tau^*-1}\varrho^i = (\varrho^{\tau^*}-1)/(\varrho-1)$ where $\varrho > 1$, we obtain

$$\begin{aligned}&\|x_S - x^{(\tau^*+1)}\|_2\\ &\le \varrho^{\tau^*}\|x_S - x^{(0)}\|_2 + (1+\varrho+\cdots+\varrho^{\tau^*-1})(c_1\|\nu'\|_2)\\ &= \varrho^{\tau^*}\|x_S - x^{(0)}\|_2 + c_1\left(\frac{\varrho^{\tau^*}-1}{\varrho-1}\right)\|\nu'\|_2.\end{aligned} \tag{55}$$

Merging (54) and (55) yields

$$\begin{aligned}\|x_S - x^{(p+1)}\|_2 &\le \beta^{p-\tau^*}(\varrho^{\tau^*}\|x_S - x^{(0)}\|_2)\\ &\quad + c_1\beta^{p-\tau^*}\left(\frac{\varrho^{\tau^*}-1}{\varrho-1}\right)\|\nu'\|_2 + \frac{c_2\|\nu'\|_2 + \epsilon}{1-\beta},\end{aligned}$$

which is exactly the error bound given in (53). Note that the coefficients $(\varrho/\beta)^{\tau^*} \le (\varrho/\beta)^k$ and $c_1\beta^{p-\tau^*}\left(\frac{\varrho^{\tau^*}-1}{\varrho-1}\right) \le c_1\left(\frac{\varrho^k-1}{\varrho-1}\right)$ since $\beta < 1$, $\varrho > 1$ and $\tau^* \le k$. Thus we deduce that

$$\begin{aligned}\|x_S - x^{(p+1)}\|_2 &\le \beta^p\left[(\varrho/\beta)^k\|x_S - x^{(0)}\|_2\right] + \tau\|\nu'\|_2\\ &\quad + \epsilon/(1-\beta).\end{aligned} \tag{56}$$

where $\tau = c_1\left(\frac{\varrho^k-1}{\varrho-1}\right) + \frac{c_2}{1-\beta}$ with constants $c_1 = 2/\sqrt{1-\delta_{ck}}$ and c_2 given in (52). The term $(\varrho/\beta)^k\|x_S - x^{(0)}\|_2$ in (56) is a fixed quantity. In (56), only the last term $\epsilon/(1-\beta)$ depends on σ, and we see from the definition of ϵ that $\epsilon \to 0$ as $\sigma \to \infty$. Therefore, the desired estimation (19) is obtained. ■

The main result established in this section provides an estimation for the distance between the significant components of the target data and the solution generated by the DOMP algorithm. The result claims that the reconstruction quality becomes better and better as the iteration proceeds provided that $|S^{(p)}| \le (c-2)k$, where k is the sparsity level of the target data and ck is the order of the RIC of the measurement matrix. In this process, the algorithm gradually identifies the correct support of the sparse data and the iterate $x^{(p)}$ gets closer and closer to the significant components of the target data. When m, n are particularly large with $n \gg m$ and the highest order RIC of A is much larger than $2k$, the proposed algorithms can perform enough iterations to guarantee that the

reconstruction error becomes so small that the finite convergence of the algorithms can be achieved, as discussed in the next subsection.

C. FINITE CONVERGENCE

From Definition 1.1, we see that $\delta_s \leq \delta_t$ for any positive integer number $s \leq t$. Thus if A has the RIP of order t, i.e., $\delta_t < 1$, then it has the RIP of any order $s \leq t$. In this section, we define the highest order of the RIP of A as

$$t^{\max} = \max\{t : \delta_t < 1\}.$$

The basic assumption in compressed sensing is that the number of measurements m is much lower than the signal length n, i.e., $m \ll n$. It is also well known that to ensure a k-sparse signal being exactly reconstructed, the signal should be sparse enough so that it is the unique sparsest solution to the underdetermined linear system $y = Ax$ (see [1]). We now consider the reconstruction problems satisfying the following assumption.

Assumption 4.8: $k \ll m \ll n$ and $2k \ll t^{\max}$, where $t^{\max}$ is the highest order of RIP of the measurement matrix A.

