

# Sparse Signal Recovery via Rescaled Matching Pursuit

Wan Li and Peixin Ye \*

Department of Applied Mathematics, School of Mathematical Sciences and LPMC, Nankai University, Tianjin 300071, China; 1120210026@mail.nankai.edu.cn

\* Correspondence: yepx@nankai.edu.cn

**Abstract:** We propose the Rescaled Matching Pursuit (RMP) algorithm to recover sparse signals in high-dimensional Euclidean spaces. The RMP algorithm has less computational complexity than other greedy-type algorithms, such as Orthogonal Matching Pursuit (OMP). We show that if the restricted isometry property is satisfied, then the upper bound of the error between the original signal and its approximation can be derived. Furthermore, we prove that the RMP algorithm can find the correct support of sparse signals from random measurements with a high probability. Our numerical experiments also verify this conclusion and show that RMP is stable with the noise. So, the RMP algorithm is a suitable method for recovering sparse signals.

**Keywords:** signal recovery; rescaled matching pursuit; restricted isometry property; computational complexity; admissible matrix

**MSC:** 41A65; 41A25; 94A15

## 1. Introduction

Compressed sensing (or compressive sensing) [1] has become a very active research direction in the field of signal processing, after the pioneering work by Candès and Tao (see [2]). It has been successfully applied in image compression, signal processing, medical imaging, and computer science (see [3–7]). Compressed sensing refers to the problem of recovering sparse signals from low dimensional measurements (see [8]). In terms of image processing, one needs to compress the image first, then the image signal becomes sparse and its reconstruction belongs to the category of compressed sensing.

Let  $\mathbb{R}^d$  denote the  $d$ -dimensional Euclidean space. Consider a signal  $x \in \mathbb{R}^d$ . We say  $x$  is a  $k$ -sparse vector if  $\#\text{supp}(x) \leq k$ . Here, we use  $\text{supp}(x)$  to denote the set of index such that  $x_i \neq 0$  for all  $i \in \text{supp}(x)$  and  $\#A$  represents the cardinality of a set  $A$ . By the compressed sensing theory, the recovery of a sparse signal can be mathematically described as

$$f = \Phi x \quad (1)$$

where  $x$  is a  $d$ -dimensional  $k$ -sparse vector, and  $\Phi \in \mathbb{R}^{N \times d}$  is the measurement matrix with  $N \ll d$ . The goal of signal recovery is establishing an algorithm to approximate the unknown signal  $x$  by using the measurement matrix  $\Phi$  and given data  $f = \Phi x$ .

A well-known method in compressed sensing is basis pursuit (BP), which is also known as  $\ell_1$  minimization (see [9]). It is defined as

$$\hat{x} = \arg \min_{x \in \mathbb{R}^d} \{\|x\|_{\ell_1} : \text{subject to } f = \Phi x\} \quad (2)$$

By imposing a “restricted orthonormality hypothesis”, which is far weaker than assuming orthonormality and defined as the following Definition 1, the  $\ell_1$  minimization program can recover  $x$  exactly. Furthermore, the target sparse signal  $x$  could be recovered with a high probability when  $\Phi$  is a Gaussian random matrix.



**Citation:** Li, W.; Ye, P. Sparse Signal Recovery via Rescaled Matching Pursuit. *Axioms* **2024**, *13*, 288. <https://doi.org/10.3390/axioms13050288>

Academic Editor: J. Alberto Conejero

Received: 16 March 2024

Revised: 7 April 2024

Accepted: 21 April 2024

Published: 24 April 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Definition 1.** A matrix  $\Phi$  is said to satisfy the restricted isometry property at sparsity level  $k$  with the constant  $\delta_k \in (0, 1)$  if

$$(1 - \delta_k)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta_k)\|x\|_2^2$$

holds for all  $k$ -sparse vectors  $x$ . And, we say that the matrix  $\Phi$  satisfies the restricted isometry property with parameters  $(k, \delta)$  if  $\delta_k < \delta$ .

The  $\ell_1$  minimization method and related algorithms found by solving (2) have been well studied, and they range widely in effectiveness and complexity (see [1,2,10–12]). It is known that a linear program problem can be solved in polynomial time by using optimization software. However, it still takes a long time if the length of the signal is large.

Compared with the BP algorithm, greedy-type algorithms are easy to implement and have less computational complexity and fast convergence speed. Thus, greedy-type algorithms are powerful methods for the recovery of signals and are increasingly used in that field; for example, see [13–17]. To recall this type of algorithm, we first recall some basic notions on sparse approximation in a Hilbert space  $H$ . We denote the inner product of  $H$  by  $\langle \cdot, \cdot \rangle$ . The norm  $\|\cdot\|$  on  $H$  is defined by  $\|f\| = \langle f, f \rangle^{1/2}$  for all  $f \in H$ . We call a set  $\mathcal{D} \subset H$  a dictionary if the closure of  $\text{span}(\mathcal{D})$  is  $H$ . Here  $\text{span}(\mathcal{D})$  is the linear space spanned by  $\mathcal{D}$ . We only consider normalized dictionaries, that is, for any  $\varphi \in \mathcal{D}$ ,  $\|\varphi\| = 1$ . Let  $\Sigma_m(\mathcal{D})$  denote the set of all elements which can be expressed as a linear combination of  $m$  elements from  $\mathcal{D}$

$$\Sigma_m(\mathcal{D}) := \left\{ g : g = \sum_{\varphi \in \Lambda} c_\varphi \varphi, \varphi \in \mathcal{D}, \#\Lambda \leq m \right\}$$

We define the minimal error of the approximation of an element  $f \in H$  as  $g \in \Sigma_m(\mathcal{D})$

$$\sigma_m(f) := \sigma_m(f, \mathcal{D}) = \inf_{g \in \Sigma_m(\mathcal{D})} \|f - g\|$$

Let a dictionary  $\mathcal{D}$  be given. We define the collection of elements by

$$\mathcal{A}_1^o(\mathcal{D}, M) := \left\{ f \in H : f = \sum_{k \in \Lambda} c_k(f) \varphi_k, \varphi_k \in \mathcal{D}, \#\Lambda \leq \infty, \sum_{k \in \Lambda} |c_k(f)| \leq M \right\}$$

Then, we define  $\mathcal{A}_1(\mathcal{D}, M)$  to be the closure of  $\mathcal{A}_1^o(\mathcal{D}, M)$  in  $H$ . Finally, we define the linear space  $\mathcal{A}_1(\mathcal{D})$  by

$$\mathcal{A}_1(\mathcal{D}) := \bigcup_{M>0} \mathcal{A}_1(\mathcal{D}, M)$$

and we equip  $\mathcal{A}_1(\mathcal{D})$  with the norm

$$\|f\|_{\mathcal{A}_1(\mathcal{D})} := \inf\{M : f \in \mathcal{A}_1(\mathcal{D}, M)\}$$

for all  $f \in \mathcal{A}_1(\mathcal{D})$ .

The simplest greedy algorithm in  $H$  is known as the Pure Greedy Algorithm (PGA( $H, \mathcal{D}$ )). We recall its definition from [18].

**Pure Greedy Algorithm (PGA( $H, \mathcal{D}$ ))**

**Step 1:** Set  $f_0 = 0$ .

**Step 2:**

-If  $f = f_{m-1}$ , then stop and define  $f_l = f_{m-1} = f$  for  $l \geq m$ .

-If  $f \neq f_{m-1}$ , then choose an element  $\varphi_m \in \mathcal{D}$  satisfying

$$|\langle f - f_{m-1}^{PGA}, \varphi_m \rangle| = \sup_{\varphi \in \mathcal{D}} |\langle f - f_{m-1}^{PGA}, \varphi \rangle|$$

Define the new approximation of  $f$  as

$$f_m^{PGA} = \sum_{i=1}^m \langle f - f_{m-1}^{PGA}, \varphi_i \rangle \varphi_i$$

and proceed to Step 2.

In [19], DeVore and Temlyakov proposed the Orthogonal Greedy Algorithm (OGA( $H, \mathcal{D}$ )) in  $H$  with a dictionary  $\mathcal{D}$ . We recall its definition.

