

Research Article

Recovery of Sparse Signals via Modified Hard Thresholding Pursuit Algorithms

Li-Ping Geng,¹ Jin-Chuan Zhou ,¹ Zhong-Feng Sun,¹ and Jing-Yong Tang²

¹School of Mathematics and Statistics, Shandong University of Technology, Zibo, China

²School of Mathematics and Statistics, Xinyang Normal University, Xinyang, China

Correspondence should be addressed to Jin-Chuan Zhou; jinchuanzhou@163.com

Received 22 May 2023; Revised 25 June 2023; Accepted 20 July 2023; Published 3 November 2023

Academic Editor: Wenling Li

Copyright © 2023 Li-Ping Geng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, we propose a modified version of the hard thresholding pursuit algorithm, called modified hard thresholding pursuit (MHTP), using a convex combination of the current and previous points. The convergence analysis, finite termination properties, and stability of the MHTP are established under the restricted isometry property of the measurement matrix. Simulations are performed in noiseless and noisy environments using synthetic data, in which the successful frequencies, average runtime, and phase transition of the MHTP are considered. Standard test images are also used to test the reconstruction capability of the MHTP in terms of the peak signal-to-noise ratio. Numerical results indicate that the MHTP is competitive with several mainstream thresholding and greedy algorithms, such as hard thresholding pursuit, compressive sampling matching pursuit, subspace pursuit, generalized orthogonal matching pursuit, and Newton-step-based hard thresholding pursuit, in terms of recovery capability and runtime.

1. Introduction

Compressed sensing has been applied in many industrial fields, such as image processing [1, 2], wireless channel estimation [3], group testing [4], and sparse signal recovery [5–7]. One of the key components in compressed sensing is the reconstruction of a signal using measurements smaller than the signal length. The underlying assumption used to achieve this is that the signal is required to be sparse. The corresponding mathematical model can be expressed as the following sparse optimization problem:

$$\min_x \left\{ \frac{1}{2} \|y - Ax\|_2^2 : \|x\|_0 \leq s \right\}, \quad (1)$$

where $\|x\|_0$ represents the total number of nonzero entries of $x \in \mathbb{R}^n$, s is the sparsity level, $A \in \mathbb{R}^{m \times n}$ is the measurement matrix with $m \ll n$, and $y \in \mathbb{R}^m$ is the measurement vector.

Many different types of algorithms exist to solve the aforementioned sparse optimization problem. The algorithms are divided into three categories: optimization [8–12], greedy [13–17], and thresholding [18–23]. In principle, which

algorithm should be selected depended on the actual situation. For example, the runtime is one of the criteria for algorithms. Basic pursuit is one of the optimization methods, and its runtime depends on the choice of algorithm used for the minimization. If sparsity s is much smaller than the signal length n , the greedy Orthogonal Matching Pursuit (OMP) is extremely fast. A major advantage of hard thresholding algorithms is that they are almost not affected by the level of sparsity. In addition, the theoretical analysis tools for these algorithms maybe different as well; see [24] for more discussion. In this study, we focused on a popular thresholding method, the hard thresholding pursuit (HTP) algorithm [21–23]. HTP is a modified version of the iterative hard thresholding (IHT) algorithm that adds an orthogonal projection/debiasing process. IHT [18] is a combination of the gradient descent direction and hard-thresholding operator. Although the gradient descent direction is used, the value of the objective function may not decrease because of the application of the hard-thresholding operator, which guarantees the sparsity of signals without simultaneously considering the reduction of the objective function value. This drawback is overcome using a new thresholding operator called Optimal

s -Thresholding, which selects s components that make the objective function reach the minimum [25, 26]. However, this modification causes a significant increase in runtime because the subproblem is not easy to solve. Recently, using a binary regularization and linearization technique, Zhao and Luo [27] proposed Natural Thresholding algorithm to reduce the computational cost. In addition, Liu and Barber introduced a new operator called the Reciprocal Thresholding [28], which lies between hard and soft thresholding. As mentioned earlier, an advantage of the IHT is that its runtime is almost uninfluenced by the sparsity level. Along with its simple structure, this property makes IHT considerably attractive. Moreover, orthogonal projection aims to find a vector that best fits the measurements in the corresponding subspace; therefore, its combination with IHT will alleviate the instability of IHT and lead to the HTP algorithm. Additional information on hard thresholding algorithms and their applications can be found in [5, 29–35].

The main contributions of this study are given as follows:

- (1) It should be noted that many algorithms for solving compressed sensing only use the gradient information of the current point x^p , such as IHT [18], HTP [22], compressive sampling matching pursuit (CoSaMP) [15], and subspace pursuit (SP) [13]. In the design of an algorithm, it is possible to improve the efficiency of the algorithm utilizing more information provided by previous iteration points. Considering this motivation, we used the information of the current point x^p and the previous point x^{p-1} to improve HTP performance. Notably, compressed sensing has a combinational structure; hence, it does not strictly belong to the field of continuous optimization. This motivated us to study the convex combination rather than the (extra) momentum technique because the hard-thresholding operator would break down the advantage imposed on the (extra) momentum technique shown in standard optimization problems. In other words, we present a novel idea of using a convex combination of x^p and x^{p-1} to generate an intermediate point u^p and then perform the gradient descent direction at u^p (not x^p). Based on this idea and using orthogonal projection, we propose a new algorithm called the modified hard thresholding pursuit (MHTP) algorithm.
- (2) To guarantee the convergence of our algorithm from a theoretical perspective, we discuss the Restricted Isometry Property (RIP) of measurement matrix A . More precisely, if the Restricted Isometry Constant (RIC), δ_{3s} , of the measurement matrix A satisfies

$$\delta_{3s} < \frac{1}{\sqrt{3}}, \quad (2)$$

the iterative sequence generated by the MHTP converges. This bound is the same as that for HTP [22]; indeed, the best RIP-based bound for HTP is

$\delta_{3s} < 1/\sqrt{3} \approx 0.577$ [22] at present. The algorithm proposed in this paper includes HTP as a special case; hence, the RIP-based bound cannot be larger than $1/\sqrt{3}$. We establish the convergence analysis, finite termination, and stability of the MHTP without decreasing this upper bound.

- (3) To demonstrate the effectiveness of the MHTP for sparse signal recovery, numerical experiments are conducted by comparing it with several mainstream algorithms such as HTP, Newton-step-based hard thresholding pursuit (NSHTP) [30], generalized orthogonal matching pursuit (gOMP) [36], CoSaMP, and SP from both synthetic data and real images. For synthetic data, simulations are conducted in noiseless and noisy environments, wherein the successful frequency, average runtime, and phase transition of algorithms are considered. The numerical results indicate that the MHTP outperforms the other algorithms in terms of the recovery capability of sparse signals and consumes less time than NSHTP, CoSaMP, SP, and gOMP. In real-world experiments, the reconstruction capability of the algorithms using several standard test images is compared in terms of the peak signal-to-noise ratio (PSNR) and total runtime. Simulations show that the MHTP is competitive with other algorithms in terms of image reconstruction quality and runtime.

The notations frequently used in this study are given as follows. For a given index set $S \subseteq \{1, 2, \dots, n\}$, $|S|$ denotes the cardinality of S and $\bar{S} := \{1, 2, \dots, n\} \setminus S$ denotes the complement of S . For a fixed vector $x \in \mathbb{R}^n$, x_S is obtained by retaining the elements of x indexed in S and zeroing out the rest of elements. The notation $\mathcal{L}_s(x)$ is an index set whose elements correspond to s -largest absolute entries of x , and $\mathcal{H}_s(x)$ represents the hard thresholding of x , that is, $\mathcal{H}_s(x) = x_{\mathcal{L}_s(x)}$. Let $\text{supp}(x)$ be the support of x , that is, $\text{supp}(x) := \{i \mid x_i \neq 0\}$. For a given real number a , $\lceil a \rceil$ denotes the ceil function of a . The remainder of this paper is organized as follows. The HTP (Algorithm 1) and MHTP (Algorithm 2) algorithms are described in Section 2. Section 3 presents a theoretical analysis of the MHTP. Numerical experiments are presented in Section 4. Finally, conclusions are drawn in Section 5.

2. Algorithm: MHTP

As mentioned earlier, the classical HTP uses the negative gradient at the current point x^p as the search direction and then resorts to a hard thresholding operator to ensure feasibility. Hence, the new iteration x^{p+1} significantly depends on the information at x^p .

To improve the performance of HTP, we attempt to use the information of the current point x^p and the previous point x^{p-1} to construct a new search direction. First, we introduce an intermediate point $u^p := x^p + (1 - \lambda)x^{p-1}$ that is a convex combination of x^p and x^{p-1} , where $\lambda \in (0, 1]$. Replacing x^p in the HTP by the intermediate point u^p leads to MHTP; clearly, MHTP is reduced to HTP when $\lambda = 1$.

Input: a measurement matrix A , a measurement vector y , a sparsity level s , a step size μ . Perform the following steps:

S1. Start with s -sparse vector $x^0 \in \mathbb{R}^n$, typically $x^0 = 0$.