The problem in such settings may be referred to as a large-scale compressed sensing problem for which the DOMP algorithm can maintain the error bound (19) with enough iterations so that the significant information of the target data can be guaranteed to reconstruct. Let c be the largest integer number such that $2k \ll ck \leq t^{\max}$, i.e., $c = \lfloor \frac{t^{\max}}{k} \rfloor$. Recall that p^* (defined in the proof of Theorem 4.2) denotes the largest number of iterations such that $|S^{(p^*+1)}| \leq (c-2)k$, i.e., $\sum_{i=1}^{p^*} |\Theta^{(i)}| \leq (c-2)k$. This means that $\sum_{i=1}^{p^*+1} |\Theta^{(i)}| > (c-2)k$ which together with $|\Theta^{(i)}| \leq k$ implies that $(p^*+1)k > (c-2)k$, i.e., $p^* \geq c-2$. Thus p^* would be a large number when $c \gg 2$ which is guaranteed by $t^{\max} \gg 2k$.

We now point out that for large-scale compressed sensing problems, the DOMP algorithm can exactly reconstruct the support of x_S in a finite number of iterations as long as $\|Ax_{\overline{S}} + \nu\|_2$ is small enough. We say that the vector x is *nondegenerated k-compressible* if it has at least k nonzero components and $\|x_{\overline{S}}\|_2$ is sufficiently small, where $S = L_k(x)$. For such a vector, $x_i \neq 0$ for all $i \in S$. Clearly, any k-sparse vector with k nonzeros must be nondegenerated k-compressible.

Theorem 4.9: Assume that the data $x \in \mathbb{R}^n$ is nondegenerated k-compressible. Let $y := Ax + \nu$ be the slightly inaccurate measurements of x in the sense that the errors ν are sufficiently small. Let $\gamma \in (0, 1]$ be a given constant in DOMP. Suppose that Assumption 4.8 holds and that A has the RIP of order ck with $t^{\max} \geq ck \gg 2k$ and δ_{ck} satisfies (19). Then when $p^* = \max\{p : |S^{(p+1)}| \leq (c-2)k\}$ is large enough, there must exist $\widehat{p}$ such that for any p with $\widehat{p} \leq p \leq p^*$, the iterate $x^{(p)}$ generated by DOMP satisfies that

$$L_k(x^{(p)}) = L_k(x).$$

Proof: (i) Under (19), Theorem 4.2 claims that

$$\|x^{(p)} - x_S\|_2 \leq \beta^{p-1} \left(\frac{\varrho}{\beta}\right)^k \|x^0 - x_S\|_2 + \tau \|\nu'\|, \tag{57}$$

where $S = L_k(x)$ and the constants (β, ϱ, τ) are given in Theorem 4.2. By assumption, the right-hand side of the above bound is sufficiently small when p is large enough. Note that $\|x^{(p)} - x_S\|_2^2 = \|(x^{(p)})_S - x_S\|_2^2 + \|(x^{(p)})_{\overline{S}}\|_2^2$. It follows from (57) that both terms $\|(x^{(p)})_S - x_S\|_2$ and $\|(x^{(p)})_{\overline{S}}\|_2$ are sufficiently small when p is large enough, which is ensured under the assumption of the theorem. Since x is 'nondegenerated k-compressible,' this implies that the largest k magnitudes of $x^{(p)}$ are concentrated on S and sufficiently close to x_S. This means $L_k(x^{(p)}) = S$. Thus after performing enough iterations, the support of x_S is correctly reconstructed by DOMP. ■

The error bound established in this paper measures how close the iterates generated by the algorithm to the target data in terms of δ_{ck}, the number of iterations p, and the measurement error $\|\nu\|_2$. The bound (19) indicates that as p increases, the reconstruction error decays provided that $p \leq p^*$. In the setting of large-scale compressed sensing, Theorem 4.9 further claims that the exact reconstruction of the sparse data can be also possible under the assumption of the theorem.