**Orthogonal Greedy Algorithm (OGA( $H, \mathcal{D}$ ))**

**Step 1:** Set  $f_0 = 0$ .

**Step 2:**

-If  $f = f_{m-1}$ , then stop and define  $f_l = f_{m-1} = f$  for  $l \geq m$ .

-If  $f \neq f_{m-1}$ , then choose an element  $\varphi_m \in \mathcal{D}$  satisfying

$$|\langle f - f_{m-1}^{OGA}, \varphi_m \rangle| = \sup_{\varphi \in \mathcal{D}} |\langle f - f_{m-1}^{OGA}, \varphi \rangle|$$

Define the new approximation of  $f$  as

$$f_m^{OGA} = P_m(f)$$

where  $P_m(f)$  is the best approximation of  $f$  from the linear space  $V_m := \text{span}\{\varphi_1, \varphi_2, \dots, \varphi_m\}$ , and proceed to Step 2.

In fact, the OGA is one of the modifications of the PGA, improving the convergence rate. The convergence rate is one of the critical issues for the approximation power of algorithms. If  $f_m$  is the approximation of the target element  $f$  by applying a greedy algorithm after  $m$  iterations, the efficiency of the algorithm can be measured by the decay of the sequence  $\{\|f - f_m\|\}$ . It has been shown that the PGA and the OGA can achieve the convergence rates  $m^{-1/6}$  and  $m^{-1/2}$  for  $\|f - f_m\|$ , respectively, if the target element  $f$  belongs to  $\mathcal{A}_1(\mathcal{D}, M)$  (see [19]).

**Theorem 1.** Suppose  $\mathcal{D}$  is a dictionary of  $H$ . Then, for every  $f \in \mathcal{A}_1(\mathcal{D}, M)$ , the following inequality holds:

$$\|f - f_m^{PGA}\| \leq Mm^{-1/6}, \quad m = 0, 1, 2, \dots$$

**Theorem 2.** Suppose  $\mathcal{D}$  is a dictionary of  $H$ . Then, for every  $f \in \mathcal{A}_1(\mathcal{D}, M)$ , the following inequality holds:

$$\|f - f_m^{OGA}\| \leq Mm^{-1/2}, \quad m = 0, 1, 2, \dots$$

It is known from [20] that the convergence rate  $m^{-1/2}$  cannot be improved. Thus, as a very general result for all dictionaries, the convergence rate is optimal.

The fundamental problem of signal recovery is to find out the support of the original target signal by using the model (1). Since  $N$  is much smaller than the dimension  $d$ , the column vectors of the measurement matrix  $\Phi$  are linearly dependent. Therefore, these vectors consist of a redundant system. This redundant system may be considered as a dictionary of  $\mathbb{R}^N$ . It is well known that  $\mathbb{R}^N$  is a Hilbert space with the usual inner product

$$\langle x, y \rangle = \sum_{j=1}^N x_j y_j, \quad \forall x \in \mathbb{R}^N, y \in \mathbb{R}^N$$

Thus, signal recovery can be considered as an approximation problem with a dictionary in a Hilbert space. In particular, the approximation of the original signal is the solution of the following minimization problem:

$$\min_{\hat{x}_1, \dots, \hat{x}_d \in \mathbb{R}} \left\| f - \sum_{i=1}^d \hat{x}_i \varphi_i \right\|_2$$

where  $\varphi_1, \dots, \varphi_d$  denote the columns of  $\Phi$ . Naturally, one can adapt greedy approximation algorithms to handle the signal recovery problem. In CS, one such method is Orthogonal Matching Pursuit (OMP) (see [21,22]), which is an application of OGA to the field of signal processing. Compared with the BP, the OMP is faster and it admits an ease of implementation (see [17,23]). We recall the definition of the OMP from [23].

**Orthogonal Matching Pursuit (OMP)**

**Input:** An  $N \times d$  matrix  $\Phi$ ,  $f \in \mathbb{R}^N$ , the sparsity index  $k$ .

**Step 1:** Set  $r_0 = f$ , the index set  $\Lambda_0 = \emptyset$ , and  $m = 1$ .

**Step 2:** Define  $\Lambda_m := \Lambda_{m-1} \cup \{i_m\}$  such that

$$|\langle r_{m-1}, \varphi_{i_m} \rangle| = \sup_{\varphi \in \Phi} |\langle r_{m-1}, \varphi \rangle|$$

Then,

$$x_m = \arg \min_{z: \text{supp}(z) \in \Lambda_m} \|f - \Phi z\|_2, \quad f_m^o = \Phi x_m$$

and

$$r_m = f - f_m^o = f - \Phi x_m$$

**Step 3:** Increment  $m$ , and return to Step 2 if  $\#\Lambda_m < k$ .

**Output:** If  $\#\Lambda_m \geq k$ , then output  $\Lambda_m$  and  $\hat{x} = x_m$ .

It is clear from the available literature that the OMP algorithm is the most popular greedy-type algorithm being used for signal recovery. The OMP algorithm, with a high probability, can exactly recover each sparse signal by using a Gaussian or Bernoulli matrix (see [17]). And, it is shown that the OMP algorithm can stably recover sparse signals in  $l^2$  under measurement noise (see [24]). Although the OMP algorithm is powerful, the computational complexity of the orthogonality step is quite high, especially for large-scale problems. See [25] for the detailed computation and the storage cost of the OMP. So, the main disadvantage of the OMP is its high computational cost. In particular, when the sparsity level  $k$  of the target signal is large, OMP may not be a good choice, since the cost of orthogonalization increases quadratically with the number of iterations. Thus, in this case, it is natural to look for other greedy-type algorithms to reduce the computational cost.

In Section 2, we propose an algorithm called Rescaled Matching Pursuit (RMP) for signal recovery. This algorithm has less computational complexity than the OMP, which means it can save resources such as storage space and computation time. In Section 3, we analyze the efficiency of the RMP algorithm under the RIP condition. In Section 4, we prove that the RMP algorithm can be used to find the correct support of sparse signal from a random measurement matrix with a high probability. In Section 5, we use numerical experiments to verify this conclusion. In Section 6, we summarize our results.

**2. Rescaled Matching Pursuit Algorithm**

In [26], Petrova presented another modification of the PGA which is called the rescaled pure greedy algorithm (RPGA). This modification is very simple. It simply rescales the approximation in each iteration of the PGA.

We recall the definition of the RPGA( $H, \mathcal{D}$ ) from [26].

**Rescaled Pure Greedy Algorithm (RPGA( $H, \mathcal{D}$ ))**

**Step 1:** Set  $f_0 = 0$ ,  $r_0 = f$ .

**Step 2:**

- If  $f = f_{m-1}$ , then stop and define  $f_k = f_{m-1} = f$  for  $k \geq m$ .
- If  $f \neq f_{m-1}$ , then choose an element  $\varphi_m \in \mathcal{D}$  satisfying

$$|\langle r_{m-1}, \varphi_m \rangle| = \sup_{\varphi \in \mathcal{D}} |\langle r_{m-1}, \varphi \rangle|$$

with

$$\lambda_m = \langle r_{m-1}, \varphi_m \rangle, \quad \widehat{f}_m := f_{m-1} + \lambda_m \varphi_m, \quad s_m = \frac{\langle f, \widehat{f}_m \rangle}{\|\widehat{f}_m\|_2^2}$$

Then, define the the new approximation of  $f$  as

$$f_m = s_m \widehat{f}_m, \quad r_m = f - f_m$$

and proceed to Step 2.

The following theorem on the convergence rate of the RPGA was also obtained in [26].

**Theorem 3.** *Suppose  $\mathcal{D}$  is a dictionary of  $H$ . Then, for every  $f \in \mathcal{A}_1(\mathcal{D})$ , the RPGA satisfies the following inequality:*

$$\|f - f_m\| \leq \|f\|_{\mathcal{A}_1(\mathcal{D})} (m + 1)^{-1/2}, \quad m = 0, 1, 2, \dots$$

Notice that the RPGA only needs to solve a one-dimensional optimization problem at each step. Thus, the RPGA is simpler than the OGA. Together with Theorem 3, it is known that the RPGA is a greedy algorithm with minimal computational complexity to achieve the optimal convergence rate  $m^{-1/2}$ . Based on this result, the RPGA has been applied successfully to solve the problems of convex optimization and regression, see [27–29]. In solving these two types of problems, the RPGA performs better than the OGA. The main reason is that the RPGA has smaller computational cost than the OGA. For more detailed results about the RPGA, one can refer to [26–30]. Besides, in [31] the authors propose the Super Rescaled Pure Greedy Learning Algorithm and investigate its behavior. The success of the RPGA inspires us to consider its application in the field of signal recovery.