S2. Repeat

$$\begin{aligned} S^{p+1} &:= \mathcal{L}_s(x^p + \mu A^T(y - Ax^p)), \\ x^{p+1} &:= \arg \min_z \{ \|y - Az\|_2 : \text{supp}(z) \subseteq S^{p+1} \}, \end{aligned}$$

until a stopping criterion is met.

Output: the s -sparse vector x^* .

ALGORITHM 1: Hard thresholding pursuit (HTP).

Input: a measurement matrix A , a measurement vector y , a sparsity level s , and two parameters $\mu > 0$, $0 < \lambda \leq 1$. Perform the following steps:

S1. Start with s -sparse vectors $x^0, x^1 \in \mathbb{R}^n$, typically $x^0 = x^1 = 0$.

S2. Repeat

$$\begin{aligned} u^p &:= \lambda x^p + (1 - \lambda)x^{p-1}, \\ S^{p+1} &:= \mathcal{L}_s(u^p + \mu A^T(y - Au^p)), \\ x^{p+1} &:= \arg \min_z \{ \|y - Az\|_2 : \text{supp}(z) \subseteq S^{p+1} \}, \end{aligned}$$

until a stopping criterion is met.

Output: the s -sparse vector x^* .

ALGORITHM 2: Modified hard thresholding pursuit (MHTP).

3. Theoretical Analysis

A variety of available tools has been introduced to analyze compressed sensing, such as coherence, RIP, null-space properties, and range-space properties. These differences and relationships are described in [24]. In this section, the RIP concept is used to establish the convergence analysis, finite convergence, and stability of the MHTP. Hence, we first define the RIC and RIP.

Definition 1. [9, 24] Let $A \in \mathbb{R}^{m \times n}$ be a matrix with $m \ll n$ and s be a positive integer. The restricted isometry constant (RIC) of order s denoted by $\delta_s := \delta_s(A)$ is the smallest number $\delta \geq 0$ such that

$$(1 - \delta)\|z\|_2^2 \leq \|Az\|_2^2 \leq (1 + \delta)\|z\|_2^2, \quad (3)$$

for any s -sparse vector $z \in \mathbb{R}^n$. The matrix A is said to satisfy the RIP of order s if $\delta_s < 1$.

Note that the RIC satisfies $\delta_t \leq \delta_s$ as $t \leq s$. The following are some basic inequalities related to RIC, which are frequently used in the theoretical analyses.

Lemma 1. [22, 24] Let $u \in \mathbb{R}^n$ and $v \in \mathbb{R}^m$ be the vectors, $S \subseteq \{1, 2, \dots, n\}$ be an index set, and s be a positive integer. The following statements hold.

- (i) If $|\text{supp}(u) \cup S| \leq s$, then $\|(I - A^T A)u\|_S \leq \delta_s \|u\|_2$.
- (ii) If $|S| \leq s$, then $\|(A^T v)_S\|_2 \leq \sqrt{1 + \delta_s} \|v\|_2$.

Because x^{p+1} is produced using the current point x^p and the previous point x^{p-1} , the following three-terms inequality plays a key role.

Lemma 2. [32] Suppose that a nonnegative sequence $\{a^p\} \subseteq \mathbb{R}$ with $p = 0, 1, \dots$ satisfies

$$a^{p+1} \leq b_1 a^p + b_2 a^{p-1} + b_3, \quad p \geq 1, \quad (4)$$

where $b_1, b_2, b_3 \geq 0$ and $b_1 + b_2 < 1$. Then

$$a^p \leq \rho^{p-1} [a^1 + (\rho - b_1)a^0] + \frac{b_3}{1 - \rho}, \quad (5)$$

with

$$0 \leq \rho := \frac{b_1 + \sqrt{b_1^2 + 4b_2}}{2} < 1. \quad (6)$$

The following result estimates the error on $\|v^p - x_S\|_2$ using $\|x^p - x_S\|_2$ and $\|x^{p-1} - x_S\|_2$, where $v^p = u^p + \mu A^T(y - Au^p)$ and $u^p = \lambda x^p + (1 - \lambda)x^{p-1}$.

Lemma 3. Suppose that the sequence $\{x^p\}$ is generated by the MHTP with an inaccurate measurement $y = Ax + e$. Let $v^p := u^p + \mu A^T(y - Au^p)$ and $S := \mathcal{L}_s(x)$. Then

$$\begin{aligned} \| (v^p - x_S)_{S \cup S^{p+1}} \|_2 &\leq \varrho \lambda \|x^p - x_S\|_2 \\ &+ \varrho(1 - \lambda) \|x^{p-1} - x_S\|_2 + \mu \sqrt{1 + \delta_{2s}} \|e'\|_2, \end{aligned} \quad (7)$$

where $\varrho := \mu \delta_{3s} + |1 - \mu|$ and $e' := Ax_S + e$.

Proof. Note that

$$\begin{aligned} v^p - x_S &= \mu\lambda(I - A^T A)(x^p - x_S) + \lambda(1 - \mu)(x^p - x_S) \\ &\quad + \mu A^T e' + \mu(1 - \lambda)(I - A^T A)(x^{p-1} - x_S) \\ &\quad + (1 - \lambda)(1 - \mu)(x^{p-1} - x_S). \end{aligned} \quad (8)$$

Since $|\text{supp}(x^i - x_S) \cup S \cup S^{p+1}| \leq 3s$ for $i = p, p-1$ and $|S \cup S^{p+1}| \leq 2s$, it follows from Lemma 1 that

$$\|(A^T e')_{S \cup S^{p+1}}\|_2 \leq \sqrt{1 + \delta_{2s}} \|e'\|_2, \quad (9)$$

$$\|(I - A^T A)(x^i - x_S)_{S \cup S^{p+1}}\|_2 \leq \delta_{3s} \|x^i - x_S\|_2, \quad i = p, p-1. \quad (10)$$

This together with Equation (8) leads to

$$\begin{aligned} &\|(v^p - x_S)_{S \cup S^{p+1}}\|_2 \\ &\leq \mu\lambda\|(I - A^T A)(x^p - x_S)_{S \cup S^{p+1}}\|_2 + \lambda\|(1 - \mu)(x^p - x_S)_{S \cup S^{p+1}}\|_2 \\ &\quad + \mu\|(A^T e')_{S \cup S^{p+1}}\|_2 + \mu(1 - \lambda)\|(I - A^T A)(x^{p-1} - x_S)_{S \cup S^{p+1}}\|_2 \\ &\quad + (1 - \lambda)\|(1 - \mu)(x^{p-1} - x_S)_{S \cup S^{p+1}}\|_2 \\ &\leq \lambda[\mu\delta_{3s} + |1 - \mu|]\|x^p - x_S\|_2 + (1 - \lambda)[\mu\delta_{3s} + |1 - \mu|]\|x^{p-1} - x_S\|_2 \\ &\quad + \mu\sqrt{1 + \delta_{2s}}\|e'\|_2, \end{aligned} \quad (11)$$

where $\mu > 0$ and $\lambda \in (0, 1]$ are used to ensure the coefficients above are nonnegative. It yields Equation (7) and the proof is complete. \square

The following property of orthogonal projection is essential in the analysis of the MHTP, see for example [22, Equation (3.21)] and Zhao [25, p. 49].

Lemma 4. *Given a vector $x \in \mathbb{R}^n$ and its corresponding index set $S := \mathcal{L}_s(x)$, let y be the inaccurate measurement of x , i.e., $y = Ax + e$, and $T \subset \{1, 2, \dots, n\}$ be an index set such that $|T| \leq s$. If z^* satisfies*

$$z^* = \arg \min_z \{\|y - Az\|_2 : \text{supp}(z) \subseteq T\}, \quad (12)$$

then

$$\|x_S - z^*\|_2 \leq \frac{\|(x_S)_{\bar{T}}\|_2}{\sqrt{1 - \delta_{2s}^2}} + \frac{\sqrt{1 + \delta_s}}{1 - \delta_{2s}} \|e'\|_2, \quad (13)$$

where $e' := Ax_{\bar{S}} + e$.

HTP is a mainstream thresholding algorithm for sparse signal recovery. The best RIP-based bound for HTP is $\delta_{3s} < 1/\sqrt{3} \approx 0.577$ [22]. The MHTP algorithm proposed in this paper includes HTP as a special case; hence, the RIP-based bound cannot exceed $1/\sqrt{3}$. The following theorem provides an affirmative answer as to whether the convergence of the MHTP can be established without decreasing the upper bound.

