axioms

Article

Efficiency of Orthogonal Matching Pursuit for Group Sparse Recovery

Chunfang Shao [1], Xiujie Wei [2], Peixin Ye [3] and Shuo Xing [3,*]

[1] College of Science, North China University of Science and Technology, Tangshan 063210, China; scfaku@163.com
[2] School of Sciences, Tianjin Chengjian University, Tianjin 300384, China; weixiujie1986@163.com
[3] School of Mathematical Sciences and LPMC, Nankai University, Tianjin 300071, China; yepx@nankai.edu.cn
* Correspondence: xingshuo9853@gmail.com

Abstract: We propose the Group Orthogonal Matching Pursuit (GOMP) algorithm to recover group sparse signals from noisy measurements. Under the group restricted isometry property (GRIP), we prove the instance optimality of the GOMP algorithm for any decomposable approximation norm. Meanwhile, we show the robustness of the GOMP under the measurement error. Compared with the P-norm minimization approach, the GOMP is easier to implement, and the assumption of γ-decomposability is not required. The simulation results show that the GOMP is very efficient for group sparse signal recovery and significantly outperforms Basis Pursuit in both scalability and solution quality.

Keywords: compressed sensing; group orthogonal matching pursuit; group sparse; group restricted isometry property; instance optimality; robustness; scalability

MSC: 41A65; 41A25; 94A15; 94A08; 94A12; 68P30

check for updates

Citation: Shao, C.; Wei, X.; Ye, P.; Xing, S. Efficiency of Orthogonal Matching Pursuit for Group Sparse Recovery. *Axioms* **2023**, *12*, 389. https://doi.org/10.3390/axioms12040389

Academic Editor: Clemente Cesarano

Received: 10 March 2023
Revised: 12 April 2023
Accepted: 14 April 2023
Published: 17 April 2023

1. Introduction

The problem of recovering sparse signals from incomplete measurements has been studied extensively in the field of compressed sensing. Many important results have been obtained, see refs. [1–5]. In this article, we consider the problem of recovering group sparse signals. We first recall some basic notions and results.

Assume that $x \in \mathbb{R}^n$ is an unknown signal, and the information we gather about $x \in \mathbb{R}^n$ can be described by

$$y = \Phi x + e, \tag{1}$$

where $\Phi \in \mathbb{R}^{m\times n}$ is the encoder matrix and $y \in \mathbb{R}^m$ is the information vector with a noise level $\|e\|_2 \le \epsilon$. To recover x from y, we use a decoder Δ, which is a mapping from $\mathbb{R}^m$ to $\mathbb{R}^n$, and denote

$$x^* := \Delta(y) = \Delta(\Phi x + e)$$

as our approximation of x.

The l_1-norm minimization, also known as Basis Pursuit [2], is one of the most popular decoder maps. It is defined as follows:

$$\Delta_{\mathrm{BP}}(y) := \arg\min_{z\in\mathbb{R}^n} \|z\|_1, \text{ s.t. } \|y - \Phi z\|_2 \le \epsilon. \tag{2}$$

Candès and Tao in Ref. [6] introduced the restricted isometry property (RIP) of a measurement matrix as follows:

A matrix Φ is said to satisfy the RIP of order k if there exists a constant $\delta \in (0,1)$ such that, for all k-sparse signals x,

$$(1-\delta)\|x\|_2^2 \le \|\Phi x\|_2 \le (1+\delta)\|x\|_2^2.$$

In particular, the minimum of all δ satisfying the above inequality is defined as the isometry constant δ_k.

Based on the RIP of the encoder matrix Φ, it has been shown that the solution to (2) recovers x exactly provided that x is sufficiently sparse [7]. We denote by Σ_k the set of all k-sparse signals, i.e.,

$$\Sigma_k := \{z \in \mathbb{R}^n : |\mathrm{supp}(z)| \leq k\},$$

where $\mathrm{supp}(z)$ is the set of i for which $z_i \neq 0$ and $|A|$ is the cardinality of the set A. Let $\sigma_k(x, \|\cdot\|_1)$ be the sparsity index of x from Σ_k, which is defined by

$$\sigma_k(x, \|\cdot\|_1) := \inf_{z \in \Sigma_k} \|x - z\|_1.$$

In Ref. [7], the authors obtained the estimation of the residual error

$$\|\Delta_{\mathrm{BP}}(y) - x\|_2 \leq \frac{C_0}{\sqrt{k}} \sigma_k(x, \|\cdot\|_1) + C_1 \epsilon, \tag{3}$$

where C_0, C_1 are constants depending only on the encoder matrix Φ but not on x or e. In many current papers on sparse recovery, the l_1-norm objective function in (2) has been changed to several other norms.

At about the same time, the research community began to propose that the number of nonzero components of a signal might not be the only reasonable measure of the sparsity of a signal. Alternate notions under the broad umbrella of "group sparsity" and "group sparse recovery" began to appear. For more detail, one can see Ref. [8]. It has been shown that group sparse signals are used widely in EEG [9], wavelet image analysis [10], gene analysis [11], multi-channel image analysis [12] and other fields.

We next recall a result on the recovering of the group sparse signals from Ref. [13]. In that paper, Ahsen and Vidyasagar estimated x from y by solving the following optimization problem:

$$\Delta_P(y) := \arg\min_{z \in \mathbb{R}^n} \|z\|_P, \text{ s.t. } \|y - \Phi z\|_2 \leq \epsilon, \tag{4}$$

where $\|\cdot\|_P$ is the penalty norm. We begin with recalling some definitions from [13]. We denote the set $\{1, 2, \ldots, n\}$ by the symbol $[n]$. The group structure $\mathcal{G} = \{G_1, G_2, \ldots, G_g\}$ is a partition of $[n]$. Here, we suppose that $|G_i| \leq k$ for all $1 \leq i \leq g$. A signal $x \in \mathbb{R}^n$ is said to be group k-sparse if its support set $\mathrm{supp}(x)$ is contained in a group k-sparse set, which is defined as follows:

Definition 1. *A subset $\Lambda \subseteq [n]$ is said to be group k-sparse if $\Lambda = G_S := \cup_{i \in S} G_i$ for some $S \subseteq \{1, 2, \ldots, g\}$ and $|\Lambda| \leq k$.*

We use GkS to denote the set of all group k-sparse subsets of $[n]$.

Definition 2. *We say a norm $\|\cdot\|$ on $\mathbb{R}^n$ is decomposable if $x, y \in \mathbb{R}^n$, satisfy $supp(x) \subseteq G_S$, $supp(y) \subseteq G_T$, with S, T being disjoint subsets of $\{1, 2, \ldots, g\}$, then one has $\|x + y\| = \|x\| + \|y\|$.*

It is clear that l_1-norm is decomposable. More examples of decomposable norms can be found in Ref. [13].

In the processing of recovering an unknown signal $x \in \mathbb{R}^n$, the approximation norm on $\mathbb{R}^n$ will be denoted by $\|\cdot\|_A$. We assume that the approximation norm is decomposable.

Definition 3. *If there exists $\gamma \in (0, 1]$ such that for any $x, y \in \mathbb{R}^n$ with $supp(x) \subseteq G_S$, $supp(y) \subseteq G_T$, and S, T are disjoint subsets of $\{1, 2, \ldots, g\}$, it is true that $\|x + y\| \geq \|x\| + \gamma\|y\|$, we say the norm $\|\cdot\|$ is γ-decomposable.*

Note that when $\gamma = 1$, γ-decomposability coincides with decomposability.

For any index set $A \subseteq [n]$, let x_A denote the part of the signal x consisting of coordinates falling on A. The sparsity index and the optimal decomposition of a signal in $\mathbb{R}^n$ are defined respectively as follows:

Definition 4. *The group k-sparse index of a signal $x \in \mathbb{R}^n$ with respect to the norm $\|\cdot\|$ and the group structure $\mathcal{G}$ is defined by*

$$\sigma_{k,\mathcal{G}}(x, \|\cdot\|) := \min_{\Lambda \in GkS} \|x - x_\Lambda\| = \min_{\Lambda \in GkS} \|x_{\Lambda^c}\|.$$

For $x \in \mathbb{R}^n$ and a norm $\|\cdot\|$ on $\mathbb{R}^n$, if there exist $\Lambda_i \in \text{GkS}$ for $i = 0, 1, 2, \ldots, s$, such that

$$\|x_{\Lambda_0^c}\| = \|x - x_{\Lambda_0}\| = \min_{\Lambda \in \text{GkS}} \|x - x_\Lambda\| = \sigma_{k,\mathcal{G}}(x)$$

and

$$\|x_{\Lambda_i^c}\| = \min_{\Lambda \in \text{GkS}} \|x - \sum_{j=0}^{i-1} x_{\Lambda_j} - x_\Lambda\|, \quad i = 1, 2, \ldots, s,$$

then we call $\{x_{\Lambda_0}, x_{\Lambda_1}, \ldots, x_{\Lambda_s}\}$ an optimal group k-sparse decomposition of x.

Definition 5. *A matrix $\Phi \in \mathbb{R}^{m \times n}$ is said to satisfy the group restricted isometry property (GRIP) of order k if there exists a constant $\delta_k \in (0, 1)$, such that*

$$1 - \delta_k \leq \min_{\Lambda \in GkS} \min_{supp(z) \subseteq \Lambda} \frac{\|\Phi z\|_2^2}{\|z\|_2^2} \leq \max_{\Lambda \in GkS} \max_{supp(z) \subseteq \Lambda} \frac{\|\Phi z\|_2^2}{\|z\|_2^2} \leq 1 + \delta_k.$$

Then, we are ready to state the result of Ahsen and Vidyasagar. We need the following constants:

$$a := \min_{\Lambda \in \text{GkS}} \min_{x_\Lambda \neq 0} \frac{\|x_\Lambda\|_P}{\|x_\Lambda\|_A}, \; b := \max_{\Lambda \in \text{GkS}} \max_{x_\Lambda \neq 0} \frac{\|x_\Lambda\|_P}{\|x_\Lambda\|_A}, \; f := \sqrt{k}$$

and

$$c := \min_{\Lambda \in \text{GkS}} \min_{x_\Lambda \neq 0} \frac{\|x_\Lambda\|_A}{\|x_\Lambda\|_2}, \; d := \max_{\Lambda \in \text{GkS}} \max_{x_\Lambda \neq 0} \frac{\|x_\Lambda\|_A}{\|x_\Lambda\|_2}. \tag{5}$$

Theorem 1. *Suppose that*

1. *The norm $\|\cdot\|_A$ is decomposable.*
2. *The norm $\|\cdot\|_P$ is γ-decomposable for some $\gamma \in (0, 1]$.*
3. *The matrix Φ satisfies GRIP of order 2k with constant δ_{2k}.*
4. *Suppose the "compressibility condition"*

$$\delta_{2k} < \frac{fa\gamma/bd}{\sqrt{2} + fa\gamma/bd}$$

holds, then

$$\|\Delta_P(\Phi x + e) - x\|_2 \leq D_1 \sigma_{k,\mathcal{G}}(x)_A + D_2 \epsilon.$$

Furthermore,

$$\|\Delta_P(\Phi x + e) - x\|_A \leq D_3 \sigma_{k,\mathcal{G}}(x)_A + D_4 \epsilon,$$

where

$$D_1 = \frac{b(1+\gamma)}{a\gamma f} \cdot \frac{1 + (\sqrt{2} - 1)\delta_{2k}}{1 - (1 + \sqrt{2}bd/a\gamma f)\delta_{2k}}, D_2 = 2(1 + bd/a\gamma f) \cdot \frac{\sqrt{1 + \delta_{2k}}}{1 - (1 + \sqrt{2}bd/a\gamma f)\delta_{2k}}$$

and

$$D_3 = \frac{b(1+\gamma)}{a\gamma} \cdot \frac{1 + (\sqrt{2}d/f - 1)\delta_{2k}}{1 - (1 + \sqrt{2}bd/a\gamma f)\delta_{2k}}, D_4 = 2(1 + bd/a\gamma) \cdot \frac{\sqrt{1 + \delta_{2k}}}{1 - (1 + \sqrt{2}bd/a\gamma f)\delta_{2k}}.$$

Recently, Ranjan and Vidyasagar presented sufficient conditions for (2) to achieve robust group sparse recovery by establishing a group robust null space property [14]. They derived the residual error bounds for the l_p-norm for $p \in [1,2]$.

In compressed sensing, an alternative decoder for (1) is Orthogonal Matching Pursuit (OMP), which was originally proposed by J. Tropp in [15]. It is defined as Algorithm 1:

Algorithm 1 Orthogonal Matching Pursuit (OMP)

Input: measurement matrix Φ, measurement vector y.
Initialization: $S^0 = \emptyset, x^0 = 0$.
Iteration: repeat until a stopping criterion is met at $n = \overline{n}$:

$$S^{n+1} = S^n \cup \{j_{n+1}\},\ j_{n+1} := \arg\max_{j\in[n]}\{|\Phi^*(y - \Phi x^n))_j|\},$$

$$x^{n+1} = \arg\min_{z\in\mathbb{R}^n}\{\|y - \Phi z\|_2,\ \mathrm{supp}(z) \subseteq S^{n+1}\}.$$

Output: the $\overline{n}$-sparse vector $x^{\#} = x^{\overline{n}}$.