In the present article, we will design an algorithm based on the RPGA to recover sparse signals and analyze its performance. Suppose that  $x$  is a  $k$ -sparse signal in  $\mathbb{R}^d$  and  $\Phi$  is an  $N \times d$  matrix. We denote the columns of  $\Phi$  as  $\varphi_1, \dots, \varphi_d$ . Then, we design the Rescaled Matching Pursuit (RMP) for sparse signal recovery.

**Rescaled Matching Pursuit (RMP)**

**Input:** Measurement matrix  $\Phi$ , vector  $f$ , the sparsity index  $k$ .

**Step 1:** Set  $r_0 = f, f_0 = 0$ , the index set  $\Lambda_0 = \emptyset$ , and  $m = 1$ .

**Step 2:** Define  $\Lambda_m := \Lambda_{m-1} \cup \{i_m\}$  such that

$$|\langle r_{m-1}, \varphi_{i_m} \rangle| = \sup_{\varphi \in \Phi} |\langle r_{m-1}, \varphi \rangle|$$

Then,

$$f_m = s_m \widehat{f}_m$$

where  $\theta_m = \langle r_{m-1}, \varphi_{i_m} \rangle, \widehat{f}_m = f_{m-1} + \theta_m \varphi_{i_m}$ , and  $s_m = \frac{\langle f, \widehat{f}_m \rangle}{\|\widehat{f}_m\|_2^2}$ . Next, update the residual

$$r_m = f - f_m$$

**Step 3:** Increment  $m$ , and return to Step 2 if  $\#\Lambda_m < k$ .

**Output:** If  $\#\Lambda_m \geq k$ , then output  $\Lambda_m$  and  $\hat{x}$ . It is easy to see that  $\hat{x}$  has nonzero indices at the components listed in  $\Lambda_m$ . If  $i_{m_1} = i_{m_2} = \dots = i_{m_s} = i_j$ , then we can figure out the value of  $\hat{x}$  in component  $i_j$  through

$$\hat{x}(i_j) = \sum_{l=1}^s \theta_{m_l} \prod_{i=m_l}^k s_i$$

The main procedure of the algorithm can be divided into selecting the columns of  $\Phi$  and constructing the approximation  $f_m$ . In Step 2, the approximation  $f_m$  is obtained by solving a one-dimensional optimization problem. In fact,  $f_m$  is an orthogonal projection of  $f$  onto the one-dimensional space  $\text{span}\{\hat{f}_m\}$ . The RMP algorithm is broadly similar to the OMP algorithm, while the computational complexity differs. For the OMP algorithm, the approximation  $f_m^o := \Phi\hat{x}$  is obtained by solving the  $m$ -dimensional least squares problem at  $m$  iteration in Step 2, which has a total cost of  $O(mN)$ . For the RMP algorithm, as argued above, the approximation  $f_m$  is obtained by solving a one-dimensional optimization problem at the  $m$  iteration in Step 2, which has a total cost of  $O(N)$ . It is not difficult to see that the computational cost is less expensive than most existing greedy algorithms, such as the OMP algorithm. That is, we can save a lot of resources such as storage space and computation time by using the RMP algorithm.

### 3. Efficiency of the RMP under the Restricted Isometry Property

Suppose that the matrix  $\Phi$  satisfies the RIP condition which is defined in Definition 1. Then, we study the performance of the RMP for  $k$ -sparse signal recovery.

Under the RIP, we obtain the convergence rate of the error between the given data vector  $f$  and its approximation  $f_m$ .

**Theorem 4.** Assume that  $x$  is an arbitrary  $k$ -sparse signal in  $\mathbb{R}^d$  and the measurement matrix  $\Phi \in \mathbb{R}^{N \times d}$  satisfies the RIP condition with parameters  $(k, \delta)$ . Then, for  $f = \Phi x$  and every positive integer  $m$ , we have

$$\|f - f_m\|_2 \leq \frac{k^{1/2} \|f\|_2}{(1 - \delta)} (m + 1)^{-1/2}, \quad m = 0, 1, \dots$$

Moreover, we derive the upper bound of the difference between the original signal  $x$  and the estimate  $\hat{x}$  of  $x$ .

**Theorem 5.** Assume that  $x$  is an arbitrary  $k$ -sparse signal in  $\mathbb{R}^d$  and the measurement matrix  $\Phi \in \mathbb{R}^{N \times d}$  satisfies the RIP condition with parameters  $(2k, \delta)$ . Given the data  $f = \Phi x$ , if  $\#\Lambda_L = k$ ,  $L \geq k$ , we have

$$\|x - \hat{x}\|_2 \leq \frac{k^{1/2} \|f\|_2}{(1 - \delta)^{3/2}} (L + 1)^{-1/2}$$

To prove Theorems 4 and 5, we will use a lemma which was obtained in [32].

**Lemma 1.** Let the positive constants  $\ell$ ,  $r$ , and  $B$  be given. Suppose  $\{a_m\}$  and  $\{r_m\}$  are finite or infinite sequences of positive real numbers such that

$$a_j \leq B, \quad a_m \leq a_{m-1} \left( 1 - \frac{r_m}{r} a_{m-1}^\ell \right), \quad m = J + 1, J + 2, \dots$$

Then,  $a_m$  has the following upper bound:

$$a_m \leq \max\{1, \ell^{-1/\ell}\} r^{1/\ell} \left( rB^{-\ell} + \sum_{k=J+1}^m r_k \right)^{-1/\ell}, \quad m = J + 1, J + 2, \dots$$

**Proof of Theorem 4:** It follows from the definitions of  $\widehat{f}_m$  and  $\theta_m$  that

$$\begin{aligned} \|f - \widehat{f}_m\|_2^2 &= \langle f - \widehat{f}_m, f - \widehat{f}_m \rangle \\ &= \langle f - f_{m-1} - \theta_m \varphi_{i_m}, f - f_{m-1} - \theta_m \varphi_{i_m} \rangle \\ &= \|f - f_{m-1}\|_2^2 - 2\theta_m \langle f - f_{m-1}, \varphi_{i_m} \rangle + \theta_m^2 \|\varphi_{i_m}\|_2^2 \\ &= \|f - f_{m-1}\|_2^2 - 2\langle f - f_{m-1}, \varphi_{i_m} \rangle^2 + \|\varphi_{i_m}\|_2^2 \langle f - f_{m-1}, \varphi_{i_m} \rangle^2 \end{aligned}$$

From the design of the RMP, we know that  $f_m$  is the orthogonal projection of  $f$  onto  $\text{span}\{\widehat{f}_m\}$ . Thus, the following must be true:

$$\begin{aligned} \|f - f_m\|_2^2 &\leq \|f - \widehat{f}_m\|_2^2 \\ &= \|f - f_{m-1}\|_2^2 - 2\langle f - f_{m-1}, \varphi_{i_m} \rangle^2 + \|\varphi_{i_m}\|_2^2 \langle f - f_{m-1}, \varphi_{i_m} \rangle^2 \end{aligned}$$

We continue to estimate the approximation error of  $f_m$  by using the RIP condition with parameters  $(k, \delta)$  of the matrix  $\Phi$ . Hence, the above inequality can be estimated as

$$\begin{aligned} \|f - f_m\|_2^2 &\leq \|f - f_{m-1}\|_2^2 - 2\langle f - f_{m-1}, \varphi_{i_m} \rangle^2 + (1 + \delta) \langle f - f_{m-1}, \varphi_{i_m} \rangle^2 \\ &= \|f - f_{m-1}\|_2^2 - (1 - \delta) \langle f - f_{m-1}, \varphi_{i_m} \rangle^2 \end{aligned} \tag{3}$$

Now, we will derive a lower bound for  $|\langle f - f_{m-1}, \varphi_{i_m} \rangle|$ . Note that  $f_m$  is orthogonal to  $f - f_m$ , that is, the following applies

$$\langle f - f_m, f_m \rangle = 0, \quad m \geq 0 \tag{4}$$

It is observed that

$$f = \Phi x = \sum_{i=1}^d x_i \varphi_i = \sum_{i \in \text{supp}(x)} x_i \varphi_i \tag{5}$$