Theorem 1. *Suppose that the RIC, δ_{3s} , of the measurement matrix A and the step size μ satisfy*

$$\delta_{3s} < \frac{1}{\sqrt{3}} \approx 0.577, \quad \frac{\hat{\eta} - 1}{\hat{\eta}(1 - \delta_{3s})} < \mu < \frac{1 + \hat{\eta}}{\hat{\eta}(\delta_{3s} + 1)}, \quad (14)$$

where $\hat{\eta} := \sqrt{2/(1 - \delta_{2s}^2)}$. Then, the sequence $\{x^p\}$ generated by the MHTP with $y = Ax + e$ satisfies

$$\|x^p - x_S\|_2 \leq C \cdot (\rho_1)^{p-1} + \frac{q_3}{1 - \rho_1} \|e'\|_2, \quad (15)$$

where $S := \mathcal{L}_s(x)$, $e' := Ax_{\bar{S}} + e$, $\rho_1 := \frac{q_1 + \sqrt{q_1^2 + 4q_2}}{2} < 1$,

$$C := \|x^1 - x_S\|_2 + (\rho_1 - q_1) \|x^0 - x_S\|_2, \quad (16)$$

and

$$\begin{aligned} q_1 &:= \lambda\hat{\eta}(\mu\delta_{3s} + |1 - \mu|), \quad q_2 := (1 - \lambda)\hat{\eta}(\mu\delta_{3s} + |1 - \mu|), \\ q_3 &:= \mu\sqrt{\frac{2}{1 - \delta_{2s}^2}} + \frac{\sqrt{1 + \delta_s}}{1 - \delta_{2s}}. \end{aligned} \quad (17)$$

Proof. For x^{p+1} and S^{p+1} in MHTP, we have $|S^{p+1}| = s$. Replacing $z^* = x^{p+1}$ and $T = S^{p+1}$ in Lemma 4 obtains

$$\|x^{p+1} - x_S\|_2 \leq \frac{\|x_{S \setminus S^{p+1}}\|_2}{\sqrt{1 - \delta_{2s}^2}} + \frac{\sqrt{1 + \delta_s}}{1 - \delta_{2s}} \|e'\|_2. \quad (18)$$

Note that $S^{p+1} = \mathcal{L}_s(v^p)$, where $v^p = u^p + \mu A^T(y - Au^p)$ is given in Lemma 3. The term $\|x_{S \setminus S^{p+1}}\|_2$ can be estimated as follows:

$$\begin{aligned} \|x_{S \setminus S^{p+1}}\|_2 &= \left\| (x_S - v^p)_{S \setminus S^{p+1}} + (v^p)_{S \setminus S^{p+1}} \right\|_2 \\ &\leq \left\| (x_S - v^p)_{S \setminus S^{p+1}} \right\|_2 + \left\| (v^p)_{S \setminus S^{p+1}} \right\|_2. \end{aligned} \quad (19)$$

Since $|S| = s$ and S^{p+1} is the index set of s -largest absolute elements of v^p , then

$$\left\| (v^p)_{S \setminus S^{p+1}} \right\|_2 \leq \left\| (v^p)_{S^{p+1} \setminus S} \right\|_2 = \left\| (x_S - v^p)_{S^{p+1} \setminus S} \right\|_2, \quad (20)$$

where the last equality is due to $(x_S)_{S^{p+1} \setminus S} = 0$. It follows from Equations (19) and (20) that

$$\begin{aligned} \|x_{S \setminus S^{p+1}}\|_2 &\leq \left\| (x_S - v^p)_{S \setminus S^{p+1}} \right\|_2 \\ &\quad + \left\| (x_S - v^p)_{S^{p+1} \setminus S} \right\|_2 \leq \sqrt{2} \|(x_S - v^p)_{S^{p+1} \cup S}\|_2. \end{aligned} \quad (21)$$

Using Lemma 3, Equation (21) becomes

$$\|x_{S \setminus S^{p+1}}\|_2 \leq \sqrt{2}(\varrho\lambda\|x^p - x_S\|_2 + \varrho(1-\lambda)\|x^{p-1} - x_S\|_2 + \mu\sqrt{1+\delta_{2s}}\|e'\|_2), \quad (22)$$

where $\varrho = \mu\delta_{3s} + |1 - \mu|$. Inserting Equation (22) into Equation (18) yields

$$\begin{aligned} &\|x^{p+1} - x_S\|_2 \\ &\leq \widehat{\eta}(\lambda(\mu\delta_{3s} + |1 - \mu|)\|x^p - x_S\|_2 + (1-\lambda)(\mu\delta_{3s} + |1 - \mu|)\|x^{p-1} - x_S\|_2) \\ &\quad + \left(\mu\sqrt{\frac{2}{1-\delta_{2s}}} + \frac{\sqrt{1+\delta_s}}{1-\delta_{2s}} \right) \|e'\|_2 \\ &= q_1\|x^p - x_S\|_2 + q_2\|x^{p-1} - x_S\|_2 + q_3\|e'\|_2, \end{aligned} \quad (23)$$

where $\widehat{\eta} = \sqrt{2/(1-\delta_{2s}^2)}$ and q_i ($i = 1, 2, 3$) are given in Equation (17).

Next, let us check whether Lemma 2 is applicable for Equation (23) under the condition Equation (14). From Equation (17) and $\mu > 0$, $\lambda \in (0, 1]$, we see that $q_i \geq 0$ ($i = 1, 2, 3$). Hence, it only needs to prove that $q_1 + q_2 < 1$. Based on the fact $\delta_{2s} \leq \delta_{3s}$ and the condition $\delta_{3s} < \frac{1}{\sqrt{3}}$ in Equation (14), we have

$$\delta_{3s} < \sqrt{\frac{1-\delta_{3s}^2}{2}} \leq \sqrt{\frac{1-\delta_{2s}^2}{2}} = \frac{1}{\widehat{\eta}}, \quad (24)$$

which implies that

$$\frac{\widehat{\eta} - 1}{\widehat{\eta}(1 - \delta_{3s})} < 1 < \frac{1 + \widehat{\eta}}{\widehat{\eta}(1 + \delta_{3s})}. \quad (25)$$

If $\widehat{\eta} - 1/\widehat{\eta}(1 - \delta_{3s}) < \mu \leq 1$, then

$$q_1 + q_2 = \widehat{\eta}[\mu\delta_{3s} + |1 - \mu|] = \widehat{\eta}[1 - \mu(1 - \delta_{3s})] < 1, \quad (26)$$

where the first equality is given by Equation (17). On the other hand, if $1 < \mu < 1 + \widehat{\eta}/\widehat{\eta}(1 + \delta_{3s})$, then

$$q_1 + q_2 = \widehat{\eta}[\mu\delta_{3s} + |1 - \mu|] = \widehat{\eta}[\mu(\delta_{3s} + 1) - 1] < 1. \quad (27)$$

It follows that $q_1 + q_2 < 1$ under Equation (14). Applying Lemma 2 to Equation (23), we obtain Equation (15) with

$$\rho_1 = \frac{q_1 + \sqrt{q_1^2 + 4q_2}}{2} < 1. \quad \square$$

Remark 1. From the above argument, the requirement on δ_{3s} is a key factor in algorithm convergence. It is well-known that if the measurement matrix A is taken as a standard Gaussian random matrix, then the restricted isometry property can be attained with high probability. More precisely, for a standard Gaussian matrix $A \in \mathbb{R}^{m \times n}$, if $0 < \eta, \epsilon < 1$ and

$$m \geq 2\eta^{-2}[s(1 + \ln(n/s)) + \ln(2\epsilon^{-1})], \quad (28)$$

then the restricted isometry constant δ_s of $1/\sqrt{m}A$ satisfies

$$\delta_s \leq [2 + \eta g(n, s)] \cdot \eta g(n, s), \quad (29)$$

with probability greater than or equal to $1 - \epsilon$, where

$$g(n, s) := 1 + [2(1 + \ln(n/s))]^{-1/2}. \quad (30)$$

Please see [24, Chapter 8] for more information.

The following result shows that, in the case of the accurate measurement $y = Ax$, the sequence $\{x^p\}$ generated by the MHTP converges to the s -sparse signal x .

Corollary 1. Suppose that the RIC, δ_{3s} , of the measurement matrix A and the step size μ satisfy Equation (14). Then the sequence $\{x^p\}$ generated by the MHTP with $y = Ax$ and $\|x\|_0 \leq s$ satisfies

$$\|x^p - x\|_2 \leq (\rho_1)^{p-1}(\|x^1 - x\|_2 + (\rho_1 - q_1)\|x^0 - x\|_2), \quad (31)$$

where $\rho_1 < 1$ and q_1 are given in Theorem 1. Moreover, x^p converges to x as $p \rightarrow \infty$.

The following results establish the finite termination properties of the MHTP. It only needs to consider the case where x is nonzero; otherwise, x is recovered at the first iteration because the initial point x_0 is taken as 0 in the MHTP.