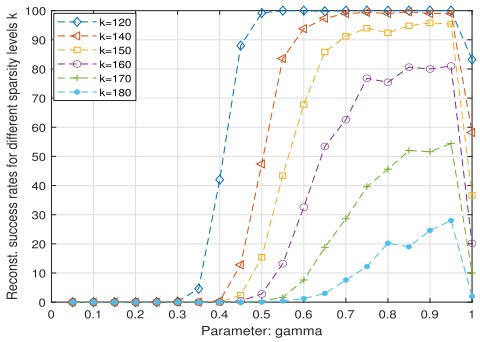
Remark 4.10: The EDOMP algorithm shares the same steps of DOMP when $|S^{(p)}| \leq k$, and performs hard thresholding to restore the k-sparsity of iterates when $|S^{(p)}| > k$. The iterate $x^{(p+1)}$, generated by EDOMP when $|S^{(p)}| > k$, is very different from the one generated by DOMP. Clearly, the relations $S^{(p)} \subseteq S^{(p+1)}$ and $|S^{(p)}| = \sum_{i=1}^{p} |\Theta^{(i)}|$ in DOMP are lost in EDOMP. Thus the analysis of DOMP presented in this section would not be generalized straightforwardly to EDOMP. While we had attempted to show the reconstruction error bound for EDOMP in the earlier draft of this report, the argument therein for EDOMP remains incomplete. It is still not completely clear at the moment about the error bound for EDOMP. This is an interesting and worthwhile future work.

V. NUMERICAL EXPERIMENTS

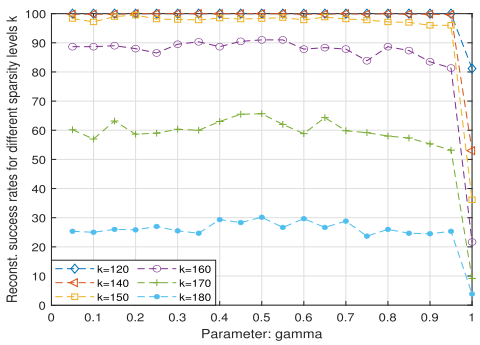
The performance of the proposed algorithms is clearly related to three main factors: (a) the choice of parameter γ, (b) the relative sparsity level k/m or k/n, and (c) the total number of iterations executed. We first carry out experiments to demonstrate how these factors might affect the performance of the proposed algorithms and thus to suggest the suitable choice of γ as well as the number of iterations to perform. All sparse vectors and measurement matrices in experiments are randomly generated. The entries of sparse vectors and matrices are assumed to be independent and identically distributed and follow the standard normal distribution, and the positions of nonzero entries of sparse vectors are uniformly distributed. All iterative algorithms in our experiments use $x^{(0)} = 0$ as the initial point. Unless otherwise stated, we adopt the following reconstruction criterion:

$$\|x^{(p)} - x\|_2 / \|x\|_2 \leq 10^{-5}, \tag{58}$$

where $x^{(p)}$ is the iterate produced by the algorithm and x is the target data to reconstruct. If $x^{(p)}$ satisfies (58), we say that the algorithm succeeds in reconstructing x.



(a) DOMP: Performance vs γ



(b) EDOMP: Performance vs γ

FIGURE 1. Comparison of performances of DOMP and EDOMP with different choices of γ.

A. EFFECT OF THE PARAMETER γ

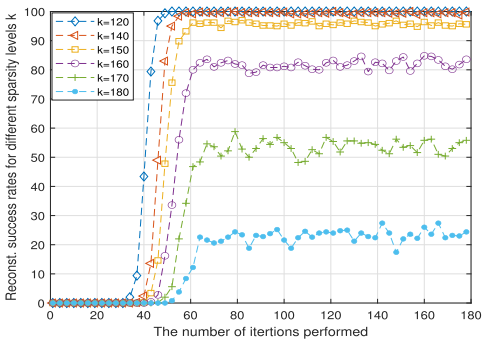
To illustrate how the value of γ affects the reconstruction ability of the proposed algorithms, we compare the success rates of the proposed algorithms in k-sparse data reconstruction using the following 20 different values: $\gamma = t/20$, where $t = 1, 2, \ldots, 20$. The size of matrix is 500×2000, and the sparsity levels $k = 120, 140, 150, 160, 170$ and 180 of the sparse vectors $x \in \mathbb{R}^{2000}$ are examined in this experiment. For every given sparsity level k, the success rates for k-sparse data reconstruction via DOMP and EDOMP with each given γ are obtained respectively by attempting 500 random pairs (A, x) with accurate measurements $y := Ax$. The algorithms are performed a total of k iterations, where k is the sparsity level of the vector to reconstruct. The results are given in Fig. 1. Intuitively, a small value of γ allows more indices to be added to $\Theta^{(p)}$ at every iteration, but it may also increase the chance for an incorrect index being added to Θ^p, and thus a small γ may lower the success rates of DOMP for data reconstruction as shown in Fig. 1(a). However, as shown in Fig. 1(b), it seems that the performance of EDOMP is less sensitive to the change of γ, compared with DOMP. On the other side, the results in Fig. 1 indicate that the success rates of the proposed algorithms with $\gamma = 1$ are not the highest ones. This does indicate that the algorithm with $\gamma = 1$ (including OMP) are not working to their potentials to identify the support of the target data. One can also see from Fig. 1(a) that the sparser the target data, the wider range of values for γ can be used in DOMP. The numerical experiments indicate that the values of γ situated between 0.7 and 1 are generally good for DOMP, while EDOMP admits more freedom in the choice of γ. As a result, we may use $\gamma = 0.9$ as a default choice for two algorithms in the remaining experiments.