From (4), (5), and the Hölder inequality, we have

$$\begin{aligned} \|f - f_{m-1}\|_2^2 &= \langle f - f_{m-1}, f - f_{m-1} \rangle \\ &= \langle f - f_{m-1}, f \rangle \\ &= \sum_{i \in \text{supp}(x)} x_i \langle f - f_{m-1}, \varphi_i \rangle \\ &\leq \sum_{i \in \text{supp}(x)} |x_i| |\langle f - f_{m-1}, \varphi_i \rangle| \\ &\leq |\langle f - f_{m-1}, \varphi_{i_m} \rangle| \sum_{i \in \text{supp}(x)} |x_i| \\ &\leq |\langle f - f_{m-1}, \varphi_{i_m} \rangle| k^{1/2} \left( \sum_{i \in \text{supp}(x)} |x_i|^2 \right)^{1/2} \\ &= k^{1/2} |\langle f - f_{m-1}, \varphi_{i_m} \rangle| \|x\|_2 \end{aligned}$$

From the RIP condition of  $\Phi$  with the parameters  $(k, \delta)$ , we obtain

$$\begin{aligned} \|f - f_{m-1}\|_2^2 &\leq k^{1/2} |\langle f - f_{m-1}, \varphi_{i_m} \rangle| \|x\|_2 \\ &\leq k^{1/2} (1 - \delta)^{-1/2} |\langle f - f_{m-1}, \varphi_{i_m} \rangle| \|\Phi x\|_2 \\ &= k^{1/2} (1 - \delta)^{-1/2} |\langle f - f_{m-1}, \varphi_{i_m} \rangle| \|f\|_2 \end{aligned}$$

which implies

$$|\langle f - f_{m-1}, \varphi_{i_m} \rangle| \geq \frac{(1 - \delta)^{1/2}}{k^{1/2} \|f\|_2} \|f - f_{m-1}\|_2^2$$

Thus, combining the above inequality with (3), we have

$$\begin{aligned} \|f - f_m\|_2^2 &\leq \|f - f_{m-1}\|_2^2 - (1 - \delta) \frac{(1 - \delta)}{k \|f\|_2^2} \|f - f_{m-1}\|_2^4 \\ &= \|f - f_{m-1}\|_2^2 \left( 1 - \frac{(1 - \delta)^2}{k \|f\|_2^2} \|f - f_{m-1}\|_2^2 \right) \\ &= \|f - f_{m-1}\|_2^2 \left( 1 - \left[ \frac{k^{1/2} \|f\|_2}{(1 - \delta)} \right]^{-2} \|f - f_{m-1}\|_2^2 \right) \end{aligned}$$

Note that

$$\|f - f_0\|^2 = \|f\|^2 \leq \frac{k}{(1 - \delta)^2} \|f\|_2^2$$

Applying Lemma 1 with  $a_m = \|f - f_m\|_2^2$ ,  $B = \left(\frac{k^{1/2} \|f\|_2}{1 - \delta}\right)^2$ ,  $r_m := 1$ ,  $r = \left(\frac{k^{1/2} \|f\|_2}{1 - \delta}\right)^2$ ,  $J = 0$ , and  $\ell = 1$ , then we have

$$\|f - f_m\|_2^2 \leq \left(\frac{k^{1/2} \|f\|_2}{1 - \delta}\right)^2 (m + 1)^{-1}, \quad m \geq 1$$

which completes the proof of Theorem 4.  $\square$

Based on the result of Theorem 4, we can prove Theorem 5.

**Proof of Theorem 5:** If the matrix  $\Phi$  satisfies the RIP condition with parameters  $(2k, \delta)$ , then by Theorem 4, we can estimate the upper bound of  $\|x - \hat{x}\|$  as follows:

$$\begin{aligned} \|x - \hat{x}\|_2^2 &\leq (1 - \delta)^{-1} \|\Phi(x - \hat{x})\|_2^2 \\ &= (1 - \delta)^{-1} \|\Phi x - \Phi \hat{x}\|_2^2 \\ &= (1 - \delta)^{-1} \|f - f_L\|_2^2 \\ &\leq \frac{k \|f\|_2^2}{(1 - \delta)^3} (L + 1)^{-1} \end{aligned}$$

Thus, the proof of Theorem 5 is finished.  $\square$

It is known from Theorem 5 that RMP can obtain a good approximation of the sparse signal  $f$ . In this case, it is worse than the OMP, since OMP can recover  $f$  exactly if the RIP condition is satisfied. The key point is that one can establish the ideal Lebesgue-type inequality for OMP but not for RMP; for example, see [22,33,34]. On the other hand, as far as we know, until now, it is impossible to construct a deterministic matrix satisfying the RIP condition. In practice, one usually uses the random matrix in signal recovery. We will show that the performance of the RMP is similar to the OMP in this case.

#### 4. Signal Recovery with Admissible Measurements

Since one could not construct a deterministic matrix satisfying the RIP condition, it is natural to consider a random matrix. In practice, almost all approaches to recovering sparse signals using greedy-type methods involve some kind of randomness. In this section, we choose a class of random matrices which have good properties. This class of matrices is

called admissible matrices. We first recall the definition of this class of matrices. Then, we prove that the support of a  $k$ -sparse target signal in  $\mathbb{R}^d$  can be recovered with a high probability by applying the RMP algorithm. This shows that RMP is a powerful method of recovering sparse signals in high-dimensional Euclidean spaces.

**Definition 2.** An admissible measurement matrix for a  $k$ -sparse signal in  $\mathbb{R}^d$  is an  $N \times d$  random matrix  $\Phi$  satisfying the following conditions:

(M1) The columns of  $\Phi$  are stochastically independent.

(M2)  $\mathbb{E}(\|\varphi_j\|_2^2) = 1$  for  $j = 1, \dots, d$ .

(M3) Let  $\{u_m\}$  be a sequence of  $k$  vectors whose  $\ell_2$  norm do not exceed 1. If  $\varphi$  is a column of  $\Phi$  independent from  $\{u_m\}$ , then

$$\text{Prob}\left\{\max_m |\langle \varphi, u_m \rangle| \leq \varepsilon\right\} \geq 1 - 2ke^{-C_3\varepsilon^2N}$$

for a positive constant  $C_3$ .

(M4) For a given  $N \times k$  submatrix  $Z$  from  $\Phi$ , the  $k$ th largest singular value  $\sigma_k(Z)$  satisfies

$$\text{Prob}\left\{\sigma_k(Z) \geq \frac{1}{2}\right\} \geq 1 - e^{-C_4N}$$

for a positive constant  $C_4$ .

We illustrate some points about the conditions (M3) and (M4). The above condition (M3) offers a limitation of the singular value of the submatrix, which is analogous to the RIP condition (see [17]). Furthermore, the condition (M4) controls the inner product, since the selection of index at Step 2 of the RMP algorithm relates to the inner product. Two typical classes of admissible measurement matrices are the Gaussian matrix and Bernoulli matrix, which were first introduced by Candès and Tao for signal recovery and defined as follows (see [2]).

**Definition 3.** A measurement matrix  $\Phi$  is called a Gaussian measurement if every entry of  $\Phi$  is selected from the Gaussian distribution  $\mathcal{N}(0, N^{-1})$ , of which the density function  $p(x)$  is

$$p(x) = \frac{1}{\sqrt{2\pi N}}e^{-x^2N/2}, \text{ for } x \in \mathbb{R}$$

**Definition 4.** A measurement matrix  $\Phi$  is called a Bernoulli measurement if every entry of  $\Phi$  is selected to be  $\pm \frac{1}{\sqrt{N}}$  with the same probability, i.e.,  $\text{Prob}\{\Phi_{ij} = 1/\sqrt{N}\} = \text{Prob}\{\Phi_{ij} = -1/\sqrt{N}\} = \frac{1}{2}$ , where  $i = 1, \dots, N, j = 1, \dots, d$ .

Obviously, for Gaussian and Bernoulli measurements, the properties (M1) and (M2) are straightforward to verify. We can check the other two properties using the following known results. See [17] for more details.