The major advantage of the OMP algorithm is that it is easy to implement, see Refs. [16–21]. In [22], under the RIP, T. Zhang proved that OMP with k iterations can exactly recover k-sparse signals and the procedure is also stable in l_2 under measurement noise. Based on this result, Xu showed that the OMP can achieve instance optimality under the RIP, cf. [23]. In [24], Cohen, Dahmen and DeVore extended T. Zhang's result to the general context of k-term approximation from a dictionary in arbitrary Hilbert spaces. They have shown that OMP generates near-best k-term approximations under a similar RIP condition.

In this article, we will generalize the OMP to recover the group sparse signals. Assume that the signal x is group k-sparse with the group structure $\mathcal{G}$, then the number of non-zero groups of $x = (x_{G_1}, x_{G_2}, \ldots, x_{G_g})$ is no more than k. We define

$$\alpha := \max_{j\in[g]} |G_j|.$$

It is obvious that $\alpha \le k$.

The group support set of x with respect to $\mathcal{G}$ is defined as follows:

$$\mathrm{Gsupp}(x) := \{j \in [g] : x_{G_j} \neq 0\}.$$

Let $|\mathrm{Gsupp}(x)|$ be the number of distinct groups on which x is supported. For any $i \in \{1, 2, \ldots, g\}$, the subvector (submatrix) x_{G_i} (Φ_{G_i}) can also be written as $x[i]$ $(\Phi[i])$.

Now we start to propose the Group Orthogonal Matching Pursuit (GOMP) algorithm. Assume that the group structure $\mathcal{G}$ is given a prior. Let Ω be the initial feature set of the group sparse signal x with the maximum allowed sparsity M. The GOMP is defined as Algorithm 2:

Algorithm 2 Group Orthogonal Matching Pursuit ($\mathrm{GOMP}_M(y)$)

Input: encoding matrix Φ, the vector y, group structure $\{G_i\}_{i=1}^g$, a set $\Omega \subseteq \{1, 2, \ldots, g\}$, maximum allowed sparsity M.
Initialization: $x^0 = \arg\min_{\widetilde{x}:\mathrm{Gsupp}(\widetilde{x})\subseteq\Omega} \|y - \Phi\widetilde{x}\|_2$, $r^0 = y - \Phi x^0$, $l = 1$, $I_0 = \Omega$.
while $l \le k \le M$ **do**
  Set $i_l = \arg\max_i \|\Phi^*[i] r^{l-1}\|_2$.
  Identity $I_l = \{i_1, i_2, \ldots, i_l\} \cup \Omega$.
  Update $x^l = \arg\min_{\widetilde{x}:\mathrm{Gsupp}(\widetilde{x})\subseteq I_l} \|y - \Phi\widetilde{x}\|_2$, $r^l = y - \Phi x^l$, $l = l + 1$.
end while
Output: $\widehat{x}_{\mathrm{GOMP}} = x^k$.

The GOMP$_M(y)$ algorithm begins by initializing the residual as $r^0 = y$, and at the lth $(l \geq 1)$ iteration, we chose one group index which matched r^{l-1} best. However, the algorithm does not obtain a least-square minimization over the group index sets that have already been selected, but updates the residual and proceeds to the next iteration.

Then, we studied the efficiency of the GOMP. We show the instance optimality and robustness of the GOMP algorithm under the GRIP of the matrix Φ. We formulate them in the following theorem.

Theorem 2. *Suppose that the norm $\|\cdot\|_A$ is decomposable and $0 < \delta \leq 1$. If the GRIP condition $\delta_{k\alpha} + (1+\delta)\delta_{\beta k\alpha} \leq \delta$ holds with $\beta := \lceil 16 + 15\delta \rceil$, then for any signal x and any permutation e with $\|e\|_2 \leq \epsilon$, the solution $x^* := GOMP_{(\beta-1)k}(\Phi x + e)$ obeys*

$$\|x^* - x\|_2 \leq D_1 \sigma_{k,\mathcal{G}}(x)_A + D_2 \epsilon, \tag{6}$$

where

$$D_1 = \frac{1}{c}[(1+\delta)(1+\sqrt{11+20\delta})+1] \quad \textit{and} \quad D_2 = \sqrt{1+\delta}(1+\sqrt{11+20\delta}). \tag{7}$$

Furthermore,

$$\|x^* - x\|_A \leq D_3 \sigma_{k,\mathcal{G}}(x)_A + D_4 \epsilon, \tag{8}$$

where

$$D_3 = \frac{d}{c}\sqrt{2\beta\alpha - 1}(1+\delta)(1+\sqrt{11+20\delta}) + 1 \quad \textit{and} \quad D_4 = d\sqrt{(2\beta\alpha - 1)(1+\delta)}(1+\sqrt{11+20\delta}). \tag{9}$$

When $e = 0$, from inequalities (6) and (8), we have

$$\|x^* - x\|_2 \leq D_1 \sigma_{k,\mathcal{G}}(x)_A$$

and

$$\|x^* - x\|_A \leq D_3 \sigma_{k,\mathcal{G}}(x)_A.$$

This implies the instance optimality of the GOMP. In other words, from the viewpoint of the error, the GOMP is almost the best. The robustness of the GOMP means that the signal can be recovered stably for different dimensions, sparsity levels, numbers of measurements and noise scenarios (which can be observed from the experiments in Section 3). Compared with Basis Pursuit (BP), our proposed algorithm performs far better, and runs extremely faster, especially when it comes large-scale problems.

Based on the observation of Theorems 1 and 2, one can see that the assumption of γ-decomposability is not required for the GOMP. Instead, we developed some new techniques to analyze the error performance of the GOMP.

We remark that the results of Theorem 2 are rather general. They cover some important cases.

Firstly, if $n = \sum_{i=1}^{g} d_i$ and we take a special partition of $[n]$ as

$$[n] = [\underbrace{1, \ldots, d_1}_{G_1}, \underbrace{d_1+1, \ldots, d_1+d_2}_{G_2}, \ldots, \underbrace{n-d_g+1, \ldots, n}_{G_g}],$$

then the unknown sparse signals appear in a few blocks, i.e., the signals are block-sparse. It is obvious that our results include the case of the block-sparse signals. In practice, there are many forms of block-sparse signal, such as multi-band signal, DNA array, radar pulse signal and the multi-measurement vector problem and so on, see Refs. [25–32].

Secondly, applying Theorem 2 to the case of conventional sparsity, we obtain the following new results on the error estimates of the OMP.

Corollary 1. *(Conventional sparsity) Suppose that* Φ *satisfies the RIP condition* $\delta_k + (1+\delta)\delta_{\beta k} \leq \delta$ *with* $\beta := \lceil 16 + 15\delta \rceil$. *Then,*

$$\|OMP_{(\beta-1)k}(\Phi x + e) - x\|_2 \leq D_1 \sigma_k(x)_A + D_2 \epsilon,$$

and

$$\|OMP_{(\beta-1)k}(\Phi x + e) - x\|_A \leq D_3 \sigma_k(x)_A + D_4 \epsilon,$$

where the constants D_i *are defined in (1.6) and (1.8) for* $i = 1, 2, 3, 4$.

The remaining part of the paper is organized as follows. In Section 2, based on a corollary of Theorem 3, we prove Theorem 2, while the proof of Theorem 3 will be given in the Appendix B. In the Appendix A, we establish some preliminary results for the proof. In Section 3, we compare the Group Orthogonal Matching Pursuit (GOMP) with the Basis Pursuit (BP) by numerical experiments. In Section 4, we draw the conclusions of our study. Moreover, we mention some other competitive algorithms which we will study in the future.

2. Proof of Theorem 2

The proof of Theorem 2 is based on the estimation of the residual error of the GOMP algorithm. To establish such estimation, we use the restricted gradient optimal constant for group sparse signals.

Definition 6. *Given* $\overline{x} \in \mathbb{R}^n$ *and* $s > 0$, *we define the restricted gradient optimal constant* $\varepsilon_s(\overline{x})$ *as the smallest non-negative value such that*

$$|\langle 2\Phi\mu, \Phi\overline{x} - y\rangle| \leq \varepsilon_s(\overline{x})\|\mu\|_2$$

for all $\mu \in \mathbb{R}^n$ *with* $|Gsupp(\mu)| \leq s$.

Some estimates about $\varepsilon_s(\overline{x})$ have been given in Ref. [22], such as $\varepsilon_s(\overline{x}) \leq 2\sqrt{s}\|\Phi^*(\Phi\overline{x} - y)\|_\infty$, $\varepsilon_s(\overline{x}) \leq 2\|\Phi^*(\Phi\overline{x} - y)\|_2$, and so on.

The following theorem is our result of the residual error of the GOMP algorithm. Its proof is quite technical and is given in the Appendix B.

Theorem 3. *Let* $\overline{x} \in \mathbb{R}^n$ *and* $\overline{F} = Gsupp(\overline{x})$. *If there exists* s, *such that*

$$s \geq |\overline{F} \cup \Omega| + \lceil 4|\overline{F} \setminus \Omega| \frac{1+\delta_\alpha}{1-\delta_{s\alpha}} \ln \frac{20(1+\delta_{|\overline{F}\setminus\Omega|\alpha})}{1-\delta_{s\alpha}} \rceil, \tag{10}$$

then when $k \geq \lceil 4|\overline{F} \setminus \Omega| \frac{1+\delta_\alpha}{1-\delta_{s\alpha}} \ln \frac{20(1+\delta_{|\overline{F}\setminus\Omega|\alpha})}{1-\delta_{s\alpha}} \rceil$, *we have*

$$\|\Phi x^k - y\|_2^2 \leq \|\Phi\overline{x} - y\|_2^2 + 2.5\varepsilon_s(\overline{x})^2/(1-\delta_{s\alpha}),$$

and

$$\|x^k - \overline{x}\|_2 \leq \sqrt{6}\varepsilon_s(\overline{x})/(1-\delta_{s\alpha}).$$

Next, we will use Theorem 3 in the case of Ω being empty to prove the following corollary, which plays an important role in the proof of Theorem 2.

Corollary 2. *Let* $\overline{x} \in \mathbb{R}^n, \overline{k} = |Gsupp(\overline{x})|$. *Let* $0 < \delta \leq 1, \beta = \lceil 16 + 15\delta \rceil$. *If the GRIP condition* $\delta_{\overline{k}\alpha} + (1+\delta)\delta_{\beta\overline{k}\alpha} \leq \delta$ *holds, then for* $k = (\beta - 1)\overline{k}$, *we have*

$$\|\Phi x^k - y\|_2^2 \leq (11 + 20\delta)\|\Phi\overline{x} - y\|_2^2.$$

Proof. If $\delta_{\bar{k}\alpha} + (1+\delta)\delta_{\beta\bar{k}\alpha} \leq \delta$, then

$$1 \leq \frac{1+\delta_\alpha}{1-\delta_{\beta\bar{k}\alpha}} \leq \frac{1+\delta_{\bar{k}\alpha}}{1-\delta_{\beta\bar{k}\alpha}} \leq 1+\delta. \tag{11}$$

Therefore, we take $s = \beta\bar{k}$ in (10). Since $0 < \delta \leq 1$, then for $k = (\beta-1)\bar{k}$, using the properties of rounding function and logarithm function, we have

$$\begin{aligned} k &\geq \lceil 4\bar{k}(1+\delta)\ln(20(1+\delta))\rceil \\ &\geq \lceil 4\bar{k}\frac{1+\delta_\alpha}{1-\delta_{\beta\bar{k}\alpha}} \ln 20\frac{1+\delta_{\bar{k}\alpha}}{1-\delta_{\beta\bar{k}\alpha}}\rceil. \end{aligned}$$

From Theorem 3 and inequality (11), we conclude

$$\begin{aligned} \|\Phi x^k - y\|_2^2 &\leq \|\Phi(\bar{x}) - y\|_2^2 + 2.5\varepsilon_{\beta\bar{k}}(\bar{x})^2/(1-\delta_{\beta\bar{k}\alpha}) \\ &\leq \|\Phi(\bar{x}) - y\|_2^2 + 2.5\cdot 4(1+\delta_{\beta\bar{k}\alpha})\|\Phi\bar{x} - y\|_2^2/(1-\delta_{\beta\bar{k}\alpha}) \\ &= \|\Phi(\bar{x}) - y\|_2^2 + 10\frac{1+\delta_{\beta\bar{k}\alpha}}{1-\delta_{\beta\bar{k}\alpha}}\|\Phi\bar{x} - y\|_2^2 \\ &\leq \|\Phi(\bar{x}) - y\|^2 + 10(1+2\delta)\|\Phi\bar{x} - y\|_2^2 \\ &= (11+20\delta)\|\Phi\bar{x} - y\|_2^2, \end{aligned}$$

where the second inequality holds, since $\varepsilon_{\beta\bar{k}}(\bar{x}) \leq 2\sqrt{1+\delta_{\beta\bar{k}\alpha}}\|\Phi\bar{x} - y\|_2$. □

Now, we prove Theorem 2 with the help of Corollary 2.