B. EFFECT OF THE NUMBER OF ITERATIONS

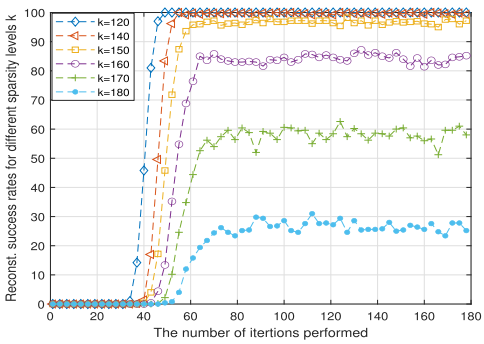
The main feature of OMP is that the support of the target data is gradually identified as the iteration proceeds. This means that the support of the target data (which is usually unknown beforehand) cannot be fully obtained if the number of iterations performed is too low. Similarly, the proposed algorithms need to perform enough iterations before being terminated. However, after the necessary iterations have been performed, can any further iterations continue to remarkably improve the performance of the proposed algorithms? To find a possible answer, it is useful to carry out an experiment to test the performance of algorithms against the number of iterations executed. According to our theory, under the RIP assumption, the DOMP and EDOMP algorithms keep finding the correct support of the target data provided that the performed iterations are lower than p^*, the largest integer number p such that $|S^{(p+1)}| \leq (c-2)k$ where $c > 2$ is the constant associated with the order of RIP. But when iterations go beyond p^*, there would be no guarantee for the algorithms to continue to find the correct support of x_S. With the same random matrices and sparse vectors used in first experiment, we compare the performance of DOMP and EDOMP when they are allowed to perform different numbers of iterations: $p = 1 + 3j$ where $j = 0, 1, \ldots, 59$. For every given sparsity level $k = 120, 140, 150, 160, 170, 180$ and the prescribed number for iterations, the success rates for data reconstruction via the two algorithms with $\gamma = 0.9$ are obtained by 500 random tries of (A, x) with accurate measurements. The results are given in Fig. 2. As expected, the algorithms need to perform necessary iterations in order to possibly reconstruct the data. The figure indicates that with the increment of iterations, the success frequencies of the two algorithms for data reconstruction increase to some highest point, after that the performance of algorithms would generally not be improved no matter how many further iterations are performed. From Fig. 2, it can be easily seen that the highest point of the reconstruction curve appears after a certain number of iterations that is much lower than the sparsity level k of the vector. This means unlike the classic OMP, the DOMP and EDOMP algorithms do not need to perform at least k iterations in order to reconstruct a k-sparse signal. Thus we carry out further simulations to examine the average number of iterations as well as the average runtime needed for the proposed algorithms to reconstruct sparse data.

C. AVERAGE NUMBER OF ITERATIONS AND AVERAGE RUNTIME

Experiments were performed to demonstrate the average number of iterations and runtime needed for the proposed algorithms to achieve the reconstruction criterion (58). The random matrices in 10 different sizes ranging from $(m, n) = (200, 1000)$ to $(2000, 10000)$ are used in this experiments.

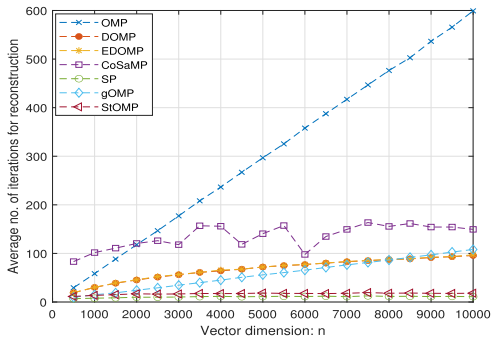


(a) DOMP: Performance vs no. of iter.