**Proposition 1.** Suppose  $\{u_m\}$  is a sequence of  $k$ -sparse vectors. The  $\ell_2$  norm of each  $u_m$  does not exceed 1. Choose  $z$  independently to be a random vector with i.i.d.  $\mathcal{N}(0, N^{-1})$  entries. Then

$$\text{Prob}\left\{\max_m |\langle z, u_m \rangle| \leq \varepsilon\right\} \geq 1 - ke^{-2\varepsilon^2N/2}$$

**Proposition 2.** Let  $Z$  be an  $N \times k$  matrix whose entries are all i.i.d.  $\mathcal{N}(0, N^{-1})$  or else i.i.d. uniform on  $\pm \frac{1}{\sqrt{N}}$ . Then

$$\frac{1}{2}\|x\|_2 \leq \|Z\|_2 \leq \frac{3}{2}\|x\|_2, \text{ for all } x \in \mathbb{R}^k$$

with a probability of at least

$$1 - 2 \cdot 24^k \cdot e^{-cN}$$

For the RMP algorithm, we derive the following result.

**Theorem 6.** *Let  $x$  be a  $k$ -sparse signal in  $\mathbb{R}^d$  and  $\Phi$  be an  $N \times d$  admissible matrix independent from the signal. For the given data  $f = \Phi x$ , if the RMP algorithm selects  $k$  elements in the first  $L \geq k$  iterations, the correct support of  $x$  can be found with the probability exceeding  $1 - 4L(d - k)e^{-cN/k}$ , where  $c$  is a constant dependent on the admissible measurement matrix  $\Phi$ .*

**Proof of Theorem 6:** Without a loss of generality, we may assume that the first  $k$  components of  $x$  are nonzero and the other remaining components of  $x$  are zeroes. Obviously,  $f$  can be expressed as a linear combination of the first  $k$  columns of measurement matrix  $\Phi$ . We decompose the measurement matrix  $\Phi$  into two parts  $\Phi = [\Phi_{opt} | \Psi]$ , where  $\Phi_{opt}$  is the submatrix formed by the first  $k$  columns of  $\Phi$ , and  $\Psi$  is the submatrix formed by the remaining  $d - k$  columns. For  $\Phi_{opt}$ , we assume that

$$\Phi_{opt} = [\varphi_1, \dots, \varphi_k], \varphi_1, \dots, \varphi_k \in \Phi$$

and

$$f = \Phi x = \sum_{i=1}^k \varphi_i x_i$$

Furthermore, we can know that  $f = \Phi x$  is independent from the random matrix  $\Psi$ .

Denote  $E_{succ}$  as the event where the RMP algorithm recovers all  $k$  elements from the support of  $x$  correctly in the first  $L \geq k$  iterations. Define  $\Sigma$  as the event where  $\sigma_k(\Phi_{opt}) \geq \frac{1}{2}$ . Then, we have

$$\begin{aligned} \text{Prob}\{E_{succ}\} &\geq \text{Prob}\{E_{succ} \cap \Sigma\} \\ &= \text{Prob}\{E_{succ} | \Sigma\} \cdot \text{Prob}\{\Sigma\} \end{aligned}$$

For an  $N$ -dimensional vector  $r$ , define

$$\rho(r) := \frac{\|\Psi^T r\|_\infty}{\|\Phi_{opt}^T r\|_\infty} = \frac{\max_\psi |\langle \psi, r \rangle|}{\|\Phi_{opt}^T r\|_\infty}$$

where  $\psi$  is a column of the matrix  $\Psi$ . If  $r$  is a residual vector derived by the Step 2 of the RMP algorithm and  $\rho(r)$  satisfies  $\rho(r) < 1$ , it means that a column from matrix  $\Phi_{opt}$  has been selected.

If we execute the RMP algorithm with input signal  $x$  and measurement matrix  $\Phi_{opt}$  for  $L$  iterations, then we obtain a sequence of residual vectors  $r_0, r_1, \dots, r_{L-1}$ . Obviously, the residual vectors are independent from the matrix  $\Psi$  and can be considered as the functions of  $x$  and  $\Phi_{opt}$ . Instead, suppose that we execute the RMP algorithm with the input signal  $x$  and matrix  $\Phi$  for  $L$  iterations. If the RMP algorithm recovers all the  $k$  elements of the support of  $x$  correctly, the columns of  $\Phi_{opt}$  should be selected at each iteration. Thus, the sequence of residual vectors should be same as when we execute the algorithm with  $\Phi_{opt}$ .

Therefore, the condition probability satisfies

$$\text{Prob}\{E_{succ} | \Sigma\} \geq \text{Prob}\{\max_m \rho(r_m) < 1 | \Sigma\}$$

where  $r_m$  is a random vector depending on  $\Phi_{opt}$  and stochastically independent from  $\Psi$ .

Assume that  $\Sigma$  occurs. Since  $\Phi_{opt}^T r$  is a  $k$ -dimensional vector, then

$$\rho(r_m) = \frac{\max_{\psi} |\langle \psi, r_m \rangle|}{\|\Phi_{opt}^T r_m\|_{\infty}} \leq \frac{\sqrt{k} \max_{\psi} |\langle \psi, r_m \rangle|}{\|\Phi_{opt}^T r_m\|_2}$$

By the basic property of singular values, we have

$$\frac{\|\Phi_{opt}^T r\|_2}{\|r\|_2} \geq \sigma_k(\Phi_{opt}) \geq \frac{1}{2}$$

for any vector  $r$  in the range of  $\Phi_{opt}$ . Thus, for vector  $r_m$ , we have  $\|u_m\|_2 \leq 1$ , where  $u_m := \frac{1}{2} \frac{r_m}{\|\Phi_{opt}^T r_m\|_2}$ . Then, they have the following:

$$\begin{aligned} \rho(r_m) &\leq \frac{\sqrt{k} \max_{\psi} |\langle \psi, r_m \rangle|}{\|\Phi_{opt}^T r_m\|_2} \\ &= 2\sqrt{k} \max_{\psi} \left| \left\langle \psi, \frac{r_m}{2\|\Phi_{opt}^T r_m\|_2} \right\rangle \right| \\ &= 2\sqrt{k} \max_{\psi} |\langle \psi, u_m \rangle| \end{aligned}$$

for each index  $m = 0, 1, \dots, L - 1$ .

Thus, we have

$$\text{Prob}\{\rho(r_m) < 1|\Sigma\} \geq \text{Prob}\left\{ \max_m \max_{\psi} |\langle \psi, u_m \rangle| < \frac{1}{2\sqrt{k}} \mid \Sigma \right\}$$

Since the columns of  $\Psi$  are stochastically independent from each other, we can exchange the two maxima to obtain

$$\text{Prob}\{\max_m \rho(r_m) < 1|\Sigma\} \geq \prod_{\psi} \text{Prob}\left\{ \max_m |\langle \psi, u_m \rangle| < \frac{1}{2\sqrt{k}} \mid \Sigma \right\}$$

Obviously, each column of  $\Psi$  is independent from  $\{u_m\}$  and  $\Sigma$ . By condition (M3), we have

$$\text{Prob}\{\max_m \rho(r_m) < 1|\Sigma\} \geq [1 - 2Le^{-C_3N/4k}]^{d-k}$$

By the property (M4), we have

$$\text{Prob}\{\Sigma\} = \text{Prob}\left\{ \sigma_k(\Phi_{opt}) \geq \frac{1}{2} \right\} \geq 1 - e^{-C_4N}$$

Therefore, the following applies:

$$\begin{aligned} \text{Prob}\{E_{succ}\} &\geq \text{Prob}\{E_{succ}|\Sigma\} \cdot \text{Prob}\{\Sigma\} \\ &\geq \text{Prob}\{\max_m \rho(r_m) < 1|\Sigma\} \cdot \text{Prob}\{\Sigma\} \\ &\geq [1 - 2Le^{-C_3N/4k}]^{d-k} \cdot [1 - e^{-C_4N}] \end{aligned}$$

By using the following known inequality for  $k \geq 1, x \leq 1$ ,

$$(1 - x)^k \geq 1 - kx$$

we have

$$\begin{aligned} \text{Prob}\{E_{succ}\} &\geq 1 - 2L(d - k)e^{-C_3N/4k} - e^{-C_4N} \\ &\geq 1 - 4L(d - k)e^{-cN/k} \end{aligned}$$

We finish the proof of this theorem.  $\square$

We remark that a similar result holds for the OMP (see [17]). So, the performance of the RMP is similar to that of the OMP in the random case. In the next section, we will verify the conclusion of Theorem 6 via numerical experiments.