Proof of Theorem 2. Let $\{x_{\Lambda_0}, x_{\Lambda_1}, x_{\Lambda_2}, \ldots, x_{\Lambda_s}\}$ be an optimal group k-sparse decomposition of x. Taking $\bar{x} = x_{\Lambda_0}$ in Corollary 2.1, we obtain that

$$\begin{aligned} \|\Phi x^* - y\|_2 &\leq \sqrt{11+20\delta}\|\Phi x_{\Lambda_0} - y\|_2 \\ &\leq \sqrt{11+20\delta}\|\Phi x_{\Lambda_0} - \Phi x - e\|_2 \\ &\leq \sqrt{11+20\delta}(\|\Phi x_{\Lambda_0} - \Phi x\|_2 + \|e\|_2), \end{aligned} \tag{12}$$

where $x^* := \mathrm{GOMP}_{(\beta-1)k}(\Phi x + e)$. Noting that $|\mathrm{Gsupp}(x^* - x_{\Lambda_0})| \leq \beta k$ and $(1+\delta)\delta_{\beta k\alpha} \leq \delta - \delta_{k\alpha}$, we have

$$\begin{aligned} (1+\delta)(1-\delta_{\beta k\alpha}) &= 1+\delta-(1+\delta)\delta_{\beta k\alpha} \\ &\geq 1+\delta-\delta+\delta_{k\alpha} \\ &\geq 1. \end{aligned} \tag{13}$$

Combining (12) with (13), we conclude

$$\begin{aligned} \|x^* - x_{\Lambda_0}\|_2 &\leq \frac{1}{\sqrt{1-\delta_{\beta k\alpha}}}\|\Phi x^* - \Phi x_{\Lambda_0}\|_2 \\ &\leq \sqrt{1+\delta}(\|\Phi x^* - y\|_2 + \|\Phi x_{\Lambda_0} - y\|_2) \\ &\leq \sqrt{1+\delta}(\|\Phi x^* - y\|_2 + \|\Phi x_{\Lambda_0} - y\|_2) \\ &\leq \sqrt{1+\delta}(\sqrt{11+20\delta}+1)\|\Phi x_{\Lambda_0} - y\|_2 \\ &\leq \sqrt{1+\delta}(\sqrt{11+20\delta}+1)(\|\Phi x_{\Lambda_0} - \Phi x\|_2 + \|e\|_2). \end{aligned} \tag{14}$$

We now estimate the term $\|\Phi x_{\Lambda_0} - \Phi x\|_2$. It is clear that

$$\|\Phi x_{\Lambda_0} - \Phi x\|_2 = \|\Phi(x_{\Lambda_0} - x)\|_2 = \|\Phi(\sum_{i=1}^{s} x_{\Lambda_i})\|_2 = \|\sum_{i=1}^{s} \Phi x_{\Lambda_i}\|_2 \leq \sum_{i=1}^{s} \|\Phi x_{\Lambda_i}\|_2. \quad (15)$$

By the GRIP of the matrix Φ, we obtain

$$\|\Phi x_{\Lambda_0} - \Phi x\|_2 \leq \sqrt{1+\delta} \sum_{i=1}^{s} \|x_{\Lambda_i}\|_2. \quad (16)$$

Using (5) and the decomposability of $\|\cdot\|_A$, we have

$$\begin{aligned} \|x - x_{\Lambda_0}\|_2 &\leq \sum_{i=1}^{s} \|x_{\Lambda_i}\|_2 \leq \frac{1}{c} \sum_{i=1}^{s} \|x_{\Lambda_i}\|_A \\ &= \frac{1}{c} \|\sum_{i=1}^{s} x_{\Lambda_i}\|_A = \frac{1}{c} \|x_{\Lambda_0^c}\|_A = \frac{1}{c} \sigma_{k,\mathcal{G}}(x)_A. \end{aligned} \quad (17)$$

Combining inequalities (14)–(17), we obtain

$$\|x^* - x_{\Lambda_0}\|_2 \leq (1+\delta)(1+\sqrt{11+20\delta}) \frac{\sigma_{k,\mathcal{G}}(x)_A}{c} + \sqrt{1+\delta}(1+\sqrt{11+20\delta})\|e\|_2. \quad (18)$$

Inequalities (17) and (18) imply that

$$\begin{aligned} \|x^* - x\|_2 &\leq \|x^* - x_{\Lambda_0}\|_2 + \|x_{\Lambda_0} - x\|_2 \\ &\leq \frac{1}{c}[(1+\delta)(1+\sqrt{11+20\delta}) + 1]\sigma_{k,\mathcal{G}}(x)_A + \sqrt{1+\delta}(1+\sqrt{11+20\delta})\|e\|_2, \end{aligned}$$

which leads to the bound in (6).

To derive inequality (8), we adopt a different strategy. Let $h = x^* - x_{\Lambda_0}$. Without loss of generality, we assume that $\mathrm{Gsupp}(x^* - x_{\Lambda_0}) = \{1, 2, \ldots, g\}$ and $|G_1| \geq |G_2| \geq \ldots |G_g|$. We construct a subset S_1 of $\{1, 2, \ldots, g\}$ as follows:

Note that $|G_1| \leq \alpha \leq k$. We picked 1 as an element of S_1; S_1 contains 2 if and only if $|G_1| + |G_2| \leq k$. Inductively suppose we have constructed the set $\Omega_{k-1} = S_1 \cap \{1, 2, \ldots, k-1\}$, then k is an element of S_1 if and only if

$$\sum_{i \in \Omega_{k-1} \cup \{k\}} |G_i| \leq k.$$

By this method, we can construct a unique subset S_1. Then, by using the same method, we can construct a unique subset:

$$S_2 \subseteq \{1, 2, \ldots, g\} - S_1.$$

Inductively suppose we have constructed subsets $S_1, \ldots, S_{k-1}$, we can form a subset $\{1, 2, \ldots, g\} - \bigcup_{i=1}^{k-1} S_i$, then we can construct a unique subset $S_k \subseteq \{1, 2, \ldots, g\} - \bigcup_{i=1}^{k-1} S_i$. In this way, we can decompose $\{1, 2, \ldots, g\}$ as follows:

$$\{1, 2, \ldots, g\} = \bigcup_{i=1}^{l} S_i.$$

From the construction, we know that when $l \neq 1$, for $i \neq l$,

$$|G_{S_i}| > \frac{k}{2}.$$

For any $1 \le i \le l$,

$$|G_{S_{i-1}}| + |G_{S_i}| > k.$$

So we have

$$\sum_{i=1}^{l} |G_{S_i}| > \frac{k}{2} l,$$

which is to say, $\beta k\alpha > \frac{k}{2} l$, so $l < 2\beta\alpha$. When $l = 1$, we also have $l < 2\beta\alpha$.

Using the Cauchy–Schwarz inequality and the decomposability of $\|\cdot\|_A$, we have

$$\|h\|_A = \sum_{i=1}^{l} \|h_{G_{S_i}}\|_A \le d \sum_{i=1}^{l} \|h_{G_{S_i}}\|_2 \le d\sqrt{l}(\sum_{i=1}^{l} \|h_{G_{S_i}}\|_2^2)^{\frac{1}{2}} = d\sqrt{l}\|h\|_2 \le d\sqrt{2\beta\alpha - 1}\|h\|_2. \tag{19}$$

Combining (18) with (19), we have

$$\begin{aligned}
\|x^* - x_{\Lambda_0}\|_A &\le d\sqrt{2\beta\alpha - 1}\|x^* - x_{\Lambda_0}\|_2 \\
&\le \frac{d}{c}\sqrt{2\beta\alpha - 1}(1+\delta)(1+\sqrt{11+20\delta})\sigma_{k,\mathcal{G}}(x)_A \\
&\quad + d\sqrt{(2\beta\alpha - 1)(1+\delta)}(1+\sqrt{11+20\delta})\|e\|_2.
\end{aligned}$$

From the above inequality, we conclude

$$\begin{aligned}
\|x^* - x\|_A &\le \|x^* - x_{\Lambda_0}\|_A + \|x_{\Lambda_0} - x\|_A \\
&= \|x^* - x_{\Lambda_0}\|_A + \sigma_{k,\mathcal{G}}(x)_A \\
&\le [\frac{d}{c}\sqrt{2\beta\alpha - 1}(1+\delta)(1+\sqrt{11+20\delta}) + 1]\sigma_{k,\mathcal{G}}(x)_A \\
&\quad + d\sqrt{(2\beta\alpha - 1)(1+\delta)}(1+\sqrt{11+20\delta})\|e\|_2.
\end{aligned}$$

This leads to the bound in (8), and hence we complete the proof of Theorem 2. □

3. Simulation Results

In this section, we test the performance of the GOMP and present the results of our experiments. We first describe the details relevant to our experiments in Section 3.1. We demonstrate the effectiveness of the GOMP in Section 3.2. In Section 3.3, we compare the GOMP with Basis Pursuit to show the efficiency and scalability of our algorithm.

3.1. Implementation

For all experiments, we considered the following model. Suppose that $x \in \mathbb{R}^N$ is an unknown N-dimensional signal and we wish to recover it by the given data

$$y = \Phi x + e, \tag{20}$$

where $\Phi \in \mathbb{R}^{M \times N}$ is a known measurement matrix with $M \ll N$ and e is a noise. Furthermore, since $M \ll N$, the column vectors of Φ are linearly dependent and the collection of these columns can be viewed as a redundant dictionary.

For arbitrary $x, y \in \mathbb{R}^N$, define

$$\langle x, y \rangle = \sum_{j=1}^{N} x_j y_j,$$

and

$$\|x\|_2 = \left(\sum_{j=1}^{N} |x_j|^2\right)^{1/2},$$

where $x = (x_j)_{j=1}^N$ and $y = (y_j)_{j=1}^N$. Obviously, $\mathbb{R}^N$ is a Hilbert space with the inner product $\langle \cdot, \cdot \rangle$.

In the experiment, we set the measurement matrix Φ to be a Gaussian matrix where each entry is selected from the $\mathcal{N}(0, M^{-1})$ distribution and the density function of this distribution is $p(x) := \frac{1}{\sqrt{2\pi M}} e^{-x^2 M/2}$. We executed the GOMP with the data vector $y = \Phi x$.

To demonstrate the performance of signal-recovering algorithms, we use the mean square error (MSE) to measure the error between the real signal x and its approximant $\widehat{x}$, which is defined as follows:

$$\mathrm{MSE} = \frac{1}{N} \sum_{j=1}^{N} (x_j - \widehat{x})^2.$$

For the Group Orthogonal Matching Pursuit (GOMP), we let $\Omega = \varnothing \in \{1, \cdots, g\}$ as the initialization. In the experiments, we constructed input signals randomly by the following steps:

i. Given a sparse level K;
ii. Produce a group structure $\{G_i\}_{i=1}^g$ randomly, satisfying $\#G_i \leq K$ for each index i;
iii. Randomly select a set $S \subseteq \{1, \cdots, g\}$, such that $\#G_S := \#\{\sum_{i \in S} G_i\} = K$;
iv. Let the set G_S be the support, and produce a signal by random numbers from normal distribution $\mathcal{N}(0, 1)$.

Some of examples of the randomly constructed group 50-sparse signals in different dimensions can be found in Figure 1.

We set Basis Pursuit (BP) as the baseline for further comparison with the GOMP to analyze the latter's properties. For the implementation of Basis Pursuit, we used the l_1 Magic toolbox (Open-sourced code: https://candes.su.domains/software/l1magic/#code, accessed on 15 February 2023) developed by E. Candès and J. Romberg in [33].

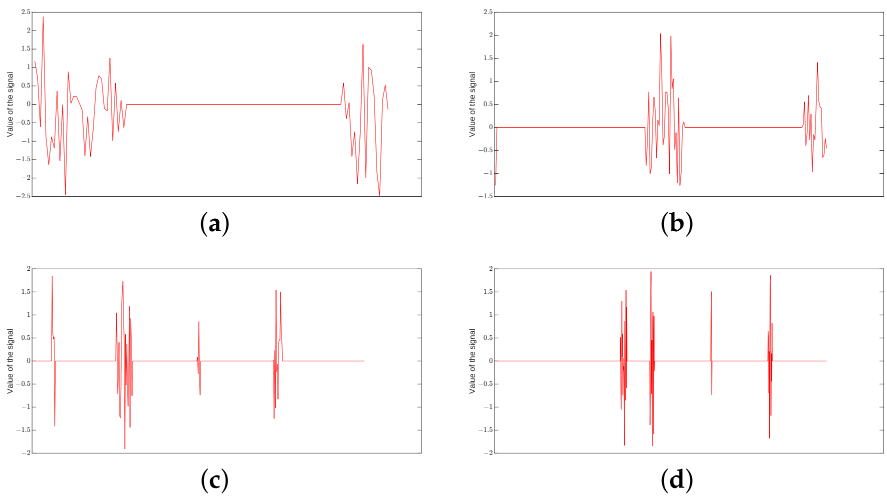
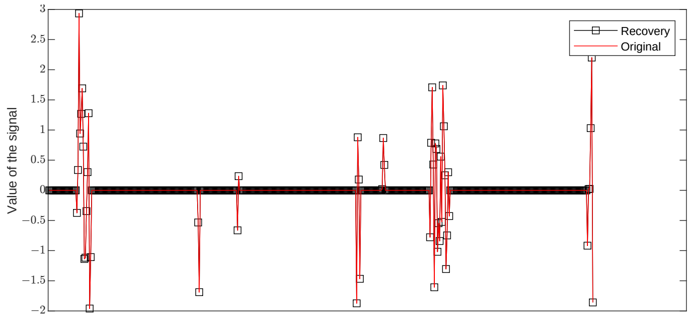


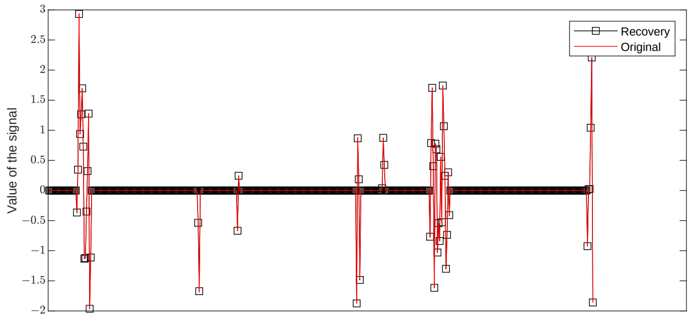
Figure 1. Examples of group 50-sparse signals in different dimensions: (**a**) An example of the group 50-sparse signal in dimension $N = 128$. (**b**) An example of the group 50-sparse signal in dimension $N = 256$. (**c**) An example of the group 50-sparse signal in dimension $N = 512$. (**d**) An example of the group 50-sparse signal in dimension $N = 1024$.