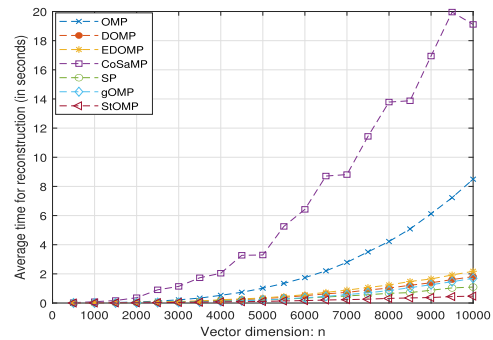


(b) EDOMP: Performance vs no. of iter.

FIGURE 2. Comparison of performances of DOMP and EDOMP with different no. of iterations.



(a) Average no. of iterations



(b) Average runtime

FIGURE 3. Comparison of average no. of iterations and runtime required by OMP, DOMP, EDOMP, CoSaMP, SP, gOMP and StOMP.

Specifically, $n = 5$ m and $m = 200j$, where $j = 1, \ldots, 10$. For every $m \times n$ matrix, the sparsity level of $x \in \mathbb{R}^n$ is set as $k = 0.3$ m. The OMP algorithm always performs at least k iterations in order to possibly reconstruct the k-sparse vector. Thus the total number of iterations required by OMP to reconstruct the k-sparse vectors ($k = 0.3$ m in this experiment) is increasing linearly with respect to the size of the matrices, as shown in Fig. 3(a). However, Fig. 3(a) indicates that the average numbers of iterations required by DOMP and EDOMP to meet the reconstruction criterion (58) are slowly increasing to the size of problems, remarkably lower than that of CoSaMP and OMP, and very comparable to gOMP ($N = 5$), but slightly more than that of SP and StOMP (with threshold parameter $t_s = 3$).

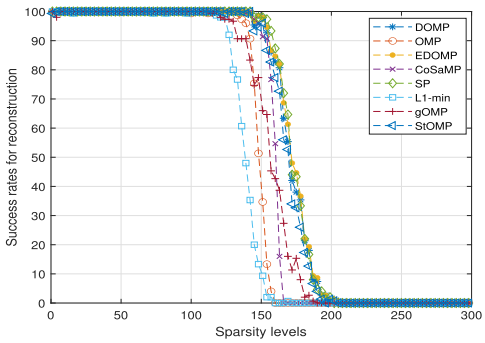
The average runtime for these algorithms to achieve the same criterion (58) is summarized in Fig. 3(b). It can be seen that the proposed algorithms require much less computational time than OMP and CoSaMP to achieve the same reconstruction criterion. The proposed algorithms take very similar amount of time as SP, StOMP and gOMP. As the problem size goes up, the runtime of the proposed algorithms increases very slowly in a linear-like manner, while the runtime for OMP and CoSaMP increases dramatically with respect to the size of problems. The simulation does indicate that the total iterations as well as the runtime required by DOMP and EDOMP to reconstruct the random sparse data can be remarkably lower than that of OMP and CoSaMP.

D. COMPARISON TO SOME EXISTING ALGORITHMS

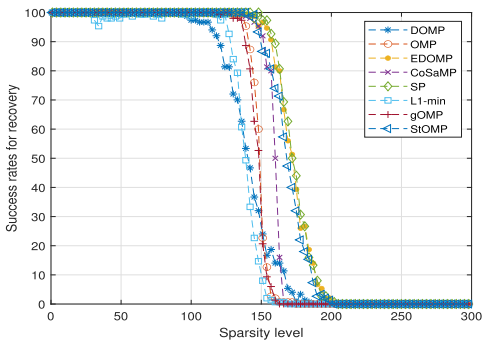
Finally, we compare the reconstruction success rates of the proposed algorithms and a few existing methods including OMP, SP CoSaMP, gOMP, StOMP and ℓ_1-minimization. The parameter $\gamma = 0.9$ is still set in DOMP and EDOMP which are performed the same number of iterations as OMP and gOMP, which is set to be the sparsity level k of the target signal. In our experiments, $N = 5$ is set in gOMP, and the threshold parameter $t_s = 3$ is set in StOMP. CoSaMP and SP are allowed to perform a total of 500 iterations for every problem instance in this experiment. For accurate measurements, all algorithms use the stopping criterion (58). For noisy measurements, however, we adopt the following criterion:

$$\|x^{(p)} - x\|_2 / \|x\|_2 \leq 10^{-3}.$$

In this comparison, the size of matrices is 500×2000, and the sparsity level k of the random data $x \in \mathbb{R}^{2000}$ is ranged from 1 to 300 with stepsize 3, i.e., $k = 1, 4, 7, \ldots, 299$. For each given sparsity level, 200 random pairs of (A, x) were realized and used to estimate the success rates of the algorithms for data reconstruction. All matrices A are normalized. The accurate measurements are taken as $y := Ax$ and inaccurate measurements are taken as $\widetilde{y} := Ax + 0.001\,h$, where h is a normalized Gaussian random vector. When an algorithm terminates, if $x^{(p)}$ satisfies the reconstruction criterion, a 'success' is counted for the algorithm. The success rates of the proposed algorithms using accurate measurements are shown



(a) Accurate measurements



(b) Inaccurate measurements

FIGURE 4. Comparison of reconstruction success rates of algorithms in noiseless and noisy cases.

in Fig. 4(a), and the results with inaccurate measurements are given in Fig. 4(b). One can see that in noiseless setting, DOMP and EDOMP outperform OMP, CoSaMP, gOMP and ℓ_1-minimization, and they perform very comparably to SP and StOMP. The similar results for EDOMP can be observed when the measurements are slightly inaccurate, while DOMP and gOMP are clearly subject to the effect of noises. In summary, EDOMP is very competitive to the existing mainstream algorithms for sparse data reconstruction. It is fast, stable, robust and very easy to implement like CoSaMP and SP.

VI. CONCLUSION

The reconstruction error bound for the proposed DOMP algorithm is established in this article under the RIP assumption. Finite convergence of the algorithm is also discussed in large-scale compressed sensing scenarios. Simulations indicate that the proposed methods (especially, EDOMP) are more efficient than OMP in sense that they may take much less average computational times and iterations to reconstruct the sparse data, and they can compete against several state-of-art methods in this field. Distinguished from the existing OMP-type methods, the main feature of the proposed algorithms is that they can dynamically and efficiently exploit the gradient information of the error metric at every iteration to recognize the support of the target data.

REFERENCES

[1] M. Elad, *Sparse and Redundant Representation: From Theory to Applications in Signal and Image Processing*. New York, NY, USA: Springer, 2010.
[2] Y. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*, Cambridge, U.K.: Cambridge Univ. Press, 2012.
[3] Z. Gao, L. Dai, S. Han, C.-L. I., Z. Wang, and L. Hanzo, "Compressive sensing techniques for next-generation wireless communications," *IEEE Wireless Commun.*, vol. 25, no. 3, pp. 144–153, Jun. 2018.
[4] H. Boche, R. Calderbank, G. Kutyniok, and J. Vybiral, *Compressed Sensing and Its Applications*. New York, NY, USA: Springer, 2019.
[5] M. Leinomen, M. Codreanu, and G. Giannakis, "Compressed sensing with applications in wireless networks," *Found. Trends Signal Process.*, vol. 13, no. 1–2, pp. 1–282, 2019.
[6] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
[7] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[8] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Basel, Switzerland: Birkhäuser, Springer, 2013.
[9] A. De Maio, Y. Eldar, and A. Haimovich, *Compressed Sensing in Radar Signal Processing*. Cambridge, U.K.: Cambridge Univ. Press, 2019.
[10] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.
[11] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
[12] E. J. Candès, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, pp. 877–905, 2008.
[13] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.
[14] A. Beck and Y. C. Eldar, "Sparsity constrained nonlinear optimization: Optimality conditions and algorithms," *SIAM J. Optim.*, vol. 23, no. 3, pp. 1480–1509, 2013.
[15] Y.-B. Zhao, *Sparse Optimization Theory and Methods*. Boca Raton, FL, USA: CRC Press, 2018.
[16] R. Khanna and A. Kyrillidis, "IHT dies hard: Provable accelerated iterative hard thresholding," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2018, pp. 188–198.
[17] Y.-B. Zhao, "Optimal k-thresholding algirithms for sparse optimization problems," *SIAM J. Optim.*, vol. 30, no. 1, pp. 31–55, 2020.
[18] Y.-B. Zhao and Z.-Q. Luo, "Analysis of optimal thresholding algorithms for compressed sensing," *Signal Process.*, vol. 187, 2021, Art. no. 108148.
[19] N. Meng and Y. B. Zhao, "Newton-step-based hard thresholding algorithms for sparse signal recovery," *IEEE Trans. Signal Process.*, vol. 68, pp. 6594–6606, 2020.
[20] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. IEEE 27th Annu. Asilomar Conf. Signals, Syst., Comput.*, 1993, pp. 40–44.
[21] G. Davis, S. Mallat, and Z. Zhang, "Adaptive time-frequency decompositions," *Opt. Eng.*, vol. 33, no. 7, pp. 2183–2191, 1994.
[22] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.
[23] G. Davis, S. Mallat, and M. Avellaneda, "Greedy adaptive approximation," *J. Constructive Approximation*, vol. 13, no. 1, pp. 57–98, 1997.
[24] R. A. DeVore, "Nonlinear approximation," *Acta Numerica*, vol. 7, pp. 51–150, 1998.
[25] V. Temlyakov, "Nonlinear methods of approximation," *Found. Comput. Math.*, vol. 3, no. 1, pp. 33–107, 2003.
[26] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.
[27] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.
[28] S. Kwon, J. Wang, and B. Shim, "Multipath matching pursuit," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2986–3001, May 2014.