### 5. Simulation Results

In the present section, we check the efficiency of the Rescaled Matching Pursuit (RMP) algorithm defined in Section 2. We adopt the following model. Assume that  $x \in \mathbb{R}^d$  is the target signal. We want to recover it by using the following information:

$$f = \Phi x \tag{6}$$

where  $\Phi \in \mathbb{R}^{N \times d}$  is a Bernoulli measurement matrix.

In our experiments, we generate a  $k$ -sparse signal  $x \in \mathbb{R}^d$  by randomly choosing  $k$  components, and each of them is selected from a standard normal distribution  $\mathcal{N}(0, 1)$ . We execute the RMP algorithm with a Bernoulli measurement matrix under the model (6). We measure the performance of the RMP algorithm with the mean square error (MSE), which is defined as

$$\text{MSE} = \frac{1}{d} \sum_{j=1}^d (x_j - \hat{x}_j)^2$$

where  $\hat{x} \in \mathbb{R}^d$  is the estimate of the target signal  $x$ .

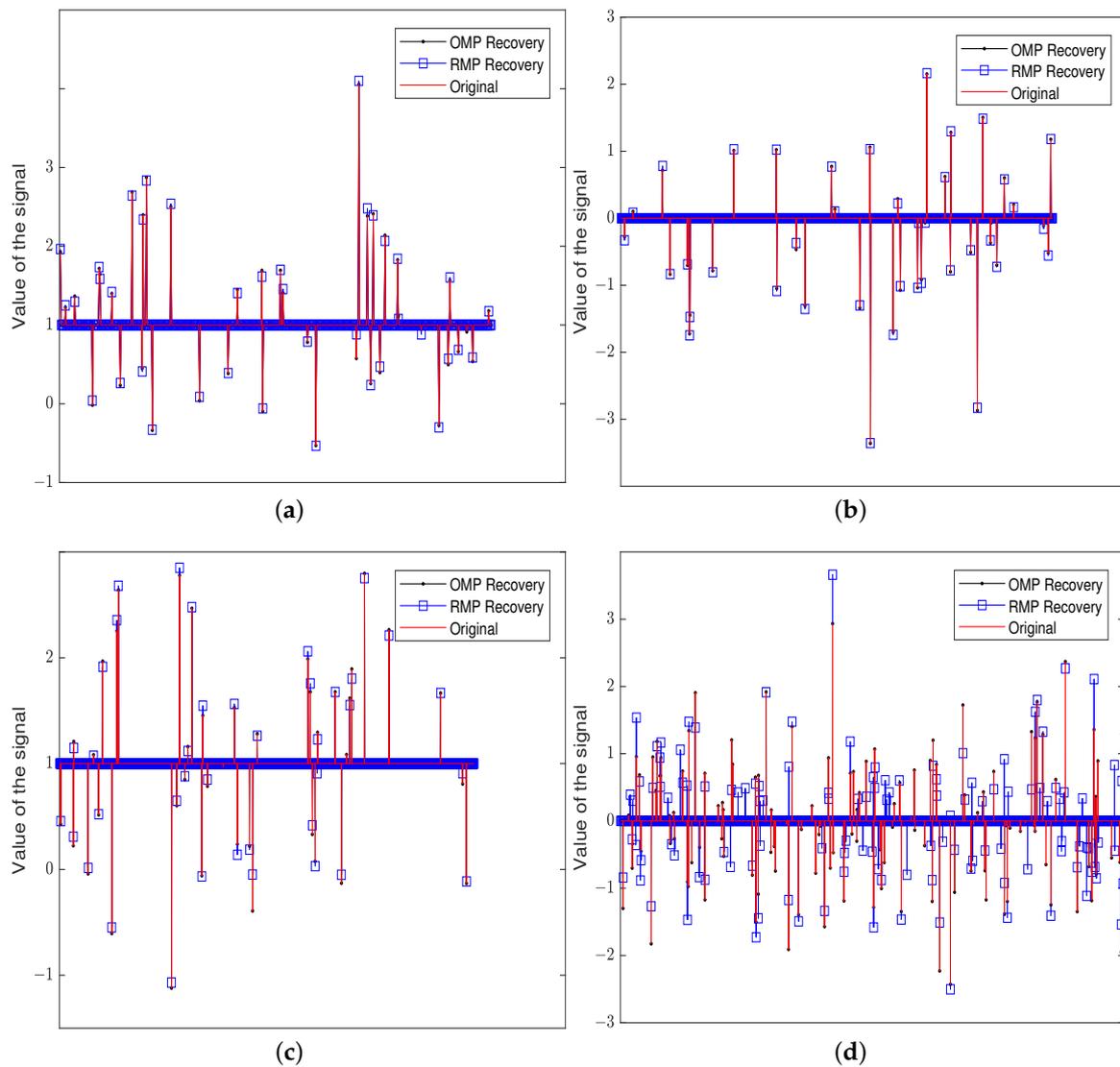
The first plot, Figure 1, displays the performances of the OMP and the RMP for an input signal with different sparsity indexes  $k$ , different dimensions  $d$ , and different numbers of measurements  $N$ . The red line denotes the original signal, the blue squares represent the approximation of the RMP, and the black dots denote the approximation of the OMP.

From Figure 1, we know that the RMP can obtain a good approximation for a sparse signal even in a high dimension or when  $N$  is much smaller than  $d$ . After repeating the test 100 times, we can figure out the mean square errors of the above four cases as follows:

- (a)  $\text{MSE}(\text{OMP}) = 4.1767 \times 10^{-6}$ ;      $\text{MSE}(\text{RMP}) = 2.3491 \times 10^{-4}$
- (b)  $\text{MSE}(\text{OMP}) = 1.5524 \times 10^{-6}$ ;      $\text{MSE}(\text{RMP}) = 1.1254 \times 10^{-4}$
- (c)  $\text{MSE}(\text{OMP}) = 1.03 \times 10^{-6}$ ;      $\text{MSE}(\text{RMP}) = 1.28 \times 10^{-4}$
- (d)  $\text{MSE}(\text{OMP}) = 5.44 \times 10^{-5}$ ;      $\text{MSE}(\text{RMP}) = 9.422 \times 10^{-3}$

Thus, taking a suitable number of measurements, the performance of the RMP is similar to the OMP within allowable limits.