3.2. *Effectiveness of the GOMP*

Figure 2 shows the performance of the GOMP for an input signal in dimension $N = 512$ with group sparsity level $K = 50$ and number of measurements $M = 200$ under different noises, where the red line represents the original signal and the black squares represent the approximation. Figure 3 shows the performance of BP. The poor performance of BP can be easily observed when the noise occurs.

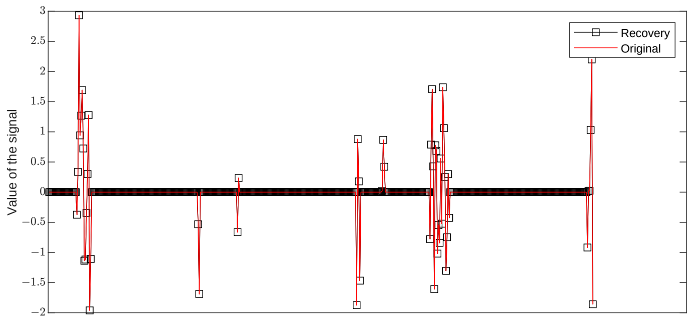


(**a**)

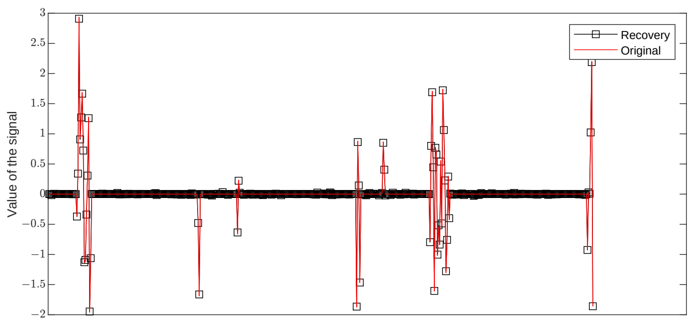


(**b**)

Figure 2. The recovery of an input signal via GOMP in dimension $N = 512$ with group sparsity level $K = 50$ and number of measurements $M = 200$ under different noises: (**a**) The recovery of an input signal via GOMP under the noise $e = 0$. (**b**) The recovery of an input signal via GOMP under a Gaussian noise e from $\mathcal{N}(0, 0.1^2)$.



(**a**)



(**b**)

Figure 3. The recovery of an input signal via BP in dimension $N = 512$ with group sparsity level $K = 50$ and number of measurements $M = 200$ under different noises: (**a**) The recovery of an input signal via BP under the noise $e = 0$. (**b**) The recovery of an input signal via BP under a Gaussian noise e from $\mathcal{N}(0, 0.1^2)$.

In general, the above example shows that the GOMP is very effective for signal recovering, i.e., it can recover the group sparse signal exactly. We will analyze the performance of the GOMP further in terms of the mean square error by a comparison with BP in the next subsection.

Figure 4 describes the situation in dimension $d = 1024$ under noise $e = 0$. It displays the relation between the percentage (of 100 input signals) of the support that can be recovered correctly and the number N of measurements. Furthermore, to some extent, it shows how many measurements are necessary to recover the support of the input group K-sparse signal $x \in \mathbb{R}^d$ with high probability. If the percentage equals 100%, it means that support of all the 100 input signals can be found, which implies the support of input signal can be exactly recovered. As expected, Figure 4 shows that when the group sparsity level K increases, it is necessary to increase the number N of measurements to guarantee signal recovery. Furthermore, we can find that the GOMP can recover the correct support of the input signal with high probability.

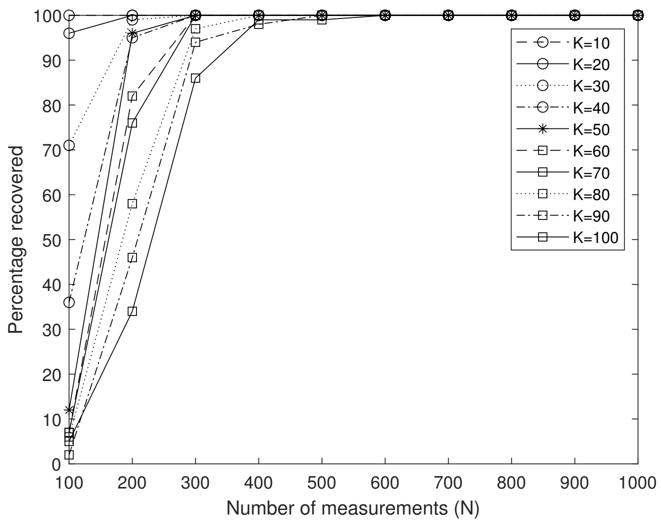


Figure 4. The percentage of the support of 100 input signals correctly recovered as a function of number N of Gaussian measurements for different group sparsity levels K in dimension $d = 1024$.

3.3. Comparison with Basis Pursuit

In order to demonstrate the efficiency and robustness of the Group Orthogonal Matching Pursuit (GOMP), we implement the Group Orthogonal Matching Pursuit (GOMP) and the Basis Pursuit (BP) to recover the group sparse signals for comparison.

We implemented the GOMP and BP for recovering the group sparse signals in dimension $N = 1024$ with number of measurements $M = 500$ under different noises to calculate the average mean square error (MSE) and the average running time by repeating the test 100 times. Figures 5 and 6, respectively, show the MSE and running time of the GOMP, implying that the error and the running time of the GOMP will constantly increase with the increased $K \in \{10, 20, \cdots, 100\}$. The figure of the error of the GOMP in terms of MSE is about 10^{-31} (noiseless) and 10^{-7} (Gaussian noise), which means that the GOMP can effectively recover the group sparse signal. Meanwhile, the computational complexity of the GOMP is also relatively low according to its running time. Figures 7 and 8 show that the MSE and the running time of the BP have the same trend as the GOMP: the error and the running time increase with the increased sparsity level, without changes to other parameters, while the MSE of BP is higher in figure when compared with the GOMP, which is around 10^{-7} (noiseless) and 10^{-5} (Gaussian noise). In order to facilitate comparison, the details of the above results in Figures 5–8 also can be found in Tables 1 and 2.

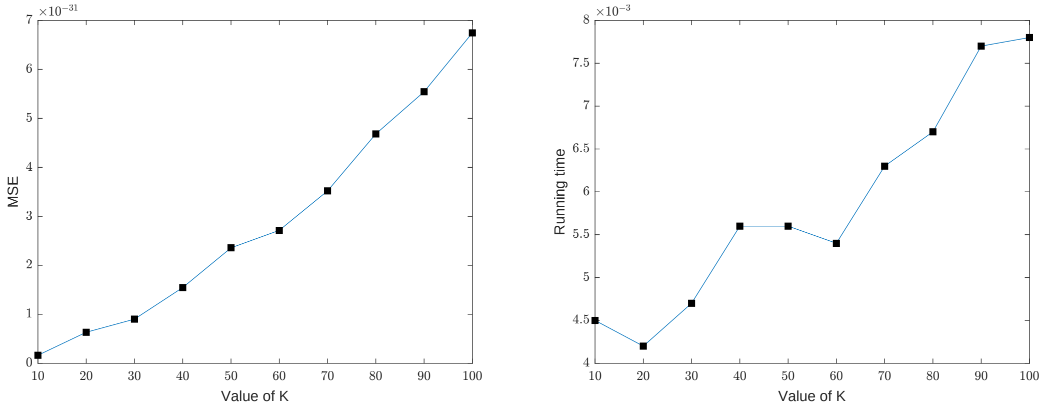


Figure 5. MSE and running time of the GOMP in dimension $N = 1024$ with number of measurements $M = 500$ under the noise $e = 0$.

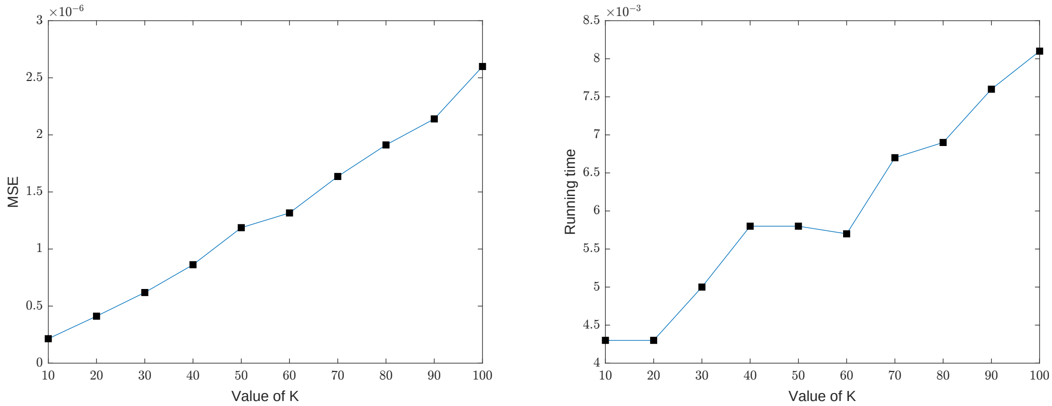


Figure 6. MSE and running time of the GOMP in dimension $N = 1024$ with number of measurements $M = 500$ under a Gaussian noise e from $\mathcal{N}(0, 0.1^2)$.

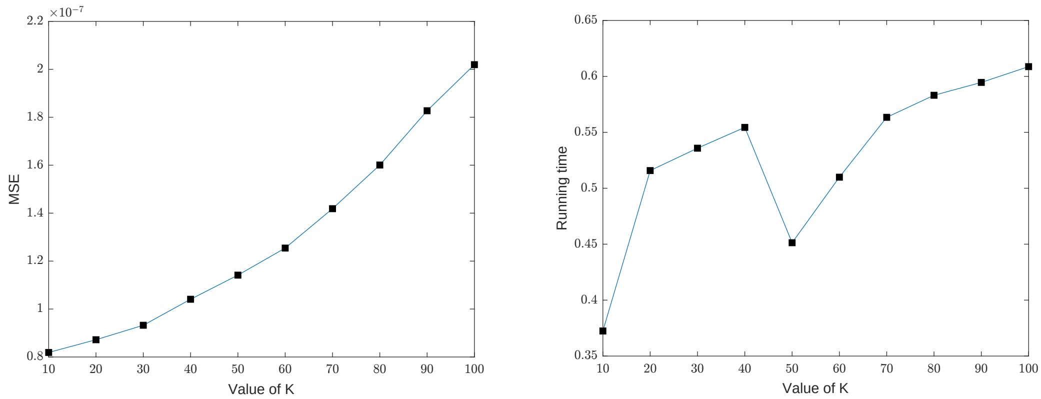


Figure 7. MSE and running time of BP in dimension $N = 1024$ with number of measurements $M = 500$ under the noise $e = 0$.

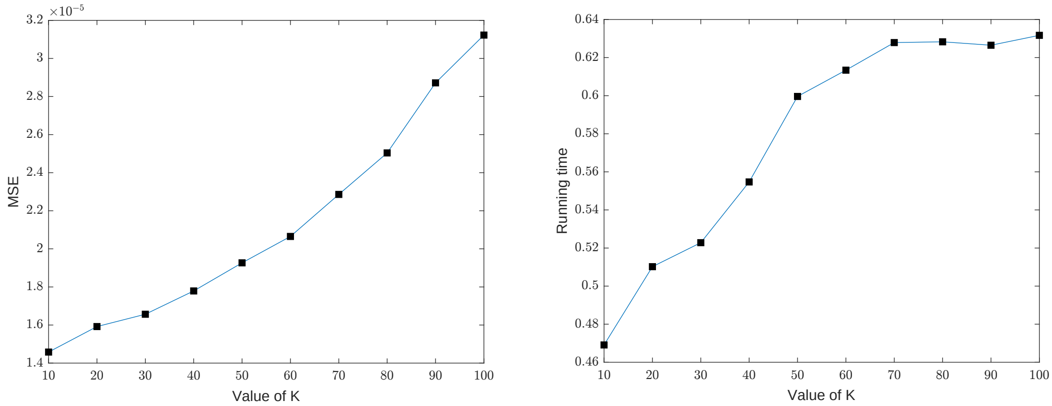


Figure 8. MSE and running time of BP in dimension $N = 1024$ with number of measurements $M = 500$ under a Gaussian noise e from $\mathcal{N}(0, 0.1^2)$.