Theorem 2. *Suppose that the RIC, δ_{3s} , of the measurement matrix A and the step size μ satisfy Equation (14). Then MHTP can recover the nonzero vector $x \in \mathbb{R}^n$ with $\|x\|_0 \leq s$ from $y = Ax$ in at most*

$$p^* := \left\lceil \frac{\ln(\sqrt{2}\widehat{C}/\theta)}{\ln(1/\rho_1)} \right\rceil + 2, \quad (32)$$

iterations, where $\theta := \min_{i \in S} |x_i|$, $S := \text{supp}(x)$, $\rho_1 < 1$ is given in Theorem 1 and

$$\widehat{C} := \|x^1 - x\|_2 + (\rho_1 - q_1)\|x^0 - x\|_2. \quad (33)$$

Proof. For $v^p = u^p + \mu A^T A(x - u^p)$, we claim that the following condition

$$\begin{aligned} & \left\| (v^p - x)_{\{k,t\}} \right\|_2 \\ & \leq \left\| \mu \lambda [(I - A^T A)(x^p - x)]_{\{k,t\}} \right\|_2 + \lambda |1 - \mu| \left\| (x^p - x)_{\{k,t\}} \right\|_2 \\ & \quad + \mu(1 - \lambda) \left\| [(I - A^T A)(x^{p-1} - x)]_{\{k,t\}} \right\|_2 + (1 - \lambda) |1 - \mu| \left\| (x^{p-1} - x)_{\{k,t\}} \right\|_2 \\ & \leq \lambda [\mu \delta_{2s+1} + |1 - \mu|] \|x^p - x\|_2 + (1 - \lambda) [\mu \delta_{2s+1} + |1 - \mu|] \|x^{p-1} - x\|_2. \end{aligned} \quad (37)$$

Since $\delta_{2s+1} \leq \delta_{3s}$ and $\widehat{\eta} = \sqrt{2/(1 - \delta_{2s}^2)} > 1$, it follows from Equations (26) and (27) that

$$\mu \delta_{2s+1} + |1 - \mu| \leq \mu \delta_{3s} + |1 - \mu| \leq \widehat{\eta} [\mu \delta_{3s} + |1 - \mu|] < 1. \quad (38)$$

Combining Equations (37) and (38) and Corollary 1, we obtain

$$\begin{aligned} \left\| (v^p - x)_{\{k,t\}} \right\|_2 & \leq \lambda \|x^p - x\|_2 + (1 - \lambda) \|x^{p-1} - x\|_2 \\ & \leq \widehat{C} [\lambda (\rho_1)^{p-1} + (1 - \lambda) (\rho_1)^{p-2}] \\ & = \widehat{C} [1 + (\rho_1 - 1) \lambda] (\rho_1)^{p-2} \\ & < \widehat{C} (\rho_1)^{p-2}, \end{aligned} \quad (39)$$

where the last inequality is ensured by $\rho_1 < 1$ and $\lambda \in (0, 1]$, and \widehat{C} is given by Equation (33). It follows that

$$|(v^p - x)_k| + |(v^p - x)_t| \leq \sqrt{2} \left\| (v^p - x)_{\{k,t\}} \right\|_2 < \sqrt{2} \widehat{C} (\rho_1)^{p-2}. \quad (40)$$

$$|(v^p)_k| > |(v^p)_t|, \quad \forall k \in S \text{ and } t \in \bar{S}, \quad (34)$$

is sufficient to ensure $\mathcal{L}_s(v^p) = S$. Indeed, since $S^{p+1} = \mathcal{L}_s(v^p)$ and x^{p+1} is the least-square solution in the subspace S^{p+1} , that is,

$$x^{p+1} = \arg \min_z \{ \|y - Az\|_2 : \text{supp}(z) \subseteq S^{p+1} \}, \quad (35)$$

then $x^{p+1} = x$ in this case. To establish the finite termination property, we just need to show the validity of Equation (34) as $p \geq p^*$, where p^* is given by Equation (32). Since $k \in S = \text{supp}(x)$, we have $|\text{supp}(x^i - x) \cup \{k, t\}| \leq 2s + 1$ as $i = p, p - 1$. From Lemma 1(i), one has

$$\left\| [(I - A^T A)(x^i - x)]_{\{k,t\}} \right\|_2 \leq \delta_{2s+1} \|x^i - x\|_2, \quad i = p, p - 1. \quad (36)$$

This together with Equation (8) yields

Since $\theta = \min_{i \in S} |x_i|$, by using the triangle inequality, we get

$$|(v^p)_k| = |(v^p - x + x)_k| \geq \theta - |(v^p - x)_k|, \quad (41)$$

for any $k \in S$. Because of $t \in \bar{S}$ and $S = \text{supp}(x)$, then $|(v^p)_t| = |(v^p - x)_t|$. For $p \geq p^*$, it follows from Equations (40) and (41) that

$$\begin{aligned} |(v^p - x)_k| + |(v^p)_t| & = |(v^p - x)_k| + |(v^p - x)_t| \\ & < \sqrt{2} \widehat{C} (\rho_1)^{p-2} < \theta \leq |(v^p - x)_k| + |(v^p)_k|, \end{aligned} \quad (42)$$

where the second inequality is ensured by Equation (32) and $p \geq p^*$. This shows the validity of Equation (34) and the proof is complete. \square

Given two integers i, j , define

$$\sigma_j(x)_i := \inf_v \{ \|x - v\|_i : \|v\|_0 \leq j \}, \quad (43)$$

which describes the error of the best j -term approximation of a vector x with respect to i -norm [24, Definition 2.2]. Recall in the following lemma a useful inequality associated with

$\sigma_j(z)_2$ and $\|z\|_1$, which is used to analyze the stability of MHTP.

Lemma 5. [24, Theorem 2.5]. *For a given $z \in \mathbb{R}^n$ and a positive integer j , we have*

$$\sigma_j(z)_2 \leq \frac{1}{2\sqrt{j}} \|z\|_1. \quad (44)$$

The stability of MHTP is given in the following result.

Theorem 3. *Suppose that the RIC, δ_{3s} , of the measurement matrix A and the step size μ satisfy Equation (14). Then for a given $x \in \mathbb{R}^n$ and $e \in \mathbb{R}^m$, the sequence $\{x^p\}$ generated by the MHTP with $y = Ax + e$ satisfies*

$$\begin{aligned} \|x - x^{p+1}\|_2 &\leq \left(\frac{1}{2} + \frac{q_3 \sqrt{1 + \delta_\tau}}{1 - \rho_1} \right) \frac{1}{\sqrt{\tau}} \sigma_k(x)_1 \\ &+ \frac{q_3}{1 - \rho_1} \|e\|_2 + C(\rho_1)^p, \end{aligned} \quad (45)$$

where ρ_1 , q_3 , C are given in Theorem 1, $k := \lceil s/2 \rceil$ and $\tau := s - k$.

Moreover, every cluster point of the sequence $\{x^p\}$, say x^* , satisfies

$$\|x - x^*\|_2 \leq \left(\frac{1}{2} + \frac{q_3 \sqrt{1 + \delta_\tau}}{1 - \rho_1} \right) \frac{1}{\sqrt{\tau}} \sigma_k(x)_1 + \frac{q_3}{1 - \rho_1} \|e\|_2. \quad (46)$$

Proof. Using the triangle inequality, we have

$$\|x - x^{p+1}\|_2 = \|x_{\bar{S}} + x_S - x^{p+1}\|_2 \leq \|x_{\bar{S}}\|_2 + \|x_S - x^{p+1}\|_2. \quad (47)$$

Denote $S_1 := \mathcal{L}_k(x)$, $S_2 := \mathcal{L}_\tau(x_{S_1}^-)$, \dots , $S_{l-1} := \mathcal{L}_\tau(x_{S_1 \cup S_2 \cup \dots \cup S_{l-2}}^-)$, $S_l := \mathcal{L}_r(x_{S_1 \cup S_2 \cup \dots \cup S_{l-1}}^-)$, where the positive integers l, r satisfy $(l-2)\tau + k + r = n$ and $r \leq \tau$. Note that $S_1 \cup S_2 = S$ and $\bar{S} = \sum_{i=3}^l S_i$. For $S_1 = \mathcal{L}_k(x)$, $S = \mathcal{L}_s(x)$ and $k = s - \tau$, we have

$$\|x_{\bar{S}}\|_2 = \sigma_{s-k}(x_{S_1}^-)_2 \leq \frac{1}{2\sqrt{s-k}} \|x_{S_1}^-\|_1 = \frac{1}{2\sqrt{\tau}} \sigma_k(x)_1, \quad (48)$$

where the equalities can be obtained from Equation (43) and the inequality follows from Lemma 5. By Theorem 1 and the triangle inequality, one has

$$\|x_S - x^{p+1}\|_2 \leq C(\rho_1)^p + \frac{q_3}{1 - \rho_1} \|Ax_{\bar{S}} + e\|_2 \quad (49)$$

$$\leq C(\rho_1)^p + \frac{q_3}{1 - \rho_1} \left(\sum_{i=3}^l \|Ax_{S_i}\|_2 + \|e\|_2 \right). \quad (50)$$

TABLE 1: The description of the parameters and the abbreviation of the algorithm name.

Abbreviation	Description
λ	The coefficient of the convex combination
μ	The coefficient of the momentum term
MHTP	Modified hard thresholding pursuit
HTP	Hard thresholding pursuit
NSHTP	Newton-step-based hard thresholding pursuit
CoSaMP	Compressive sampling matching pursuit
SP	Subspace pursuit
gOMP	Generalized orthogonal matching pursuit

It follows from [24, Lemma 6.10] that

$$\|x_{S_j}\|_2 \leq \frac{1}{\sqrt{\tau}} \|x_{S_{j-1}}\|_1, \quad j = 3, \dots, l. \quad (51)$$

Combining Equation (3) with Equation (51) yields

$$\begin{aligned} \sum_{i=3}^l \|Ax_{S_i}\|_2 &\leq \sqrt{1 + \delta_\tau} \sum_{i=3}^l \|x_{S_i}\|_2 \leq \sqrt{1 + \delta_\tau} \frac{1}{\sqrt{\tau}} \sum_{i=2}^{l-1} \|x_{S_i}\|_1 \\ &\leq \sqrt{1 + \delta_\tau} \frac{1}{\sqrt{\tau}} \sigma_k(x)_1, \end{aligned} \quad (52)$$

where the last inequality is due to the fact

$$\sum_{i=2}^{l-1} \|x_{S_i}\|_1 \leq \|x_{S_1}^-\|_1 = \sigma_k(x)_1. \quad (53)$$

Combining Equations (47)–(49) with Equation (52) yields Equation (45).