[29] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1094–1121, Feb. 2012.
[30] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Found. Comput. Math.*, vol. 9, no. 3, pp. 317–334, 2009.
[31] D. Needell and R. Vershynin, "Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 310–316, Apr. 2010.
[32] J. Wang, S. Kwon, and B. Shim, "Generalized orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6202–6216, Dec. 2012.
[33] V. Temlyakov, "Weak greedy algorithms," *Adv. Comput. Math.*, vol. 12, no. 2–3, pp. 213–227, 2000.
[34] W. Cao, W. Xi, H. Yao, and W. Chen, "Stage-wise weak orthogonal matching pursuit reconstruction of power system disturbance signal," in *Proc. IEEE 5th Asia Conf. Power Elect. Eng.*, 2020, pp. 1698–1702.
[35] J. Shen and S. Mousavi, "Exact support and vector recovery of constrained sparse vectors via constrained matching pursuit," Mar. 2020, *arXiv:1903.07236v3*.
[36] A. C. Gilbert, M. Muthukrishnan, and M. J. Strauss, "Approximation of functions over redundant dictionaries using coherence," in *Proc. 14th Annu. ACM-SIAM Symp. Discrete Algorithms*, 2003, pp. 243–252.
[37] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
[38] J. A. Tropp, "Just relax: Convex programming methods for subset selection and sparse approximation," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.
[39] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
[40] M. A. Davenport and M. B. Wakin, "Analysis of orthogonal matching pursuit using the restricted isometry property," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4395–4401, Sep. 2010.
[41] S. Huang and J. Zhu, "Recovery of sparse signals using OMP and its variants: Convergence analysis based on RIP," *Inverse Problems*, vol. 27, no. 3, 2011, Art. no. 035003.
[42] Q. Mo and Y. Shen, "A remark on the restricted isometry property in orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3654–3656, Jun. 2012.
[43] J. Wang and B. Shim, "On the recovery limit of sparse signals using orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4973–4976, Sep. 2012.
[44] L.-H. Chang and J.-Y. Wu, "An improved RIP-based performance guarantee for sparse signal recovery via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5702–5715, Sep. 2014.
[45] Q. Mo, "A sharp restricted isometry constant bound of orthogonal matching pursuit," Jan. 2015, *arXiv:1501.01708*.
[46] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4680–4688, Jul. 2011.
[47] R. Wu, W. Huang, and D.-R. Chen, "The exact support recovery of sparse signals with noise via orthogonal matching pursuit," *IEEE Signal Process. Lett.*, vol. 20, no. 4, pp. 403–406, Apr. 2013.
[48] J. Wen, Z. Zhou, D. Li, and X. Tang, "A novel sufficient condition for generalized orthogonal matching pursuit," *IEEE Commun. Lett.*, vol. 21, no. 4, pp. 805–808, Apr. 2017.
[49] T. Zhang, "Sparse recovery with orthogonal matching pursuit under RIP," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 6215–6221, Sep. 2011.
[50] L.-H. Chang and J.-Y. Wu, "An improved RIP-based performance guarantee for sparse signal reconstruction via subspace pursuit," in *Proc. IEEE 8th Sensor Array Multichannel Signal Process. Workshop*, 2014, pp. 405–408.
[51] C.-B. Song, S.-T. Xia, and X.-J. Liu, "Improved analysis for subspace pursuit algorithm in terms of restricted isometry constant," *IEEE Signal Process. Lett.*, vol. 21, no. 11, pp. 1365–1369, Nov. 2014.
[52] Y.-B. Zhao and Z.-Q. Luo, "Improved RIP-based bounds for guaranteed performance of two compressed sensing algorithms," *Sci. China Math.*, vol. 66, no. 5, pp. 1123–1140, May 2023.
[53] E. Liu and V. N. Temlyakov, "The orthogonal super greedy algorithm and applications in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2040–2047, Apr. 2012.
[54] R. Maleh, "Improved RIP analyis of orthogonal matching pursuit," Feb. 2011, *arXiv:1102.4311v1*.
[55] S. Satpathi, R. L. Das, and M. Chakraborty, "Improving the bound on the RIP constant in generalized orthogonal matching pursuit," *IEEE Signal Process. Lett.*, vol. 20, no. 11, pp. 1074–1077, Nov. 2013.
[56] Y. Shen, B. Li, W. Pan, and J. Li, "Analysis of orthogonal matching pursuit using restricted isometry constant," *Electron. Lett.*, vol. 50, no. 14, pp. 1020–1022, 2014.
[57] B. Li, Y. Shen, Z. Wu, and J. Li, "Sufficient conditions for pursuit in noisy case," *Signal Process.*, vol. 108, pp. 111–123, 2015.
[58] J. Wang, "Support recover with orthogonal matching pursuit in the presence of noise," *IEEE Trans. Signal Process.*, vol. 63, no. 21, pp. 5868–5877, Nov. 2015.
[59] Y.-B. Zhao, "RSP-based analysis for sparsest and least ℓ_1-norm solutions to underdetermined linear systems," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5777–5788, Nov. 2013.