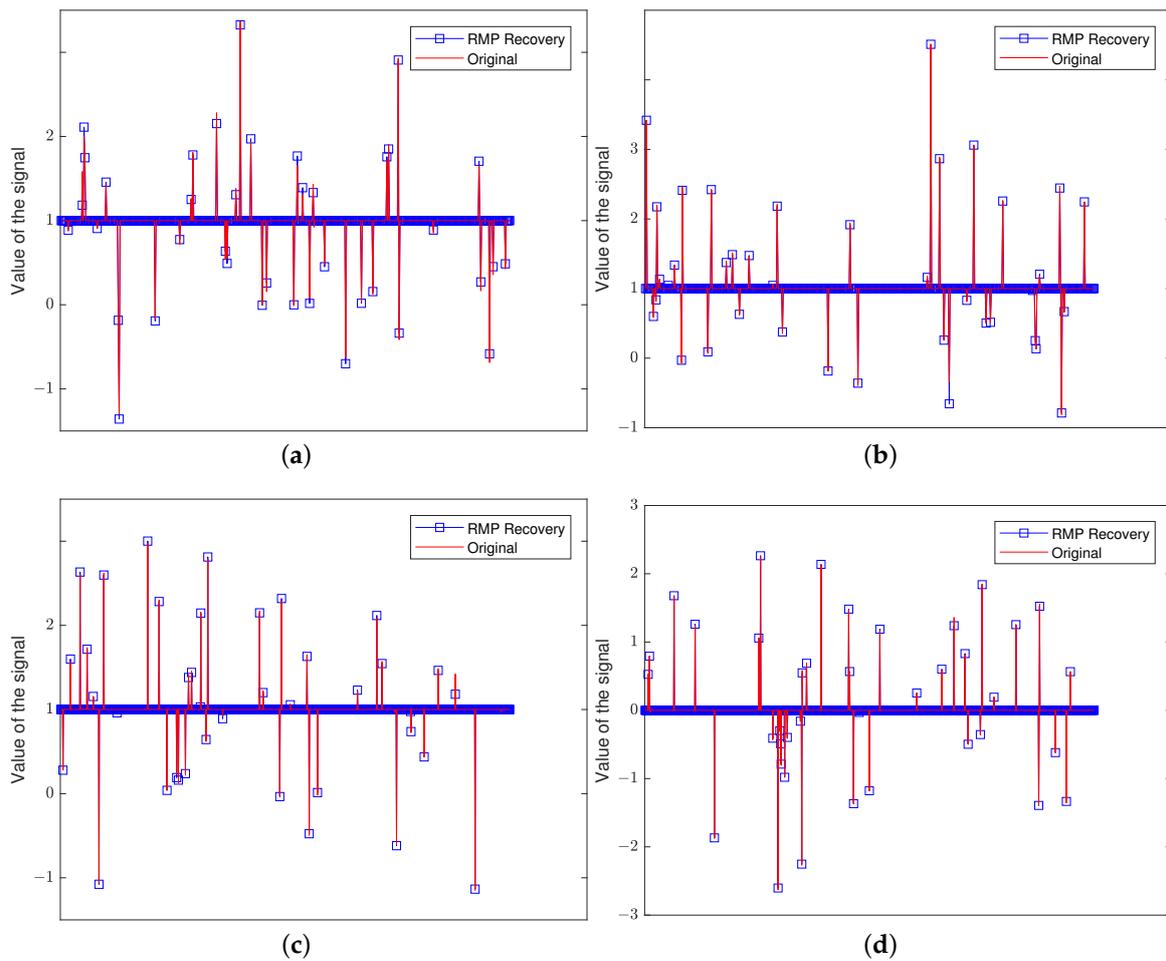
Figure 2 shows that the RMP can approximate the 40 sparse signal well under noise, even in a high dimension, which implies that the RMP is stable with the noise.



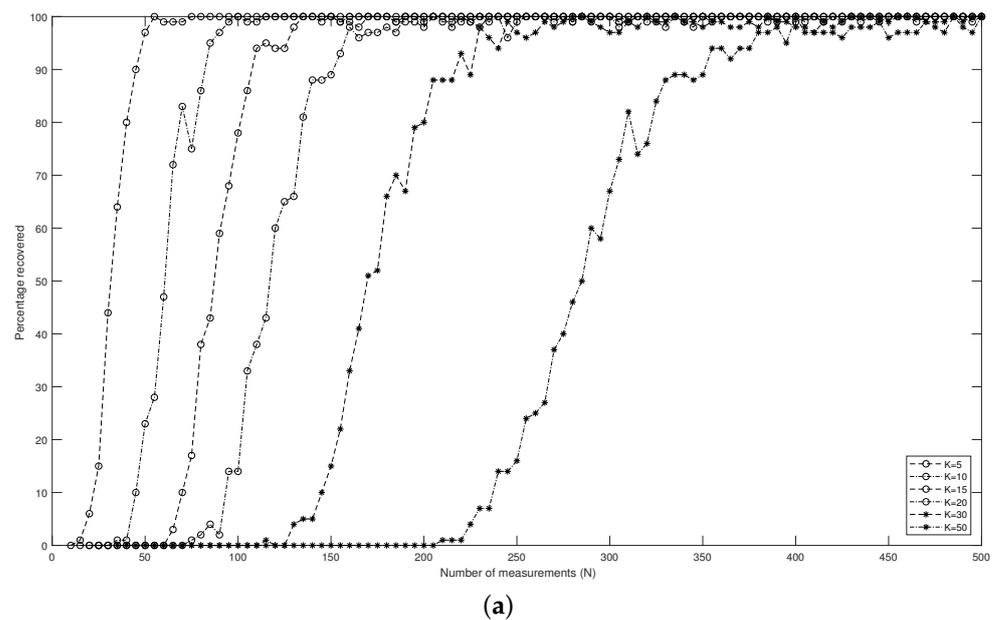
**Figure 1.** The recovery of an input signal for the OMP and the RMP under different sparsity levels  $k$ , different dimensions  $d$ , and different numbers of measurements  $N$ . (a)  $k = 40$ ,  $N = 250$ ,  $d = 512$ ; (b)  $k = 40$ ,  $N = 250$ ,  $d = 1024$ ; (c)  $k = 40$ ,  $N = 250$ ,  $d = 2048$ , (d)  $k = 125$ ,  $N = 500$ ,  $d = 5000$ .

Figure 3 displays the relation between the percentage (of 100 input signals) of the support that can be determined correctly and the number  $N$  of measurements in different dimensions  $d$ . It reveals how many measurements are necessary to reconstruct a  $k$ -sparse signal  $x \in \mathbb{R}^d$  with a high probability. A percentage equal to 100% means that support for all 100 input signals can be found. That is to say, the support of the input signal can be exactly recovered. Furthermore, Figure 3 indicates that to guarantee the success of recovering the target signal, the number  $N$  of the measurements must increase when the sparsity level  $k$  increases.

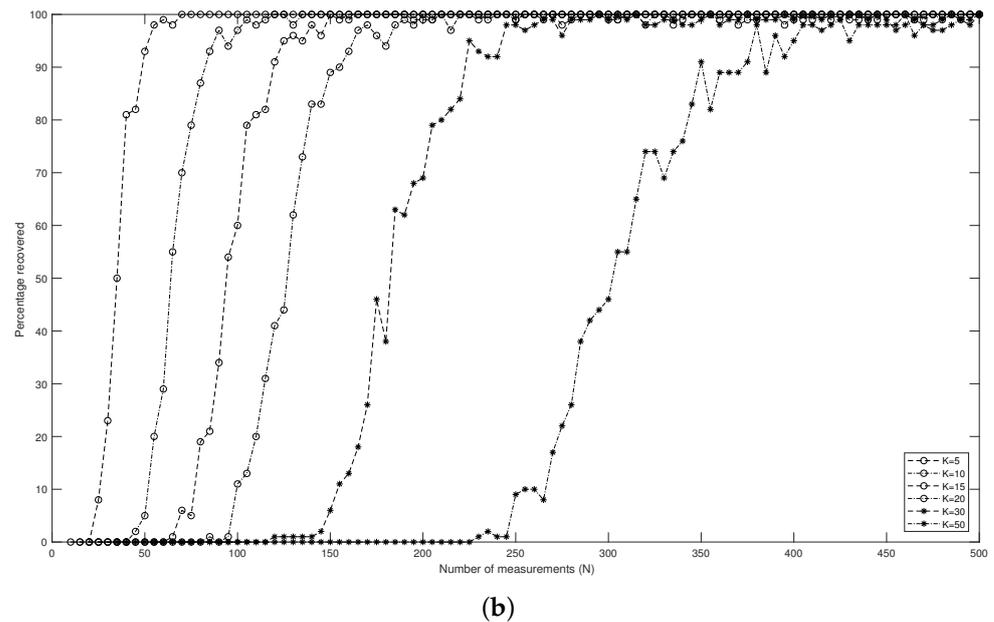
The numerical experiments show that the RMP is efficient for the recovery of the sparse signals even in the noised case. The RMP can determine the support of the target signal in a high-dimension Euclidean space with a high probability, if the number  $N$  is large enough.



**Figure 2.** The recovery of an input signal with sparsity level  $k = 40$  and number of measurements  $N = 250$  via RMP in different dimensions  $d$  and different noise levels: (a)  $d = 512$ , no noise; (b)  $d = 512$ , Gaussian noise; (c)  $d = 1024$ , no noise; (d)  $d = 1024$ , Gaussian noise. The red line denotes the original signal. The blue squares stand for the approximation obtained by applying the RMP.



**Figure 3.** Cont.



**Figure 3.** The percentage of the support of 100 input signals being recovered correctly as a function of the number  $N$  of Bernoulli measurements for different sparsity levels  $k$  in different dimensions  $d$  via RMP: (a)  $d = 512$ ; (b)  $d = 1024$ .