Table 1. The average MSE and running time (repeating 100 times) of the GOMP and BP in 1024 dimension with number of measurements $M = 500$ under the noise $e = 0$.

Group Sparsity Level	GOMP		BP	
	MSE	**Running Time**	**MSE**	**Running Time**
K = 10	1.6350×10^{-32}	0.0045s	8.1887×10^{-08}	0.3724 s
K = 20	6.3295×10^{-32}	0.0042 s	8.7195×10^{-8}	0.5158 s
K = 30	9.0108×10^{-32}	0.0047 s	9.3230×10^{-8}	0.5358 s
K = 40	1.5460×10^{-31}	0.0056 s	1.0408×10^{-7}	0.5544 s
K = 50	2.3565×10^{-31}	0.0056 s	1.1414×10^{-7}	0.4513 s
K = 60	2.7140×10^{-31}	0.0054 s	1.2543×10^{-7}	0.5099 s
K = 70	3.5189×10^{-31}	0.0063 s	1.4186×10^{-7}	0.5634 s
K = 80	4.6822×10^{-31}	0.0067 s	1.6007×10^{-7}	0.5831 s
K = 90	5.5434×10^{-31}	0.0077 s	1.8272×10^{-7}	0.5946 s
K = 100	6.7451×10^{-31}	0.0078 s	2.0194×10^{-7}	0.6087 s

Table 2. The average MSE and running time (repeating 100 times) of the GOMP and BP in 1024 dimension with number of measurements $M = 500$ under a Gaussian noise e from $\mathcal{N}(0, 0.1^2)$.

Group Sparsity Level	GOMP		BP	
	MSE	**Running Time**	**MSE**	**Running Time**
K = 10	2.1440×10^{-7}	0.0043 s	1.4580×10^{-5}	0.4691 s
K = 20	4.1167×10^{-7}	0.0043 s	1.5919×10^{-5}	0.5102 s
K = 30	6.1894×10^{-7}	0.0050 s	1.6566×10^{-5}	0.5228 s
K = 40	8.6262×10^{-7}	0.0058 s	1.7790×10^{-5}	0.5547 s
K = 50	1.1870×10^{-6}	0.0058 s	1.9266×10^{-5}	0.5996 s
K = 60	1.3162×10^{-6}	0.0057 s	2.0652×10^{-5}	0.6134 s
K = 70	1.6356×10^{-6}	0.0067 s	2.2859×10^{-5}	0.6279 s
K = 80	1.9119×10^{-6}	0.0069 s	2.5038×10^{-5}	0.6283 s
K = 90	2.1399×10^{-6}	0.0076 s	2.8718×10^{-5}	0.6265 s
K = 100	2.5990×10^{-6}	0.0081 s	3.1227×10^{-5}	0.6317 s

According to the mean square error, obviously our algorithm outperforms the BP by a lot on both noiseless and noisy data. This trend is expected because the solution of the GOMP is updated by solving least squares problems. What is more, the BP normally cannot exactly recover the real signal since the produced solution may not be 0 on the indices which are not in the support, according to the observation during the experiments (can also be found in Figure 3b. For the GOMP, the solution's support has been restricted in the group sparse set based on the selected indices of the group structure $\{G\}_{i=1}^{g}$ at each iteration. That also leads to the better performance of the GOMP compared with BP.

Although the performance of the GOMP can be affected by the group sparsity level and noise, our method still has pretty low mean square errors and performs better than BP, showing the robustness and effectiveness of our proposed algorithm. We can also observe that the GOMP runs much quicker than the BP, demonstrating the scalability and efficiency of our algorithm.

Further more, when it comes to large-scale group sparse signal recovery, the execution of BP occupies more memory and takes far more time to obtain the solution compared with our method (details can be found in Tables 3 and 4 for different noise scenarios). This is because the scale of least square problems that the GOMP needs to solve to obtain the solution are bounded by the sparsity level K, as seen in the GOMP defined in Section 1. At the same time, the GOMP still outperforms BP a lot in terms of MSE. In addition, we can also find that the running speed of BP will be slightly slower when the noise occurs, while for the GOMP it will not be affected.

Table 3. The average MSE and running time (repeating 100 times) of the GOMP and BP with number of measurements $M = 500$ and group sparsity level $K = 50$ under the noise $e = 0$.

Dimension	GOMP		BP	
	MSE	Running Time	MSE	Running Time
N = 1024	2.3846×10^{-31}	0.0057 s	1.1414×10^{-7}	0.4513 s
N = 2048	1.1178×10^{-31}	0.0114 s	4.4095×10^{-8}	1.6611 s
N = 3072	8.5425×10^{-32}	0.0195 s	2.7722×10^{-8}	5.8170 s
N = 4096	5.9586×10^{-32}	0.0413 s	1.9752×10^{-8}	11.3636 s

Table 4. The average MSE and running time (repeating 100 times) of the GOMP and BP with number of measurements $M = 500$ and group sparsity level $K = 50$ under a Gaussian noise e from $\mathcal{N}(0, 0.1^2)$.

Dimension	GOMP		BP	
	MSE	Running Time	MSE	Running Time
N = 1024	1.1143×10^{-6}	0.0066 s	1.9539×10^{-5}	0.4521 s
N = 2048	5.7826×10^{-7}	0.0115 s	7.8597×10^{-6}	2.6113 s
N = 3072	3.9382×10^{-7}	0.0200 s	5.3449×10^{-6}	6.3476 s
N = 4096	3.0434×10^{-7}	0.0406 s	4.4554×10^{-6}	12.1959 s

4. Conclusions

We propose the Group Orthogonal Matching Pursuit (GOMP) to recover group sparse signals. We analyze the error of the GOMP algorithm in the process of recovering group sparse signals from noisy measurements. We show the instance optimality and robustness of the GOMP under the group restricted isometry property (GRIP) of the decoder matrix. Compared with the P-norm minimization approach, the GOMP has two advantages. One is its easier implementation, that is, it runs quickly and has lower computational complexity. The other is that we do not need the concept of γ-decomposability. Furthermore, our simulation results show that the GOMP is very efficient for group sparse signal recovery and significantly outperforms Basis Pursuit in both scalability and solution quality.

On the other hand, there are several algorithms that we will study in the future, such as Sparsity Adaptive Matching Pursuit (SAMP) [34], Constrained Backtracking Matching Pursuit (CBMP) [35] and Group-based Sparse Representation-Joint Regularization (GSR-JR) [36], which could be potentially competitive with the GOMP.

Author Contributions: Conceptualization, all authors; methodology, all authors; software, all authors; validation, all authors; formal analysis, all authors; investigation, all authors; resources, all authors; data curation, all authors; writing—original draft preparation, all authors; writing—review and editing, all authors; visualization, all authors; supervision, all authors; project administration, all authors; funding acquisition, all authors. All authors have read and agreed to the published version of the manuscript.

Funding: The work is supported by the National Natural Science Foundation of China Grant No. 11671213 and No. 11701411.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Acknowledgments: The authors would like to thank the referees and the editors for their very useful suggestions, which improved significantly this paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

In this appendix, we establish three lemmas which will be used in the proof of Theorem 3. The first one is concerned with the reduction of the residuals for the GOMP algorithm.

Lemma A1. *Let $(r^k)_{k\geq 0}$ be the residual error sequences generated by the GOMP applied to y, $\overline{F} \subseteq \{1, 2, \ldots, g\}$, and let $\overline{x} = \arg\min_{z:Gsupp(z)\subseteq \overline{F}} \|y - \Phi z\|_2$. If $\overline{F}$ is not contained in I_k, then we have*

$$\|r^{k+1}\|_2^2 \leq \|r^k\|_2^2 - \frac{1-\delta_{|\overline{F}\cup I_k|\alpha}}{(1+\delta_\alpha)|\overline{F}\setminus I_k|} \max\{0,\ \|r^k\|_2^2 - \|y - \Phi\overline{x}\|_2^2\}. \tag{A1}$$

Proof. The proof of Lemma A1 is based on the ideas of Cohen, Dahmen and DeVore [24]. We may assume that $\|r^k\|_2 \geq \|y - \Phi\overline{x}\|_2$. Otherwise, inequality (A1) is trivially satisfied.

Denote

$$H_{k+1} = \text{span}\{\varphi_i : i \in G_{I_{k+1}}\} \quad \text{and} \quad F_{k+1} = \text{span}\{\varphi_j : j \in G_{i_{k+1}}\}.$$

It is clear that $F_{k+1} \subseteq H_{k+1}$. Note that

$$\begin{aligned} r^{k+1} = y - \Phi x^{k+1} &= y - \Phi x^k + \Phi x^k - \Phi x^{k+1} \\ &= r^k + \Phi x^k - P_{H_{k+1}}(y) \\ &= r^k + \Phi x^k - P_{H_{k+1}}(r^k + \Phi x^k) \\ &= r^k - P_{H_{k+1}}(r^k), \end{aligned}$$

which implies

$$\begin{aligned} \|r^{k+1}\|_2^2 &= \|r^k - P_{H_{k+1}}(r^k)\|_2^2 \\ &\leq \|r^k - P_{F_{k+1}}(r^k)\|_2^2 \\ &= \|r^k\|_2^2 - \|P_{F_{k+1}}(r^k)\|_2^2. \end{aligned} \tag{A2}$$

Now, we estimate $\|P_{F_{k+1}}(r^k)\|_2$. Since Φ satisfies the GRIP of order α with the isometry constant δ_α, we have

$$\|\sum_{j\in G_{i_{k+1}}} a_j\varphi_j\|_2^2 \leq (1+\delta_\alpha) \sum_{j\in G_{i_{k+1}}} |a_j|^2, \tag{A3}$$

which results in the following lower estimate

$$
\begin{aligned}
\|P_{F_{k+1}}(r^k)\|_2 &= \sup_{\psi\in F_{k+1},\|\psi\|\le 1} |\left\langle P_{F_{k+1}}(r^k),\, \psi\right\rangle| \\
&= \sup_{\{c_j\}_{j\in G_{i_{k+1}}},\|\sum_{j\in G_{i_{k+1}}} c_j\varphi_j\|\le 1} |\left\langle P_{F_{k+1}}(r^k),\, \sum_{j\in G_{i_{k+1}}} c_j\varphi_j\right\rangle| \\
&= \sup_{\{c_j\}_{j\in G_{i_{k+1}}},\|\sum_{j\in G_{i_{k+1}}} c_j\varphi_j\|\le 1} |\sum_{j\in G_{i_{k+1}}} c_j\left\langle r^k,\, \varphi_j\right\rangle| \\
&\ge \sup_{\{c_j\}_{j\in G_{i_{k+1}}},\sum_{j\in G_{i_{k+1}}} |c_j|^2\le(1+\delta_\alpha)^{-1}} |\sum_{j\in G_{i_{k+1}}} c_j\left\langle r^k,\, \varphi_j\right\rangle| \\
&= (1+\delta_\alpha)^{-\frac{1}{2}}(\sum_{j\in G_{i_{k+1}}} |\left\langle r^k,\, \varphi_j\right\rangle|^2)^{\frac{1}{2}},
\end{aligned}
\tag{A4}
$$

where the first inequality is obtained by (A3) and the Cauchy–Schwarz inequality. Using inequality (A4), we can continue to estimate (A2) and conclude

$$
\|r^{k+1}\|_2^2 \le \|r^k\|_2^2 - (1+\delta_\alpha)^{-1}(\sum_{j\in G_{i_{k+1}}} |\left\langle r^k,\, \varphi_j\right\rangle|^2). \tag{A5}
$$

Therefore, in view of inequalities (A5) and (A1), it remains to prove that

$$
(1+\delta_\alpha)^{-1}\|\Phi^*[i_{k+1}]r^k\|_2^2 \ge \frac{1-\delta_{|\overline{F}\cup I_k|\alpha}}{(1+\delta_\alpha)|\overline{F}\setminus I_k|}(\|r^k\|_2^2 - \|y-\Phi\overline{x}\|_2^2),
$$

which is equivalent to

$$
\|r^k\|_2^2 - \|y-\Phi\overline{x}\|_2^2 \le \frac{|\overline{F}\setminus I_k|\cdot\|\Phi^*[i_{k+1}]r^k\|_2^2}{1-\delta_{|\overline{F}\cup I_k|\alpha}}. \tag{A6}
$$

We note that

$$
\begin{aligned}
2\sqrt{\|r^k\|_2^2 - \|y-\Phi\overline{x}\|_2^2}\,\|\Phi\overline{x}-\Phi x^k\|_2 &\le \|r^k\|_2^2 - \|y-\Phi\overline{x}\|_2^2 + \|\Phi\overline{x}-\Phi x^k\|_2^2 \\
&= \|r^k\|_2^2 - \|r^k+\Phi x^k-\Phi\overline{x}\|_2^2 + \|\Phi\overline{x}-\Phi x^k\|_2^2 \\
&\le 2|\left\langle r^k,\, \Phi x^k-\Phi\overline{x}\right\rangle| \\
&= 2|\left\langle r^k,\, \Phi\overline{x}\right\rangle|.
\end{aligned}
$$