From Equation (45) and $\rho_1 < 1$, we conclude that the sequence $\{x^p\}$ is bounded and hence there exists at least one cluster point. Assume without loss of generality that x^* is a cluster point of $\{x^p\}$. The Formula (46) follows by taking $p \rightarrow \infty$ in Equation (45). \square

4. Numerical Experiments

All experiments are conducted on a personal computer with the processor Intel(R) Core(TM) i5-7200U CPU @ 2.50 GHz and 4 GB memory. To demonstrate the effectiveness of MHTP for sparse signal recovery, we compare it with several mainstream algorithms in synthetic and real-world experiments. For the sake of convenience, we provide the following parameter descriptions and abbreviated forms for these comparing algorithms used in numerical experiments in Table 1. There are a total of six algorithms involved.

Let $x^1 = x^0 = 0$ be the initial points for the MHTP and $x^0 = 0$ be the initial point for other algorithms. Because the

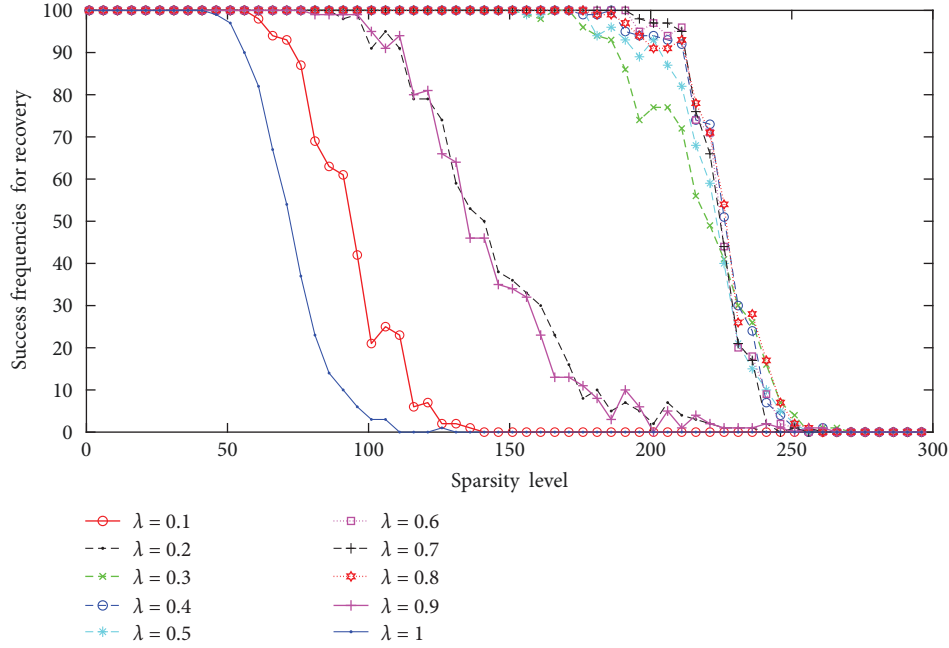


FIGURE 1: Successful frequencies of the modified hard thresholding pursuit (MHTP) with different λ by accurate measurements.

dimensions in this study are larger than those in [30], the algorithmic parameters in the NSHTP are taken as $\mu_1 = 22$ and $\epsilon = \max\{\sigma_1^2 + 1, \mu_1 - \sigma_m^2\}$ through numerical tests to acquire the better recover capability, where σ_1 and σ_m are the largest and smallest singular values of the realized sensing matrix, respectively. Let $\min\{s, m/N\}$ be the maximum number of iterations for the gOMP [36], where N represents the index corresponding to N largest correlation in magnitude. For other algorithms, the maximum numbers of iterations are set to 50.

4.1. Synthetic Data. In this subsection, we discuss the performances of these algorithms using synthetic data. All measurement matrices $A \in \mathbb{R}^{m \times n}$ and sparse vectors $x^* \in \mathbb{R}^n$ are randomly generated with $n = 2000$, whose elements are independently and identically distributed. Specifically, A is taken as a standard Gaussian random matrix, and the non-zero entries of x^* and their positions follow standard normal and uniform distributions, respectively. The experiments are performed with accurate measurement $y := Ax^*$ and inaccurate measurement $y := Ax^* + \epsilon h$, where ϵ is the noise level and h is a standard Gaussian noise vector. Let

$$\|x^p - x^*\|_2 / \|x^*\|_2 \leq 10^{-3}, \quad (54)$$

be the criterion for a successful recovery.

4.1.1. Successful Frequencies and Average Recover Time. In this section, we compare the recover capability of these algorithms with a fixed $m = 500$, and the sparsity level s ranges from 1–296 with a step size of 5. For each given s , 100 random instances are used to test the successful frequencies. We first consider the influence of the parameter λ on the recover capability of the MHTP with a step size of $\mu = 0.006$.

The corresponding results are shown in Figures 1 and 2, wherein Figures 1 and 2 correspond to the accurate and inaccurate measurement with the noise level $\epsilon = 0.005$. The novel idea of the MHTP is the combination of the current point x^p and the previous point x^{p-1} , where the parameter λ is the coefficient of this convex combination. In general, different values of λ represent different weights of information used for x^p and x^{p-1} . It is preferable that λ is not selected too close to 1 or 0, which helps us make more use of the information provided by the two points x^p and x^{p-1} . The numerical experiment illustrates this phenomenon: the recover capability of the MHTP is weaker as $\lambda \in \{0.1, 0.2, 0.9, 1\}$ whether in the absence or presence of noise. However, the recover capability of the MHTP is powerful and does not significantly change with respect to λ because it lies in the interval $[0.3, 0.8]$. Based on this observation, we set $\lambda = 0.6$ and $\mu = 0.006$ for MHTP in the remainder of this subsection.

Let $a := \widehat{\eta}(\mu\delta_{3s} + |1 - \mu|)$ and $b := 1/a$. Note that $a < 1$ and $b > 1$ under Equation (14) in Theorem 1. Since $q_1 = \lambda a$ and $q_2 = (1 - \lambda)a$, then

$$\rho_1 = \frac{\lambda a + \sqrt{\lambda^2 a^2 + 4(1 - \lambda)a}}{2} = \frac{a}{2} \left(\lambda + \sqrt{\lambda^2 + 4(1 - \lambda)b} \right). \quad (55)$$

Let $F(\lambda) := \lambda + \sqrt{\lambda^2 + 4(1 - \lambda)b}$. Then

$$F'(\lambda) = 1 + \frac{2\lambda - 4b}{2\sqrt{\lambda^2 + 4(1 - \lambda)b}} < 0. \quad (56)$$

Thus $F(\lambda)$ strictly decreases in the interval $[0, 1]$. In theory, under the convergence condition Equation (14), ρ_1

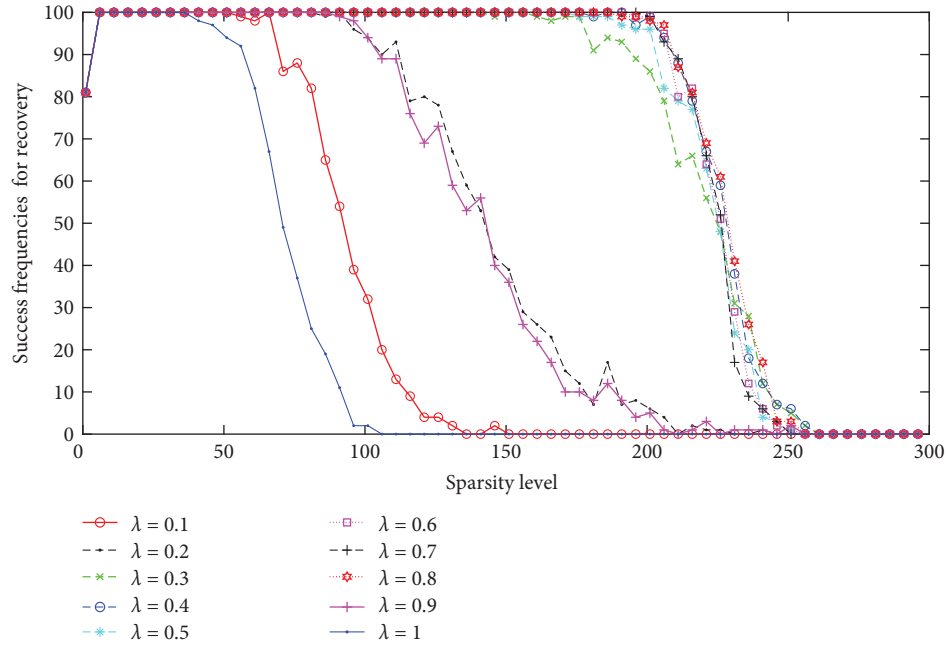


FIGURE 2: Successful frequencies of the MHTP with different λ by inaccurate measurements with $\epsilon = 0.005$.