YUN-BIN ZHAO (Member, IEEE) received the Ph.D. degree in operations research and control theory from the Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing, China, in 1998. He is currently a Professor of mathematical optimization with the Shenzhen Research Institute of Big Data (SRIBD), Chinese University of Hong Kong, Hong Kong. From 2003 to 2007, he was an Associate Professor with the Academy of Mathematics and Systems Science, Chinese Academy of Sciences. From 2007 to 2012, he was a Lecturer and from 2012 to 2020, he was an Associate Professor with the School of Mathematics, University of Birmingham, U.K. He joined to SRIBD since March 2020. He has authored/coauthored about 70 refereed journal papers, and he is the author of the book *Sparse Optimization Theory and Methods*, CRC Press, Boca Raton, FL, 2018. His research interests include the operations research, computational optimization, numerical analysis, compressed sensing, and signal processing. His latest research is to develop theory and efficient computational methods for sparse optimization problems and their applications to signal processing.

ZHI-QUAN LUO (Fellow, IEEE) received the B.S. degree in applied mathematics from Peking University, Beijing, China, and the Ph.D. degree in operations research from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1989. From 1989 to 2003, he held a Faculty position with the Department of Electronics and Communication Engineering, McMaster University, Hamilton, ON, Canada. From 2001 to 2003, he held a tier-1 Canada Research Chair in information processing. After that, he was a Full Professor with the Department of Electronics and Communication Engineering, University of Minnesota, Minneapolis, MN, USA, and held an endowed ADC Chair in digital technology. He is currently the Vice President (Academic) of The Chinese University of Hong Kong, Shenzhen, China, and Director of Shenzhen Research Institute of Big Data (SRIBD). He has authored or coauthored more than 300 refereed papers, books, and special issues. His main research interests include mathematical issues in information sciences, with particular focus on the design, analysis and applications of large-scale optimization algorithms. He is a Fellow of SIAM and selected to the Royal Society of Canada. He was the recipient of the four best paper awards from the IEEE Signal Processing Society, One Best Paper Award from EUSIPCO, Farkas Prize from INFORMS and prize of Paul Y. Tseng Memorial Lectureship in continuous optimization, and some best paper awards from international conferences. He has served as an Associate Editor for many internationally recognized journals and the Editor-in-Chief of the IEEE TRANSACTIONS ON SIGNAL PROCESSING.