This is the same as

$$
\|r^k\|_2^2 - \|y-\Phi\overline{x}\|_2^2 \le \frac{|\left\langle r^k,\, \Phi\overline{x}\right\rangle|^2}{\|\Phi\overline{x}-\Phi x^k\|_2^2}. \tag{A7}
$$

On the one hand, by the GRIP of Φ, we have

$$
\begin{aligned}
\|\Phi\overline{x}-\Phi x^k\|_2^2 &= \|\sum_{i\in\overline{F}}\Phi[i]\overline{x}[i] - \sum_{i\in I_k}\Phi[i]x^k[i]\|_2^2 \\
&= \|\sum_{i\in\overline{F}\setminus I_k}\Phi[i]\overline{x}[i] + \sum_{i\in\overline{F}\cap I_k}\Phi[i](\overline{x}[i]-x^k[i]) - \sum_{i\in I_k\setminus\overline{F}}\Phi[i]x^k[i]\|_2^2 \\
&\ge (1-\delta_{|\overline{F}\cup I_k|\alpha})(\sum_{i\in\overline{F}\setminus I_k}\|\overline{x}[i]\|_2^2 + \sum_{i\in\overline{F}\cap I_k}\|\overline{x}[i]-x^k[i]\|_2^2 + \sum_{i\in I_k\setminus\overline{F}}\|x^k[i]\|_2^2) \\
&\ge (1-\delta_{|\overline{F}\cup I_k|\alpha})\sum_{i\in\overline{F}\setminus I_k}\|\overline{x}[i]\|_2^2.
\end{aligned}
\tag{A8}
$$

On the other hand, we have

$$
\begin{aligned}
|\left\langle r^k, \Phi\overline{x}\right\rangle|^2 &= |\left\langle r^k, \sum_{i\in\overline{F}} \Phi[i]\overline{x}[i]\right\rangle|^2 = |\left\langle r^k, \sum_{i\in\overline{F}\setminus I_k} \Phi[i]\overline{x}[i]\right\rangle|^2 \\
&= |\sum_{i\in\overline{F}\setminus I_k} \left\langle r^k, \Phi[i]\overline{x}[i]\right\rangle|^2 = |\sum_{i\in\overline{F}\setminus I_k} \left\langle \Phi^*[i]r^k, \overline{x}[i]\right\rangle|^2.
\end{aligned}
$$

According to the greedy step of the GOMP, we further derive

$$
\begin{aligned}
|\left\langle r^k, \Phi\overline{x}\right\rangle|^2 &= |\sum_{i\in\overline{F}\setminus I_k} \left\langle \Phi^*[i]r^k, \overline{x}[i]\right\rangle|^2 \\
&\leq (\sum_{i\in\overline{F}\setminus I_k} \|\Phi^*[i]r^k\|_2 \cdot \|\overline{x}[i]\|_2)^2 \\
&\leq \|\Phi^*[i_{k+1}]r^k\|_2^2 \cdot (\sum_{i\in\overline{F}\setminus I_k} \|\overline{x}[i]\|_2)^2 \\
&\leq \|\Phi^*[i_{k+1}]r^k\|_2^2 \cdot |\overline{F}\setminus I_k| \cdot \sum_{i\in\overline{F}\setminus I_k} \|\overline{x}[i]\|_2^2,
\end{aligned}
\tag{A9}
$$

where we have used the Cauchy–Schwarz inequality.

Combining inequalities (A7)–(A9), we obtain

$$
\begin{aligned}
\|r^k\|_2^2 - \|y-\Phi\overline{x}\|_2^2 &\leq \frac{|\left\langle r^k, \Phi\overline{x}\right\rangle|^2}{\|\Phi\overline{x}-\Phi x^k\|_2^2} \\
&\leq \frac{|\overline{F}\setminus I_k| \cdot \|\Phi^*[i_{k+1}]r^k\|_2^2 \cdot \sum_{i\in\overline{F}\setminus I_k}\|\overline{x}[i]\|_2^2}{(1-\delta_{|\overline{F}\cup I_k|\alpha})\sum_{\overline{F}\setminus I_k}\|\overline{x}[i]\|_2^2} \\
&= \frac{|\overline{F}\setminus I_k| \cdot \|\Phi^*[i_{k+1}]r^k\|_2^2}{(1-\delta_{|\overline{F}\cup I_k|\alpha})},
\end{aligned}
$$

and hence inequality (A6) holds. We complete the proof of Lemma A1. □

Corollary A1. *Assume that* $min_{z:Gsupp(z)\subseteq\overline{F}_j}\|\Phi z-y\|_2^2 \leq \|\Phi\overline{x}-y\|_2^2 + q_j$ $(j=0,1,2,\ldots,L)$ *for* $q_0 \geq q_1 \geq \ldots \geq q_L \geq 0$ *and subsets* $\overline{F}\cap\Omega = \overline{F}_0 \subseteq \overline{F}_1 \subseteq \overline{F}_2 \subseteq \ldots \subseteq \overline{F}_L \subseteq \overline{F}$. *If*

$$
k = \sum_{j=1}^{L}\lceil|\overline{F}_j\setminus\Omega|\frac{1+\delta_\alpha}{1-\delta_{s\alpha}}\ln(2\mu)\rceil \quad \textit{with} \quad \mu \geq \sup_{j=1,2,\ldots,L-1}\frac{q_{j-1}}{q_j}
$$

and

$$
s \geq |\overline{F}\cup I_k|,
$$

then

$$
\|r^k\|_2^2 \leq \|\Phi\overline{x}-y\|_2^2 + q_L + \mu^{-1}q_{L-1}. \tag{A10}
$$

Proof. By Lemma A1, for $l=0,1,2,\ldots,L$, we have either $|\overline{F}_l\setminus I_k|=0$ or inequality (A1) holds. Inequality (A1), along with $\|r^{k+1}\|_2^2 \leq \|r^k\|_2^2$, imply that either $|\overline{F}_l\setminus I_k|=0$ or

$$
\begin{aligned}
&\max\{0, \|r^{k+1}\|_2^2 - \|\Phi\overline{x}-y\|_2^2 - q_l\} \\
\leq\ &[1-\frac{1-\delta_{s\alpha}}{(1+\delta_\alpha)|\overline{F}_l\setminus I_k|}]\max\{0, \|r^k\|_2^2 - \|\Phi\overline{x}-y\|_2^2 - q_l\} \\
\leq\ &\exp[-\frac{1-\delta_{s\alpha}}{(1+\delta_\alpha)|\overline{F}_l\setminus I_k|}]\max\{0, \|r^k\|^2 - \|\Phi\overline{x}-y\|_2^2 - q_l\},
\end{aligned}
$$

where we use the inequality $\exp(-x) > 1-x$ for $x > 0$.

Therefore, for any $k^{'} \leq k$ and $l = 0, 1, 2, \ldots, L$, we have either $|\overline{F}_l \setminus I_{k'}| = 0$ or

$$\|r^k\|_2^2 - \|\Phi\overline{x} - y\|_2^2 - q_l \leq \exp[-\frac{(1-\delta_{s\alpha})(k-k^{'})}{(1+\delta_\alpha)|\overline{F}_l \setminus I_{k'}|}] \max\{0, \|r^{k^{'}}\|_2^2 - \|\Phi\overline{x} - y\|_2^2 - q_l\}. \tag{A11}$$

We now prove this corollary by induction on L. If $L = 1$, then we can set $k^{'} = 0$ and consider $\mu > 0$. When

$$k = \lceil |\overline{F}_1 \setminus \Omega| \frac{1+\delta_\alpha}{1-\delta_{s\alpha}} \ln(2\mu) \rceil,$$

we have by inequality (A11)

$$\begin{aligned}
\|r^k\|_2^2 - \|\Phi\overline{x} - y\|_2^2 - q_1 &\leq \exp[-\frac{(1-\delta_{s\alpha})k}{(1+\delta_\alpha)|\overline{F}_1 \setminus \Omega|}] \max\{0, \|r^0\|^2 - \|\Phi\overline{x} - y\|_2^2 - q_1\} \\
&\leq \exp[-\frac{(1-\delta_{s\alpha})k}{(1+\delta_\alpha)|\overline{F}_1 \setminus \Omega|}] q_0 \\
&\leq (2\mu)^{-1} q_0,
\end{aligned}$$

and hence

$$\|r^k\|_2^2 \leq \|\Phi\overline{x} - y\|_2^2 + q_1 + (2\mu)^{-1} q_0.$$

Note that this inequality also holds when $|\overline{F}_1 \setminus I_k| = 0$, since

$$\|r^k\|_2^2 \leq \min_{z:\mathrm{Gsupp}(z) \subseteq \overline{F}_1} \|\Phi z - y\|_2^2 \leq \|\Phi\overline{x} - y\|_2^2 + q_1.$$

Therefore, inequality (A10) always holds when $L = 1$.

Assume that the corollary holds at $L-1$ for some $L > 1$. This is, with

$$k^{'} = \sum_{j=1}^{L-1} \lceil |\overline{F}_j \setminus \Omega| \frac{1+\delta_\alpha}{1-\delta_{s\alpha}} \ln(2\mu) \rceil,$$

we have

$$\|r^{k^{'}}\|_2^2 \leq \|\Phi\overline{x} - y\|_2^2 + q_{L-1} + \mu^{-1} q_{L-2}.$$

Note that for $\mu = \sup_{j=1,2,\ldots,L-1} \frac{q_{j-1}}{q_j}$, we have

$$\|r^{k^{'}}\|_2^2 - \|\Phi\overline{x} - y\|_2^2 - q_L \leq q_{L-1} + \mu^{-1} q_{L-2} - q_L \leq 2q_{L-1}. \tag{A12}$$

Inequality (A11) implies that for $|\overline{F}_L \setminus I_{k'}| \neq 0$,

$$\|r^k\|_2^2 - \|\Phi\overline{x} - y\|_2^2 - q_L \leq \exp[-\frac{(1-\delta_{s\alpha})(k-k^{'})}{(1+\delta_\alpha)|\overline{F}_L \setminus I_{k'}|}] \max\{0, \|r^{k^{'}}\|_2^2 - \|\Phi\overline{x} - y\|_2^2 - q_L\}. \tag{A13}$$

Combining (A12) with (A13), we conclude that

$$\begin{aligned}
\|r^k\|_2^2 - \|\Phi\overline{x} - y\|_2^2 - q_L &\leq \exp[-\frac{(1-\delta_{s\alpha})(k-k^{'})}{(1+\delta_\alpha)|\overline{F}_L \setminus I_{k'}|}] 2q_{L-1} \\
&\leq \exp[-\frac{(1-\delta_{s\alpha})(k-k^{'})}{(1+\delta_\alpha)|\overline{F}_L \setminus \Omega|}] 2q_{L-1} \\
&\leq \mu^{-1} q_{L-1}.
\end{aligned}$$

Again, this inequality also holds when $|\overline{F}_L \setminus I_k| = 0$, since

$$\|r^k\|_2^2 \leq \min_{\mathrm{Gsupp}(z) \subseteq \overline{F}_L} \|\Phi z - y\|_2^2 \leq \|\Phi\overline{x} - y\|_2^2 + q_L.$$

We complete the proof of this corollary. □

The second and third ones are technical lemmas. For $F \subseteq \{1, \cdots, g\}$, set

$$x = \arg\min_{z:\mathrm{Gsupp(z)} \subseteq \mathrm{F}} \|\Phi z - y\|_2^2.$$

Lemma A2. *For all $s \geq |\overline{F} \setminus F|$, we have*

$$\|\Phi x - y\|_2^2 - \|\Phi\overline{x} - y\|_2^2 \leq 1.5(1+\delta_{s\alpha})\|\overline{x}_{\overline{F}\setminus F}\|_2^2 + 0.5\varepsilon_s(\overline{x})^2/(1+\delta_{s\alpha}).$$

Proof. Let $x^{'} = \overline{x}_{\overline{F}\cap F}$. Then $\|\Phi x - y\|_2^2 \leq \|\Phi x^{'} - y\|_2^2$. Therefore,

$$\begin{aligned} &\|\Phi x - y\|_2^2 - \|\Phi\overline{x} - y\|_2^2 \\ &\leq \|\Phi x^{'} - y\|_2^2 - \|\Phi\overline{x} - y\|_2^2 \\ &= \|\Phi x^{'} - \Phi\overline{x} + \Phi\overline{x} - y\|_2^2 - \|\Phi\overline{x} - y\|_2^2 \\ &\leq \|\Phi(x^{'} - \overline{x})\|_2^2 + |2\Big\langle\Phi(x^{'} - \overline{x}), \Phi\overline{x} - y\Big\rangle|. \end{aligned} \tag{A14}$$