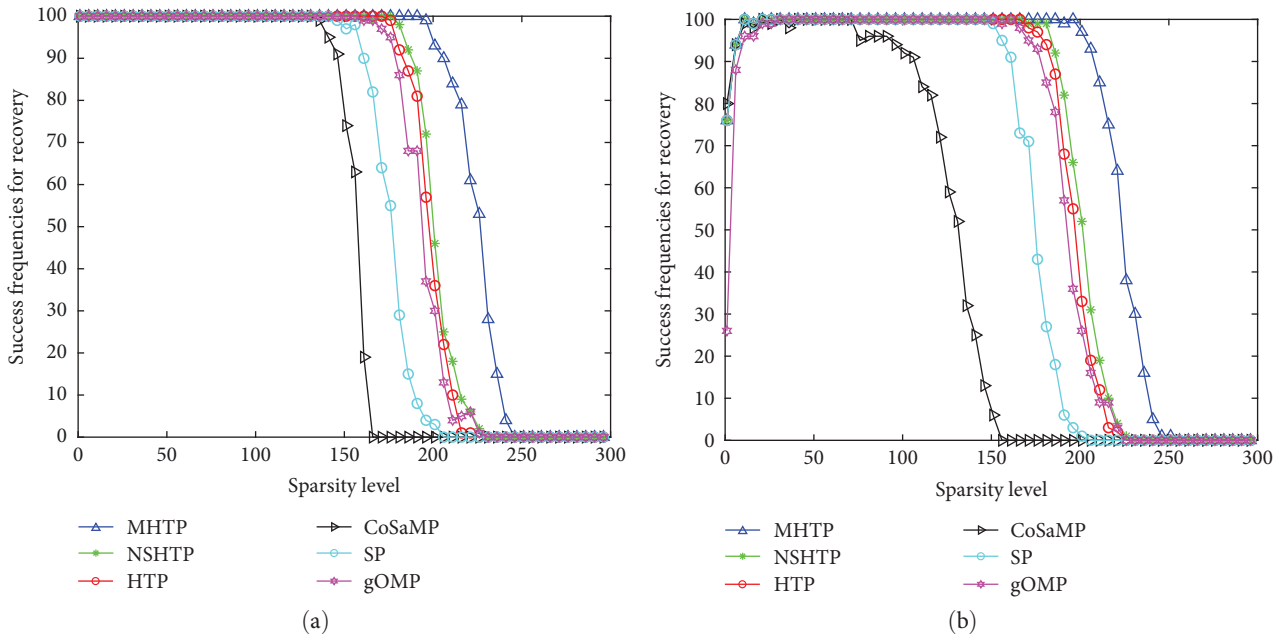


FIGURE 3: Comparison of successful frequencies. (a) Accurate measurements and (b) inaccurate measurements.

decreases with a increased λ , which corresponds to a better recovery effect. This is consistent with the phenomenon shown in Figures 1 and 2, where $\lambda \in [0.1, 0.8]$. However, checking condition Equation (14) is intractable because of the difficulty of computing the RIC δ_{3s} . By contrast, the theoretical analysis above is not consistent with the numerical results shown in Figures 1 and 2 with $\lambda \in [0.9, 1]$, which may be the results of the range of μ in Equation (14) not being satisfied or other numerical limitations, such as the maximum number of iterations. This is not clear at present

and will be further discussed in the future. Note that there often exists a gap between the theoretical and numerical results for sparse optimization algorithms, including the MHTP, owing to the theoretical analysis technique or numerical limitations.

In Figure 3, the performances of the algorithms, including MHTP, NSHTP, HTP, SP, gOMP, and CoSaMP are compared under noiseless and noisy settings. Let $\mu_2 = 0.003$ be the step size for the HTP and let the parameter for the gOMP be $N = 5$. These parameters are selected via

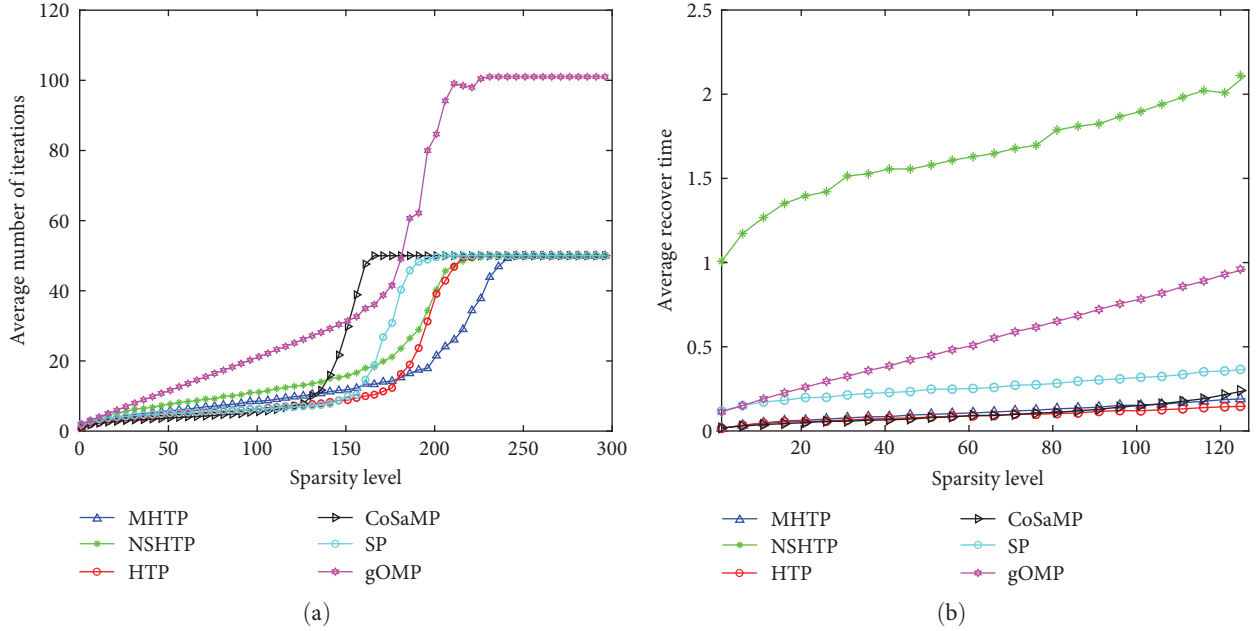


FIGURE 4: Comparison of recovery efficiency. (a) Average number of iterations and (b) average recover time.

TABLE 2: The complexity of algorithms after p iterations.

Algorithms	Complexity
CoSaMP	$O(pmn + pm^3)$
gOMP	$O(pmn + p^2m)$
HTP	$O(pmn + pm^3)$
MHTP	$O(pmn + pm^3)$
NSHTP	$O(pmn^2 + pn^3 + ps^3)$
SP	$O(pmn + pm^3)$

numerical tests based on the recover capabilities of the algorithms. The results reveal that the MHTP performs better than other algorithms, even in the noisy settings, and all algorithms are stable under small perturbation with $\epsilon = 0.01$ except for CoSaMP.

Corresponding to Figure 3(a), comparisons of the average number of iterations and average recover time (CPU) of these algorithms in noiseless environments are shown in Figure 4. From Figure 4(a), it is clearly observed that the average number of gOMP depends on the sparsity level s and its maximum is $m/5 = 100$, whereas the average number of other algorithms is not more than 50. The algorithm is regarded as failing to recover the sparse vector provided that the average number reaches the maximum. Figure 4(a) indicates that the average numbers of all algorithms except the MHTP reach their maximums at $s \leq 220$, that is, only the MHTP can recover the sparse signal as $s \in [220, 240]$, which is consistent with Figure 3(a).

A comparison of the average recover time for these algorithms with $s \leq 126$ is shown in Figure 4(b). The average recover time is closely related to the computational complexity of the algorithm. Given a measurement matrix $A \in \mathbb{R}^{m \times n}$ and a s -sparse vector x^* , the complexity of these algorithms

after p iterations is shown in Table 2. It is not difficult to see that the complexities of HTP, MHTP, CoSaMP, and SP are of the same order, while that of NSHTP is the largest and that of gOMP is the smallest (because the iteration step for gOMP being dependent on s and less than m). In Figure 4(b), NSHTP requires more time than the other algorithms, which is consistent with the results in Table 2. The result of Figure 4(b) shows that the average time consumed by gOMP and SP is much longer than that consumed by MHTP and HTP, and HTP always requires less time than MHTP. Recall from Figure 4(a) that the average number of iterations of the MHTP is greater than that of the HTP for successful recovery as $s < 180$. However, the average time of the MHTP is almost the same as that of the CoSaMP as $s \leq 80$, but the latter increases quickly with an increase in s as $s \geq 80$. In summary, the MHTP is competitive with other algorithms in terms of recover capability and average recover time.