## 6. Conclusions

We design the Rescaled Matching Pursuit (RMP) algorithm for the recovery of the target signal in Euclidean space with a high dimension. The RMP algorithm performs with great advantages in terms of computational complexity over other greedy-based compressive sensing algorithms, such as the OMP algorithm, since its total cost is  $O(N)$  at  $m$  iterations. By using the RIP condition, we derive the convergence rate of the error between the original signal and its approximation for the RMP algorithm. Moreover, we prove that the RMP algorithm can recover the support of the original sparse signal by using an admissible matrix with a high probability. Our numerical simulation results also verify this conclusion. And, we show that the RMP can obtain a good approximation for original sparse signal under noise, namely, the RMP is stable with noise. Thus, in both theory and practice, we show that the RMP algorithm is powerful in random cases. Although the computational cost of the RMP is smaller than that of the OMP, the performance of the RMP is not better than the OMP in the deterministic case. However, in randomized cases, the two algorithms perform similarly. Thus, if the sparsity level  $k$  of the target signal is large and computational resources are limited, then the RMP with random measurements may be a good choice.

**Author Contributions:** W.L. and P.Y. contribute evenly to this paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was partially funded by the NSFC of China, grant no. 11671213.

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

**Acknowledgments:** The authors thank Li Haifeng and Shen Yi for their useful comments. The authors also thank the four referees for their suggestions. The suggestions and comments helped the authors to improve the manuscript.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Donoho, D.L. Compressed sensing. *IEEE Trans. Inf. Theory* **2006**, *52*, 1289–1306. [[CrossRef](#)]
2. Candès, E.J.; Romberg, J.; Tao, T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory* **2006**, *52*, 489–509. [[CrossRef](#)]
3. Yuan, X.; Haimi-Cohen, R. Image compression based on compressive sensing: End-to-end comparison with JPEG. *IEEE Trans. Multimed.* **2020**, *22*, 2889–2904. [[CrossRef](#)]
4. Berger, C.R.; Wang, Z.; Huang, J.; Zhou, S. Application of compressive sensing to sparse channel estimation. *IEEE Commun. Mag.* **2010**, *48*, 164–174. [[CrossRef](#)]
5. Barranca, V.J. Reconstruction of sparse recurrent connectivity and inputs from the nonlinear dynamics of neuronal networks. *J. Comput. Neurosci.* **2023**, *51*, 43–58. [[CrossRef](#)] [[PubMed](#)]
6. Dai, W.; Sheikh, M.A.; Milenkovic, O.; Baraniuk, R.G. Compressive sensing DNA microarrays. *EURASIP J. Bioinform. Syst. Biol.* **2009**, *2009*, 162824. [[CrossRef](#)]
7. Gross, D.; Liu, Y.K.; Flammia, S.T.; Becker, S.; Eisert, J. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.* **2010**, *105*, 150401. [[CrossRef](#)] [[PubMed](#)]
8. Candès, E.J.; Wakin, M.B. An introduction to compressive sampling. *IEEE Signal Proc. Mag.* **2008**, *25*, 21–30. [[CrossRef](#)]
9. Candès, E.J.; Tao, T. Decoding by linear programming. *IEEE Trans. Inf. Theory* **2005**, *51*, 4203–4215. [[CrossRef](#)]
10. Cai, T.T.; Zhang, A.R. Sparse representation of a polytope and recovery of sparse signals and low-rank matrices. *IEEE Trans. Inf. Theory* **2014**, *60*, 122–132. [[CrossRef](#)]
11. Kim, S.J.; Koh, K.; Lustig, M.; Boyd, S.; Gorinevsky, D. An interior-point method for large-scale  $\ell_1$ -regularized least squares. *IEEE J. Sel. Top. Signal Process.* **2007**, *1*, 606–617. [[CrossRef](#)]
12. Candès, E.J.; Tao, T. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inf. Theory* **2006**, *52*, 5406–5425. [[CrossRef](#)]
13. Aziz, A.; Salim, A.; Osamy, W. Sparse signals reconstruction via adaptive iterative greedy algorithm. *Int. J. Comput. Appl.* **2014**, *90*, 5–11.
14. Goyal, P.; Singh, B. Greedy algorithms for sparse signal recovery based on temporally correlated experimental data in WSNs. *Arab. J. Sci. Eng.* **2018**, *43*, 7253–7264. [[CrossRef](#)]
15. Chae, J.; Hong, S. Greedy algorithms for sparse and positive signal recovery based on bit-wise MAP detection. *IEEE Trans. Signal Process.* **2020**, *68*, 4017–4029. [[CrossRef](#)]
16. Lv, X.L.; Wan, C.R.; Bi, G.A. Block orthogonal greedy algorithm for stable recovery of block-sparse signal representations. *Signal Process.* **2010**, *90*, 3265–3277. [[CrossRef](#)]
17. Tropp, J.A.; Gilbert, A.C. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inf. Theory* **2007**, *53*, 4655–4666. [[CrossRef](#)]
18. Huber, P.J. Projection pursuit. *Ann. Stat.* **1985**, *13*, 435–475. [[CrossRef](#)]
19. DeVore, R.A.; Temlyakov, V.N. Some remarks on greedy algorithms. *Adv. Comput. Math.* **1996**, *5*, 173–187. [[CrossRef](#)]
20. Barron, A.R.; Cohen, A.; Dahmen, W.; DeVore, R.A. Approximation and learning by greedy algorithms. *Ann. Stat.* **2008**, *36*, 64–94. [[CrossRef](#)]
21. Pati, Y.C.; Rezaifar, R.; Krishnaprasad, P.S. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In Proceedings of the 27th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 1–3 November 1993; Volume 1, pp. 40–44.
22. Shao, C.F.; Wei, X.J.; Ye, P.X.; Xing, S. Efficiency of orthogonal matching pursuit for group sparse recovery. *Axioms* **2023**, *12*, 389. [[CrossRef](#)]
23. Davis, G.; Mallat, S.; Avellaneda, M. Adaptive greedy approximations. *Constr. Approx.* **1997**, *13*, 57–98. [[CrossRef](#)]
24. Zhang, T. Sparse recovery with orthogonal matching pursuit under RIP. *IEEE Trans. Inf. Theory* **2011**, *57*, 6215–6221. [[CrossRef](#)]
25. Eldar, Y.C.; Kutyniok, G. *Compressed Sensing: Theory and Applications*; Cambridge University Press: Cambridge, UK, 2012.
26. Petrova, G. Rescaled pure greedy algorithm for Hilbert and Banach spaces. *Appl. Comput. Harmon. Anal.* **2016**, *41*, 852–866. [[CrossRef](#)]
27. Gao, Z.M.; Petrova, G. Rescaled pure greedy algorithm for convex optimization. *Calcolo* **2019**, *56*, 15. [[CrossRef](#)]
28. Guo, Q.; Liu, X.H.; Ye, P.X. The learning performance of the weak rescaled pure greedy algorithms. *J. Inequal. Appl.* **2024**, *2024*, 30. [[CrossRef](#)]
29. Zhang, W.H.; Ye, P.X.; Xing, S. Optimality of the rescaled pure greedy learning algorithms. *Int. J. Wavelets Multiresolut. Inf. Process.* **2023**, *21*, 2250048. [[CrossRef](#)]
30. Jiang, B.; Ye, P.X. Efficiency of the weak rescaled pure greedy algorithm. *Int. J. Wavelets Multiresolut. Inf. Process.* **2021**, *19*, 2150001. [[CrossRef](#)]
31. Zhang, W.H.; Ye, P.X.; Xing, S.; Xu, X. Optimality of the approximation and learning by the rescaled pure super greedy algorithms. *Axioms* **2022**, *11*, 437. [[CrossRef](#)]
32. Nguyen, H.; Petrova, G. Greedy strategies for convex optimization. *Calcolo* **2017**, *54*, 207–224. [[CrossRef](#)]

33. Temlyakov, V.N. *Greedy Approximation*; Cambridge University Press: Cambridge, UK, 2011.
34. Shao, C.F.; Chang, J.C.; Ye, P.X.; Zhang, W.H.; Xing, S. Almost optimality of the orthogonal super greedy algorithm for  $\mu$ -coherent dictionaries. *Axioms* **2022**, *11*, 186. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.