We continue to estimate (A14). Note that $|\mathrm{Gsupp}(x^{'} - \overline{x})| \leq s$. By the definition of $\varepsilon_s(\overline{x})$ and the GRIP of Φ, we have

$$\begin{aligned} &\|\Phi x - y\|_2^2 - \|\Phi\overline{x} - y\|_2^2 \\ &\leq (1+\delta_{s\alpha})\|\overline{x}_{\overline{F}\setminus F}\|_2^2 + \varepsilon_s(\overline{x})\|\overline{x}_{\overline{F}\setminus F}\|_2 \\ &\leq (1+\delta_{s\alpha})\|\overline{x}_{\overline{F}\setminus F}\|_2^2 + 0.5\varepsilon_s(\overline{x})^2/(1+\delta_{s\alpha}) + 0.5(1+\delta_{s\alpha})\|\overline{x}_{\overline{F}\setminus F}\|_2^2 \\ &= 1.5(1+\delta_{s\alpha})\|\overline{x}_{\overline{F}\setminus F}\|_2^2 + 0.5\varepsilon_s(\overline{x})^2/(1+\delta_{s\alpha}), \end{aligned}$$

where we have used the mean value inequality in the second inequality. □

Lemma A3. *For all $s \geq |F \cup \overline{F}|$, we have*

$$(1-\delta_{s\alpha})\|x - \overline{x}\|_2^2 \leq 2(\|\Phi x - y\|_2^2 - \|\Phi\overline{x} - y\|_2^2) + \varepsilon_s(\overline{x})^2/(1-\delta_{s\alpha}).$$

Proof. Notice that

$$\begin{aligned} &\|\Phi x - y\|_2^2 - \|\Phi\overline{x} - y\|_2^2 \\ =&\|\Phi x - \Phi\overline{x} + \Phi\overline{x} - y\|_2^2 - \|\Phi\overline{x} - y\|_2^2 \\ =&\|\Phi x - \Phi\overline{x}\|_2^2 + \|\Phi\overline{x} - y\|_2^2 + 2\langle\Phi(x - \overline{x}), \Phi\overline{x} - y\rangle - \|\Phi\overline{x} - y\|_2^2 \\ \geq&\|\Phi(x - \overline{x})\|_2^2 - |2\langle\Phi(x - \overline{x}), \Phi\overline{x} - y\rangle|. \end{aligned} \tag{A15}$$

By the definition of $\varepsilon_s(\overline{x})$ and the GRIP of Φ, we derive

$$\begin{aligned} &\|\Phi x - y\|_2^2 - \|\Phi\overline{x} - y\|_2^2 \\ &\geq (1-\delta_{s\alpha})\|x - \overline{x}\|_2^2 - \varepsilon_s(\overline{x})\|\overline{x} - x\|_2 \\ &\geq (1-\delta_{s\alpha})\|x - \overline{x}\|_2^2 - 0.5\varepsilon_s(\overline{x})^2/(1-\delta_{s\alpha}) - 0.5(1-\delta_{s\alpha})\|x - \overline{x}\|_2^2 \\ &= 0.5(1-\delta_{s\alpha})\|x - \overline{x}\|_2^2 - 0.5\varepsilon_s(\overline{x})^2/(1-\delta_{s\alpha}), \end{aligned}$$

where we have used again the mean value inequality in the second inequality. □

Appendix B

Proof of Theorem 3. We prove this theorem by induction on $|\overline{F} \setminus \Omega|$. If $|\overline{F} \setminus \Omega| = 0$, then

$$\|r^0\|_2^2 = \min_{\tilde{x}:\mathrm{Gsupp}(\tilde{x}) \subseteq \Omega}\|y - \Phi\tilde{x}\|_2^2 \leq \|\Phi\overline{x} - y\|_2^2.$$

Assume that the inequality holds with $|\overline{F} \setminus \Omega| \leq p-1$ for some integer $p > 0$. Then we consider the case of $|\overline{F} \setminus \Omega| = p$. Without loss of generality, we assume for notational convenience that $\overline{F} \setminus \Omega = \{1, 2, \ldots, p\}$ and $\|\overline{x}_{G_j}\|_2$ for $j \in \overline{F} \setminus \Omega$ is arranged in descending order so that $\|\overline{x}_{G_1}\|_2 \geq \|\overline{x}_{G_2}\|_2 \geq \ldots \geq \|\overline{x}_{G_p}\|_2$. Let L be the smallest positive integer, such that for all $1 \leq l < L$, we have

$$\sum_{i=2^{l-1}}^{p} \|\overline{x}_{G_i}\|_2^2 < \mu \sum_{i=2^l}^{p} \|\overline{x}_{G_i}\|_2^2, \tag{A16}$$

but

$$\sum_{i=2^{L-1}}^{p} \|\overline{x}_{G_i}\|_2^2 \geq \mu \sum_{i=2^L}^{p} \|\overline{x}_{G_i}\|_2^2, \tag{A17}$$

where $\mu = 10(1+\delta_{p\alpha})/(1-\delta_{s\alpha})$. Note that $2^L \leq p$, so $L \leq \lceil \log_2 p \rceil + 1$. Moreover, if the second inequality is always satisfied for all $L \geq 1$, then we can simply take $L = 1$.

We define

$$\overline{F}_l = \{i : 1 \leq i \leq 2^l - 1\} \cup (\overline{F} \cap \Omega), \quad \text{for} \quad l = 0, 1, 2, \ldots, L.$$

Then $\overline{F} \cap \Omega = \overline{F}_0 \subseteq \overline{F}_1 \subseteq \overline{F}_2 \subseteq \ldots \subseteq \overline{F}_L \subseteq \overline{F}$. Lemma 3.2 implies that for $l = 0, 1, \ldots, L-1$,

$$\min_{x:\text{Gsupp(x)} \subseteq \overline{F}_l} \|\Phi x - y\|_2^2 \leq \|\Phi \overline{x} - y\|_2^2 + q_l,$$

where

$$\begin{aligned} q_{l-1} &= 1.5(1+\delta_{p\alpha})\|\overline{x}_{\overline{F} \setminus \overline{F}_{l-1}}\|_2^2 + 0.5\varepsilon_p(\overline{x})^2/(1+\delta_{p\alpha}) \\ &= 1.5(1+\delta_{p\alpha}) \sum_{i=2^{l-1}}^{p} \|\overline{x}_{G_i}\|_2^2 + 0.5\varepsilon_p(\overline{x})^2/(1+\delta_{p\alpha}). \end{aligned}$$

By (A16), we have

$$\begin{aligned} q_{l-1} &< 1.5(1+\delta_{p\alpha})\mu \sum_{i=2^l}^{p} \|\overline{x}_{G_i}\|_2^2 + 0.5\mu\varepsilon_p(\overline{x})^2/(1+\delta_{p\alpha}) \\ &= \mu q_l \end{aligned} \tag{A18}$$

for $p \geq |\overline{F} \setminus \overline{F}_l|$. Thus $\mu \geq \sup_{j=1,2,\ldots,L-1} \frac{q_{j-1}}{q_j}$ and $q_0 \geq q_1 \geq q_2 \geq \ldots \geq q_L \geq 0$. It follows from Corollary A1 that for

$$\begin{aligned} k &= \sum_{j=1}^{L} \lceil |\overline{F}_j \setminus \Omega| \frac{1+\delta_\alpha}{1-\delta_{s\alpha}} \ln(2\mu) \rceil \\ &\leq \sum_{j=1}^{L} (2^j - 1) \frac{1+\delta_\alpha}{1-\delta_{s\alpha}} \ln(2\mu) + L \\ &\leq 2^{L+1} \frac{1+\delta_\alpha}{1-\delta_{s\alpha}} \ln(2\mu) - (2+L) \frac{1+\delta_\alpha}{1-\delta_{s\alpha}} \ln(2\mu) + L \\ &\leq 2^{L+1} \frac{1+\delta_\alpha}{1-\delta_{s\alpha}} \ln(2\mu) - 1, \end{aligned}$$

and

$$s \geq |\overline{F} \cup I_k|,$$

we have

$$\|r^k\|_2^2 \leq \|\Phi \overline{x} - y\|_2^2 + q_L + \mu^{-1} q_{L-1}. \tag{A19}$$

Combining (A17) with (A18), we obtain

$$\begin{aligned}\|r^k\|_2^2 \le{}& \|\Phi\overline{x} - y\|_2^2 + 1.5(1+\delta_{p\alpha}) \sum_{i=2^L}^{p} \|\overline{x}_{G_i}\|_2^2 + 0.5\varepsilon_p(\overline{x})^2/(1+\delta_{p\alpha}) + \\ & 1.5\mu^{-1}(1+\delta_{p\alpha}) \sum_{i=2^{L-1}}^{p} \|\overline{x}_{G_i}\|_2^2 + 0.5\mu^{-1}\varepsilon_p(\overline{x})^2/(1+\delta_{p\alpha}) \\ \le{}& \|\Phi\overline{x} - y\|_2^2 + 1.5(1+\delta_{p\alpha})\mu^{-1} \sum_{i=2^{L-1}}^{p} \|\overline{x}_{G_i}\|_2^2 + 0.5\varepsilon_p(\overline{x})^2/(1+\delta_{p\alpha}) + \\ & 1.5\mu^{-1}(1+\delta_{p\alpha}) \sum_{i=2^{L-1}}^{p} \|\overline{x}_{G_i}\|_2^2 + 0.5\mu^{-1}\varepsilon_p(\overline{x})^2/(1+\delta_{p\alpha}) \\ ={}& \|\Phi\overline{x} - y\|_2^2 + 3\mu^{-1}(1+\delta_{p\alpha}) \sum_{i=2^{L-1}}^{p} \|\overline{x}_{G_i}\|_2^2 + 0.5(1+\mu^{-1})\varepsilon_p(\overline{x})^2/(1+\delta_{p\alpha}).\end{aligned} \tag{A20}$$

If $2\mu^{-1}(1+\delta_{p\alpha})\sum_{i=2^{L-1}}^{p} \|\overline{x}_{G_i}\|_2^2 \le (1+\mu^{-1})\varepsilon_s(\overline{x})^2/(1-\delta_{s\alpha})$, then

$$\begin{aligned}\|r^k\|_2^2 - \|\Phi\overline{x} - y\|_2^2 &\le 1.5(1+\mu^{-1})\varepsilon_s(\overline{x})^2/(1-\delta_{s\alpha}) + 0.5(1+\mu^{-1})\varepsilon_p(\overline{x})^2/(1+\delta_{p\alpha}) \\ &\le 2(1+\mu^{-1})\varepsilon_s(\overline{x})^2/(1-\delta_{s\alpha}) \\ &\le 2.5\varepsilon_s(\overline{x})^2/(1-\delta_{s\alpha}),\end{aligned}$$

where we have used $\mu \ge 10$ in the last inequality.

If $2\mu^{-1}(1+\delta_{p\alpha})\sum_{i=2^{L-1}}^{p} \|\overline{x}_{G_i}\|_2^2 > (1+\mu^{-1})\varepsilon_s(\overline{x})^2/(1-\delta_{s\alpha})$, then combining inequality (A20) with Lemma A3, we have

$$\begin{aligned}(1-\delta_{s\alpha})\|x^k - \overline{x}\|_2^2 &\le 2(\|r^k\|_2^2 - \|\Phi\overline{x} - y\|_2^2) + \varepsilon_s(\overline{x})/(1-\delta_{s\alpha}) \\ &\le 6\mu^{-1}(1+\delta_{p\alpha}) \sum_{i=2^{L-1}}^{p} \|\overline{x}_{G_i}\|_2^2 + (2+\mu^{-1})\varepsilon_s(\overline{x})^2/(1-\delta_{s\alpha}) \\ &< 10\mu^{-1}(1+\delta_{p\alpha}) \sum_{i=2^{L-1}}^{p} \|\overline{x}_{G_i}\|_2^2 \\ &= (1-\delta_{s\alpha}) \sum_{i=2^{L-1}}^{p} \|\overline{x}_{G_i}\|_2^2,\end{aligned}$$

which implies that

$$\sum_{i=p-|\overline{F}\setminus I_k|+1}^{p} \|\overline{x}_{G_i}\|_2^2 \le \sum_{i\in\overline{F}\setminus I_k} \|\overline{x}_{G_i}\|_2^2 \le \|x^k - \overline{x}\|_2^2 < \sum_{i=2^{L-1}}^{p} \|\overline{x}_{G_i}\|_2^2.$$

Therefore, $p - |\overline{F}\setminus I_k| + 1 > 2^{L-1}$, that is, $|\overline{F}\setminus I_k| \le p - 2^{L-1}$. Then we can take I_k for Ω and run the GOMP algorithm again. By induction, after another

$$k_1 = \lceil 4(p - 2^{L-1})\frac{1+\delta_\alpha}{1-\delta_{s\alpha}} \ln \frac{20(1+\delta_{|\overline{F}\setminus I_k|\alpha})}{1-\delta_{s\alpha}} \rceil$$