4.1.2. Phase Transition Curves. In the following experiments, we display the performance of these algorithms in terms of Phase Transition Curves (PTC) [32, 37] using logistic regression curves with a 50% success rate. Different from the logistic regression model used in [32, 37], we select ‘glmfit’ function in Matlab to gain PTC directly. Let $\kappa := m/n$ be the undersampling rate and $\varpi := s/m$ be the oversampling rate. The (κ, ϖ) -plane is often called the phase transition plane, where the region below the PTC corresponds to the successful recovery region, and the region above corresponds to the failure region. Detailed information regarding the PTC and recovery regions can be found in [37]. To generate PTC, we collect 13 groups of κ by equally dividing the interval $[0.1, 0.7]$, and determine the values of ϖ in PTC using ‘glmfit’ function.

Let us further discuss the influence of λ on MHTP from a PTC perspective. Take 10 different values of $\lambda \in [0.1, 1]$ to

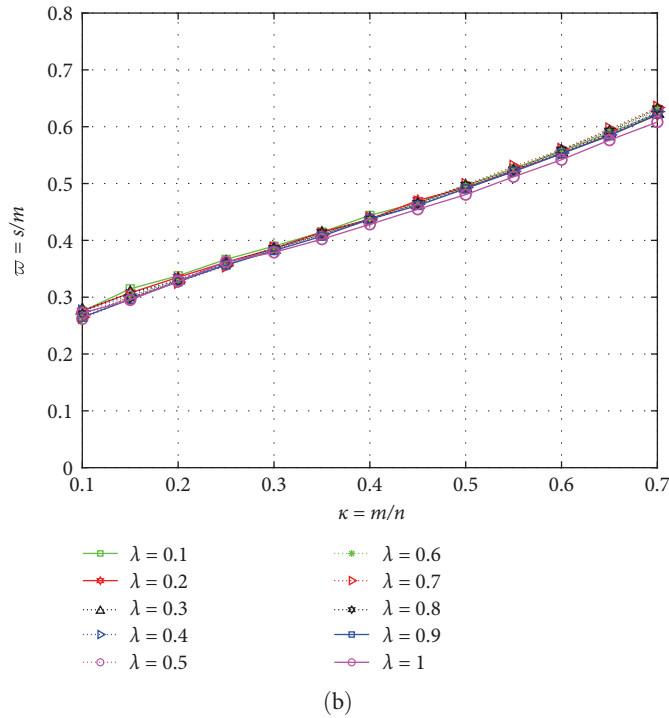
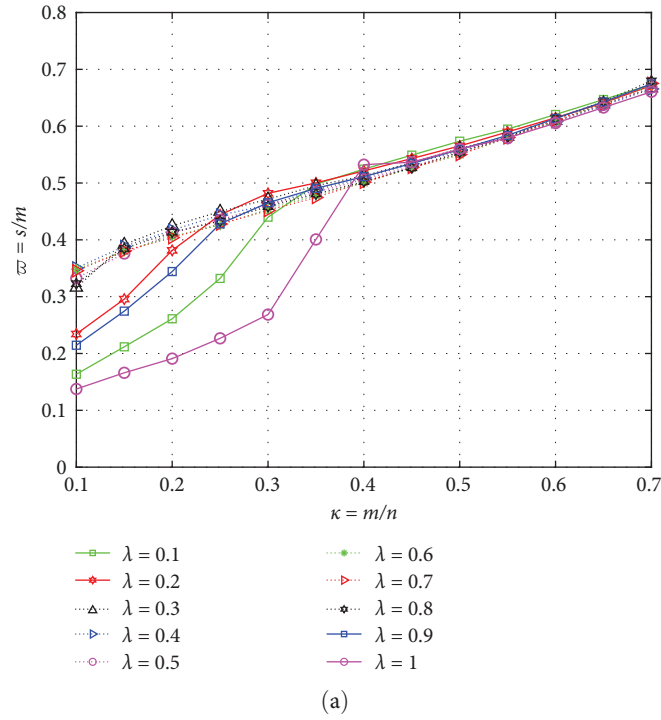


FIGURE 5: The Phase Transition Curves (PTC) of the MHTP for different λ . (a) $\mu = 2.5/m$ and (b) $\mu = 1/m$.

generate the PTCs in Figure 5, where Figures 5(a) and 5(b) correspond to $\mu = 2.5/m$ and $\mu = 1/m$, respectively. In Figure 5(a), we find that the PTC with $\lambda \in \{0.1, 0.2, 0.9, 1\}$ is lower than that of the others as κ is small, which indicates that the recover capability of the MHTP in these cases is weaker. As $\lambda \in [0.3, 0.8]$, the corresponding PTC does not change significantly. This phenomenon shows that the MHTP admits a relatively larger step size μ with a suitable

λ compared with that of the HTP, which is a special case of the MHTP with $\lambda = 1$. However, as shown in Figure 5(b), the recovery performance of the MHTP is insensitive to changes in λ when $\mu = 1/m$. By picking $\lambda = 0.6$, we select different μ to generate the PTC as shown in Figure 6, indicating that the recover capability of the MHTP with $\mu = 2.5/m$ is stronger than that of $\mu = 1/m$. That is, a stronger recover capability of the MHTP can be obtained by increasing the step size μ

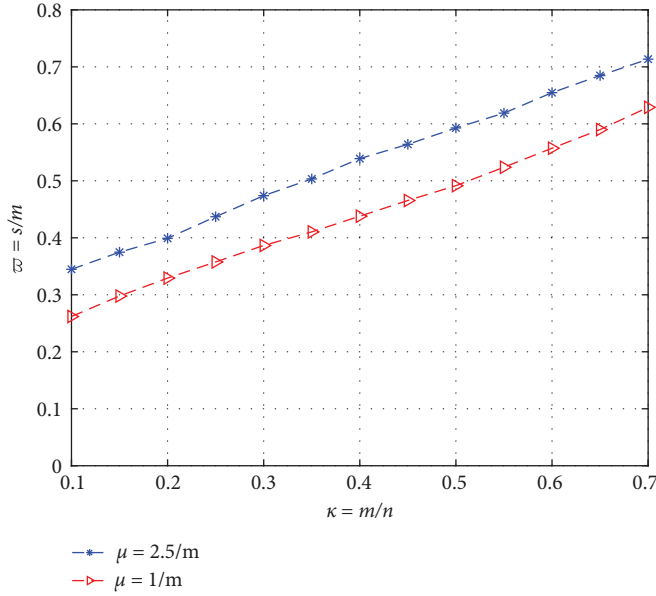


FIGURE 6: The PTC of the MHTP for different μ with $\lambda = 0.6$.

properly. However, acquiring the optimal step size μ is intractable in practice. Subsequently, we discuss the relationship between μ and m based on the recover capability.

It is well known that the step size plays a vital role in the gradient descent method, and hence, the choice of step size μ in the MHTP is crucial for the recovery of sparse signals. Base on recover capability, we select the appropriate parameter μ for MHTP through numerical experiments as m ranged from 200 to 1,400 with a step size 200. In this process, the recommended value of μ is given as

$$\mu \in \{0.012, 0.006, 0.0050, 0.0045, 0.004, 0.0038, 0.0036\}. \quad (57)$$

Using Lagrange interpolation, we establish the following relation between m and μ from the above data approximately

$$\begin{aligned} \mu(m) := & 3.906m^6 \times 10^{-20} - 2.474m^5 \times 10^{-16} \\ & + 6.276m^4 \times 10^{-13} \\ & - 8.151m^3 \times 10^{-10} + 5.723m^2 \times 10^{-7} \\ & - 2.09m \times 10^{-4} + 0.0365, \end{aligned} \quad (58)$$

and the corresponding curve is shown in Figure 7(a). Similarly, the recommended value of μ_2 in HTP is given by

$$\mu_2 \in \{0.006, 0.004, 0.003, 0.0029, 0.0028, 0.0024, 0.0023\}, \quad (59)$$

which results in

$$\begin{aligned} \mu_2(m) := & 3.038m^6 \times 10^{-20} - 1.771m^5 \times 10^{-16} \\ & + 4.132m^4 \times 10^{-13} \\ & - 4.906m^3 \times 10^{-10} + 3.109m^2 \times 10^{-7} \\ & - 1.006m \times 10^{-4} + 0.016, \end{aligned} \quad (60)$$

see Figure 7(b).