GOMP iterations, we have

$$\|r^{k+k_1}\|_2^2 \le \|\Phi\overline{x} - y\|_2^2 + 2.5\varepsilon_s(\overline{x})^2/(1-\delta_{s\alpha}),$$

while $s \geq |\overline{F} \cup I_k| + \lceil 4|\overline{F} \setminus I_k| \frac{1+\delta_\alpha}{1-\delta_{s\alpha}} \ln \frac{20(1+\delta_{|\overline{F}\setminus I_k|\alpha})}{1-\delta_{s\alpha}} \rceil$. Since

$$
\begin{aligned}
&|\overline{F} \cup I_k| + \lceil 4|\overline{F} \setminus I_k| \frac{1+\delta_\alpha}{1-\delta_{s\alpha}} \ln \frac{20(1+\delta_{|\overline{F}\setminus I_k|\alpha})}{1-\delta_{s\alpha}} \rceil \\
\leq &|\overline{F} \cup \Omega| + k + \lceil 4(p-2^{L-1}) \frac{1+\delta_\alpha}{1-\delta_{s\alpha}} \ln \frac{20(1+\delta_{|\overline{F}\setminus\Omega|\alpha})}{1-\delta_{s\alpha}} \rceil \\
\leq &|\overline{F} \cup \Omega| + 2^{L+1} \frac{1+\delta_\alpha}{1-\delta_{s\alpha}} \ln 2\mu - 1 + 4(p-2^{L-1}) \frac{1+\delta_\alpha}{1-\delta_{s\alpha}} \ln 2\mu + 1 \\
\leq &|\overline{F} \cup \Omega| + 4p \frac{1+\delta_\alpha}{1-\delta_{s\alpha}} \ln 2\mu \\
= &|\overline{F} \cup \Omega| + \lceil 4|\overline{F} \setminus \Omega| \frac{1+\delta_\alpha}{1-\delta_{s\alpha}} \ln \frac{20(1+\delta_{|\overline{F}\setminus\Omega|\alpha})}{1-\delta_{s\alpha}} \rceil,
\end{aligned}
$$

so for

$$
\begin{aligned}
k_0 &\geq \lceil 4p \frac{1+\delta_\alpha}{1-\delta_{s\alpha}} \ln \frac{20(1+\delta_{p\alpha})}{1-\delta_{s\alpha}} \rceil \\
&\geq 2^{L+1} \frac{1+\delta_\alpha}{1-\delta_{s\alpha}} \ln(2\mu) - 1 + \lceil 4(p-2^{L-1}) \frac{1+\delta_\alpha}{1-\delta_{s\alpha}} \ln \frac{20(1+\delta_{|\overline{F}\setminus I_k|\alpha})}{1-\delta_{s\alpha}} \rceil \\
&\geq k + k_1
\end{aligned}
$$

and

$$
s \geq |\overline{F} \cup \Omega| + 4|\overline{F} \setminus \Omega| \frac{1+\delta_\alpha}{1-\delta_{s\alpha}} \ln \frac{20(1+\delta_{|\overline{F}\setminus\Omega|\alpha})}{1-\delta_{s\alpha}},
$$

we have

$$
\|r^{k_0}\|^2 \leq \|\Phi\overline{x} - y\|^2 + 2.5\varepsilon_s(\overline{x})^2/(1-\delta_{s\alpha}).
$$

This finishes the induction step for the case $|\overline{F} \setminus \Omega| = p$.

For the second inequality, combining the first inequality with Lemma A3, we have

$$
\begin{aligned}
(1-\delta_{s\alpha})\|x^k - \overline{x}\|_2^2 &\leq 2(\|\Phi x^k - y\|_2^2 - \|\Phi\overline{x} - y\|_2^2) + \varepsilon_s(\overline{x})^2/(1-\delta_{s\alpha}) \\
&\leq 6\varepsilon_s(\overline{x})^2/(1-\delta_{s\alpha}).
\end{aligned}
$$

Thus we complete the proof of this theorem. □

When Ω is the empty set, we have the following corollary of Theorem 3.

Corollary A2. *If the GRIP condition $2\delta_{31|\overline{F}|\alpha} + \delta_{|\overline{F}|\alpha} \leq 1$ holds, then we have for $k = 30|\overline{F}|$,*

$$
\|r^k\|_2^2 \leq \|\Phi(\overline{x}) - y\|_2^2 + 2.5\varepsilon_s(\overline{x})^2/(1-\delta_{s\alpha})
$$

and

$$
\|x^k - \overline{x}\|_2 \leq 2\sqrt{6}\sqrt{1+\delta_{s\alpha}}\|\Phi\overline{x} - y\|_2/(1-\delta_{s\alpha}),
$$

where $s = 31|\overline{F}|$.

Proof. If $2\delta_{31|\overline{F}|\alpha} + \delta_{|\overline{F}|\alpha} \leq 1$, then we have

$$
\frac{1+\delta_\alpha}{1-\delta_{31|\overline{F}|\alpha}} \leq \frac{1+\delta_{|\overline{F}|\alpha}}{1-\delta_{31|\overline{F}|\alpha}} \leq 2.
$$

Therefore

$$s = 31|\overline{F}| = |\overline{F}| + 30|\overline{F}| \geq |\overline{F}| + 4|\overline{F}| \cdot 2\ln(20 \cdot 2)$$
$$\geq |\overline{F}| + 4|\overline{F}| \frac{1+\delta_\alpha}{1-\delta_{s\alpha}} \ln 20 \frac{1+\delta_{|\overline{F}|\alpha}}{1-\delta_{s\alpha}}. \tag{A21}$$

Combining (A21) with Theorem 3, we obtain the first inequality of Corollary A2 and

$$\|x^k - \overline{x}\|_2 \leq \sqrt{6}\varepsilon_s(\overline{x})/(1-\delta_{s\alpha}). \tag{A22}$$

Notice that for any $\mu \in \mathbb{R}^n$ with $|\mathrm{Gsupp}(\mu)| \leq s$, we have

$$|\langle 2\Phi\mu, \Phi\overline{x} - y\rangle| \leq 2\|\Phi\mu\|_2 \cdot \|\Phi\overline{x} - y\|_2 \leq 2\sqrt{1+\delta_{s\alpha}}\|\mu\|_2 \cdot \|\Phi\overline{x} - y\|_2,$$

and hence

$$\varepsilon_s(\overline{x}) \leq 2\sqrt{1+\delta_{s\alpha}}\|\Phi\overline{x} - y\|_2. \tag{A23}$$

Substituting (A23) into (A22), we obtain the second inequality of Corollary A2. The proof of Corollary A2 is completed. □

References

1. Foucart, S.; Rauhut, H. *A Mathematical Introduction to Compressive Sensing*; Springer: Berlin/Heidelberg, Germany, 2013.
2. Cai, T.; Zhang, A. Sparse representation of a polytope and recovery of sparse signals and low-rank matrices. *IEEE Trans. Inf. Theory* **2014**, *60*, 122–132.
3. Günther, M.; Böhringer, S.; Wieczorek, D.; Würtz, R.P. Reconstruction of images from Gabor graphs with applications in facial image processing. *Int. J. Wavelets Multiresolut. Inf. Process.* **2015**, *13*, 1550019. https://doi.org/10.1142/S0219691315500198.
4. Ye, P.X.; Wei, X.J. Efficiency of weak greedy algorithms for m-term approximations. *Sci. China. Math.* **2016**, *59*, 697–714.
5. Yang, B.; Yang, C.; Huang, G. Efficient image fusion with approximate sparse representation. *Int. J. Wavelets Multiresolut. Inf. Process.* **2016**, *14*, 1650024. https://doi.org/10.1142/s0219691316500247.
6. Candès, E.; Kalton, N.J. The restricted isometry property and its implications for compressed sensing. *C. R. Acad. Sci. Sér. I* **2008**, *346*, 589–592.
7. Candès, E.J.; Tao, T. Decoding by linear programming. *IEEE Trans. Inf. Theory* **2005**, *51*, 4203–4215.
8. Yuan, M.; Lin, Y. Model selection and estimation in regression with grouped variables. *J. R. Stat. Ser. B* **2006**, *68*, 49–67.
9. Ou, W.; Hämäläinen, M.S.; Golland, P. A distributed spatio-temporal EEG/MEG inverse solver. *NeuroImage* **2009**, *44*, 932–946.
10. Antoniadis, A.; Fan, J. Regularization of wavelet approximations. *J. Amer. Stat. Assoc.* **2001**, *96*, 939–967.
11. He, Z.; Yu, W. Stable feature selection for biomarker discovery. *Comput. Biol. Chem.* **2010**, *34*, 215–225.
12. Mishali, M.; Eldar, Y.C. Blind multiband signal reconstruction: Compressed sensing for analog signals. *IEEE Trans. Signal Process.* **2009**, *57*, 993–1009.
13. Ahsen, M.E.; Vidyasagar, M. Error bounds for compressed sensing algorithms with group sparsity: A unified approach. *Appl. Comput. Harmon. Anal.* **2017**, *42*, 212–232.
14. Ranjan, S.; Vidyasagar, M. Tight performance bounds for compressed sensing with conventional and group sparsity. *IEEE Trans. Signal Process* **2019**, *67*, 2854–2867.
15. Tropp, J.A. Greedy is good: Algorithmic results for sparse approximation. *IEEE Trans. Inform. Theory* **2004**, *50*, 2231–2242.
16. Lin, J.H.; Li, S. Nonuniform support recovery from noisy measurements by orthogonal matching pursuit. *J. Approx. Theory* **2013**, *165*, 20–40.
17. Mo, Q.; Shen, Y. A remark on the restricted isometry property in orthogonal matching pursuit. *IEEE Trans. Inf. Theory* **2012**, *58*, 3654–3656.
18. Dan, W. Analysis of orthogonal multi-matching pursuit under restricted isometry property. *Sci. China Math.* **2014**, *57*, 2179–2188.
19. Xu, Z.Q. The performance of orthogonal multi-matching pursuit under RIP. *J. Comp. Math.* **2015**, *33*, 495–516.
20. Shao, C.F.; Ye, P.X. Almost optimality of orthogonal super greedy algorithms for incoherent dictionaries. *Int. J. Wavelets Multiresolut. Inf. Process.* **2017**, *15*, 1750029. https://doi.org/10.1142/S0219691317500291.
21. Wei, X.J.; Ye, P.X. Efficiency of orthogonal super greedy algorithm under the restricted isometry property. *J. Inequal. Appl.* **2019**, *124*, 1–21.
22. Zhang, T. Sparse recovery with orthogonal matching pursuit under RIP. *IEEE Trans. Inf. Theory* **2011**, *57*, 6215–6221.
23. Xu, Z.Q. A remark about orthogonal matching pursuit algorithm. *Adv. Adapt. Data Anal.* **2012**, *04*, 1250026. https://doi.org/10.1142/S1793536912500264.
24. Cohen, A.; Dahmen, W.; DeVore, R. Orthogonal matching pursuit under the restricted isometry property. *Constr. Approx.* **2017**, *45*, 113–127.

25. Eldar, Y.C.; Mishali, M. Robust recovery of signals from a structured union of subspaces. *IEEE Trans. Inf. Theory* **2009**, *55*, 5302–5316.
26. Wright, J.; Yang, A.Y.; Ganesh, A.; Sastry, S.S.; Ma, Y. Robust face recognition via sparse representation, *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 210–227.
27. Van Den Berg, E.; Friedlander, M.P. Theoretical and empirical results for recovery from multiple measurements. *IEEE Trans. Inf. Theory* **2010**, *56*, 2516–2527.
28. Eldar, Y.C.; Kuppinger, P.; Bölcskei, H. Block-sparse signals: Uncertainty relations and efficient recovery. *IEEE Trans. Signal Process.* **2010**, *58*, 3042–3054.
29. Lai, M.J.; Liu, Y. The null space property for sparse recovery from multiple measurement vectors. *Appl. Comput. Harmon. Anal.* **2011**, *30*, 402–406.
30. Zhang, J.; Wang, J.; Wang, W. A perturbation analysis of block-sparse compressed sensing via mixed l_2/l_1 minimization. *Int. J. Wavelets, Multiresolut. Inf. Process.* **2016**, *14*, 1650026. https://doi.org/10.1142/S0219691316500260.
31. Chen, W.G.; Ge, H.M. A sharp recovery condition for block sparse signals by block orthogonal multi-matching pursuit. *Sci. China Math.* **2017**, *60*, 1325–1340.
32. Wen, J.M.; Zhou, Z.C.; Liu, Z.L.; Lai, M.J.; Tang, X.H. Sharp sufficient conditions for stable recovery of block sparse signals by block orthogonal matching pursuit. *Appl. Comput. Harmon. Anal.* **2019**, *47*, 948–974.
33. Candès, E.; Romberg, J. l_1-Magic: Recovery of Sparse Signals via Convex Programming. Available online: https://candes.su.domains/software/l1magic/#code (accessed on 15 February 2023).
34. Do, T.T.; Gan, L.; Nguyen, N.; Tran, T.D. Sparsity adaptive matching pursuit algorithm for practical compressed sensing. In Proceedings of the 2008 42nd Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 26–29 October 2008; pp. 581–587.
35. Bi, X.; Leng, L.; Kim, C.; Liu, X.; Du, Y.; Liu, F. Constrained Backtracking Matching Pursuit Algorithm for Image Reconstruction in Compressed Sensing. *Appl. Sci.* **2021**, *11*, 1435. https://doi.org/10.3390/app11041435.
36. Wang, R.; Qin, Y.; Wang, Z.; Zheng, H. Group-Based Sparse Representation for Compressed Sensing Image Reconstruction with Joint Regularization. *Electronics* **2022**, *11*, 182. https://doi.org/10.3390/electronics11020182.