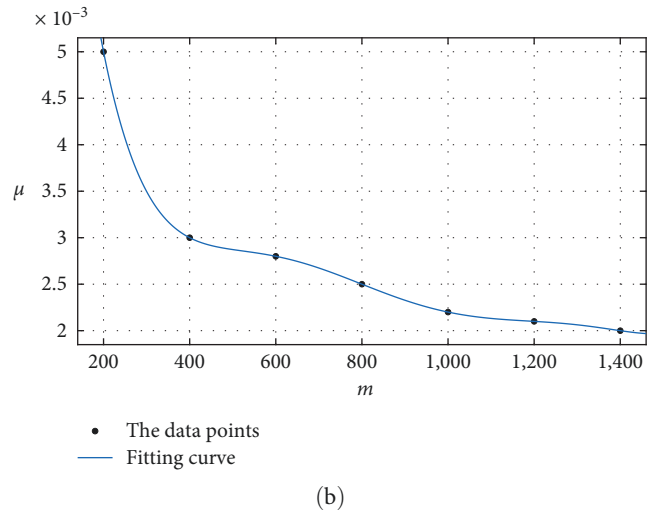
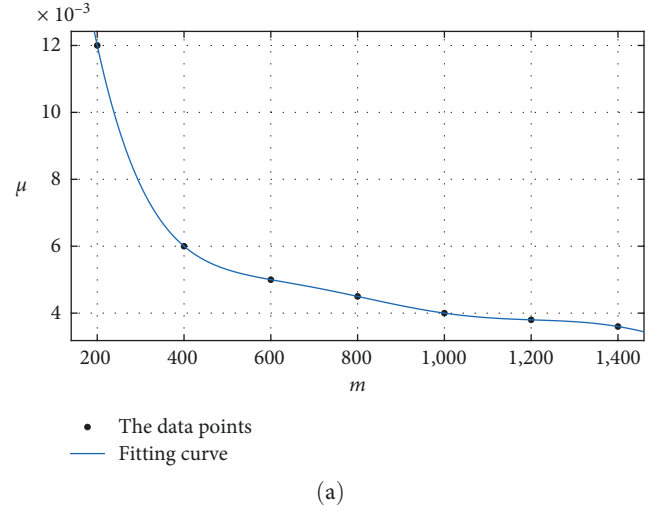


FIGURE 7: The relation between m and μ . (a) MHTP and (b) HTP.

In the following experiments, the parameters $\mu(m)$ and $\mu_2(m)$ are obtained from Equations (58) and (60), respectively. The corresponding curves are presented in Figure 7, which show that the value of μ is nonlinear and decreases as m increases. The trends for the MHTP and HTP are similar. Numerically, the value of $\mu(m)$ is almost twice that of $\mu_2(m)$ for the fixed m . A comparison of the PTC is shown in Figure 8, where Figure 8(a) corresponds to accurate measurements, whereas Figure 8(b) corresponds to inaccurate measurements with a noise level $\epsilon = 0.01$. Figure 8 shows that the PTC of the MHTP is the highest in the noiseless and noisy settings, indicating that recover capability of the MHTP is stronger than other algorithms. Compared with Figures 8(a) and 8(b), we find that all algorithms, except CoSaMP, are stable under a small disturbance account $\epsilon = 0.01$.

In the common recovery region of the algorithms, we compare their average recover time in noiseless environments, see Figures 9–11, where ϖ is ranged from 0.02 to 0.8 with a step size 0.02. For a given (κ, ϖ) , 10 random

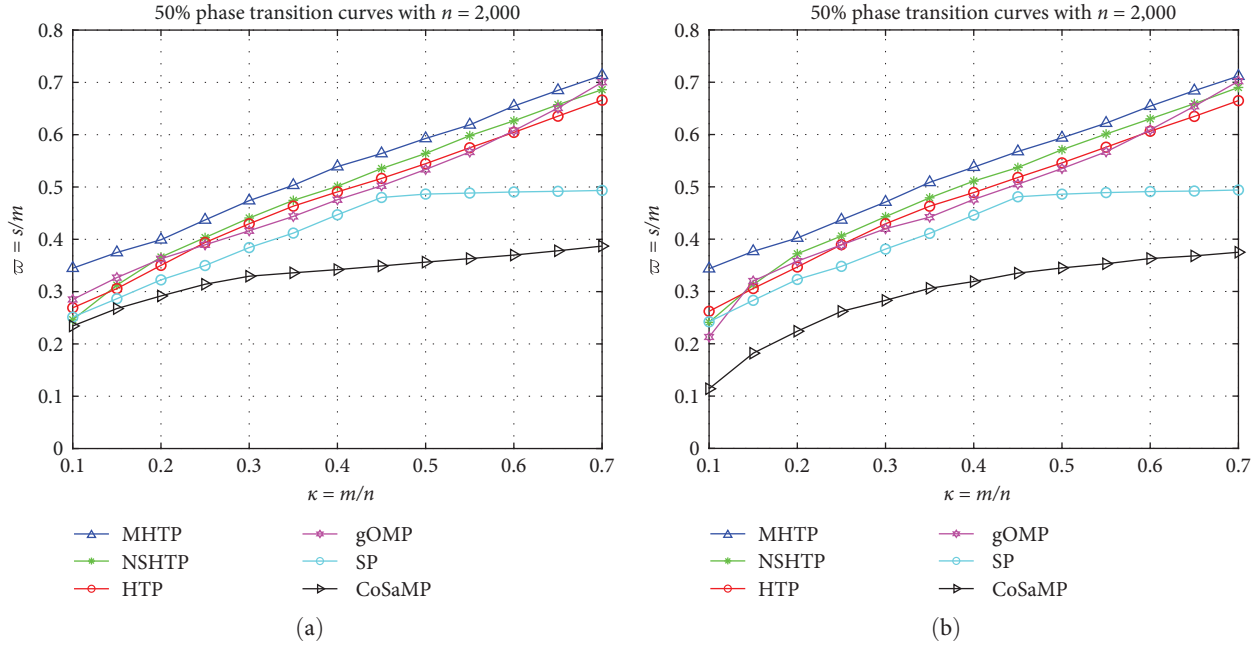


FIGURE 8: The PTC of algorithms. (a) Accurate measurements and (b) inaccurate measurements.

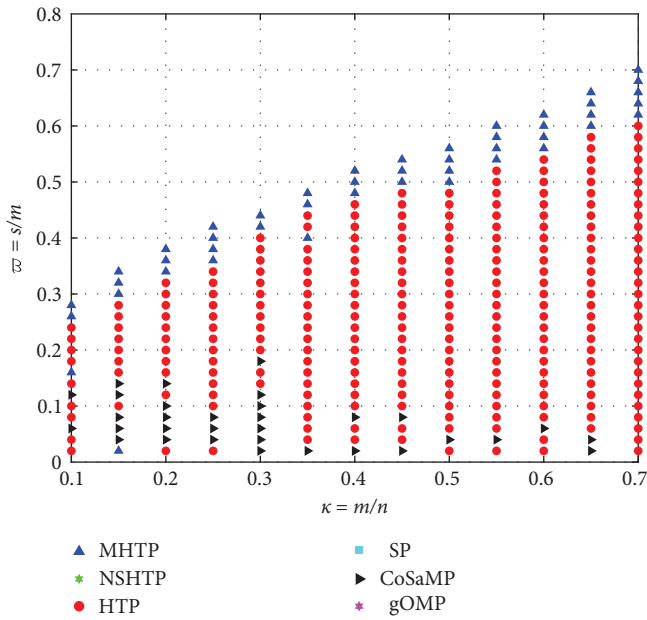


FIGURE 9: Algorithm selection map in noiseless environments.

data (A, x^*) are tested, and the average runtime for recovery is recorded when the success rate is greater than or equal to 50%. The algorithm with the shortest average recover time is selected, as shown in Figure 9, often called an Algorithm Selection Map [32, 37]. It should be noted that only three algorithms including MHTP, HTP, and CoSaMP are shown in the Figure 9 because the other algorithms are always slower than one of them. Figure 9 shows that the MHTP is the fastest algorithm when ϖ is large, whereas HTP and

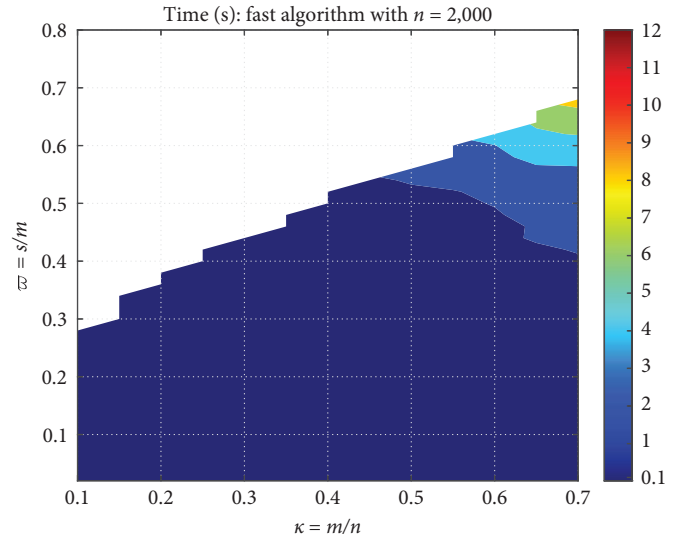


FIGURE 10: Average runtime for the fastest algorithm.

CoSaMP are the fastest algorithms when ϖ is medium and small.

The average runtime for successful recovery of the fastest algorithm is shown in Figure 10, which reveals that the average time consumed by the fastest algorithm is less than one second as $\varpi \leq 0.4$, whereas it significantly increases in other regions when ϖ is large. The ratios of average runtime of the other algorithms to that of the fastest algorithm are summarized in Figure 11. The ratio of MHTP is less than 1.5 in most areas, and is close to 1 for large ϖ , which implies that the advantage of MHTP lies in the region with larger ϖ . Figure 11(c) shows that the average runtime of HTP is close

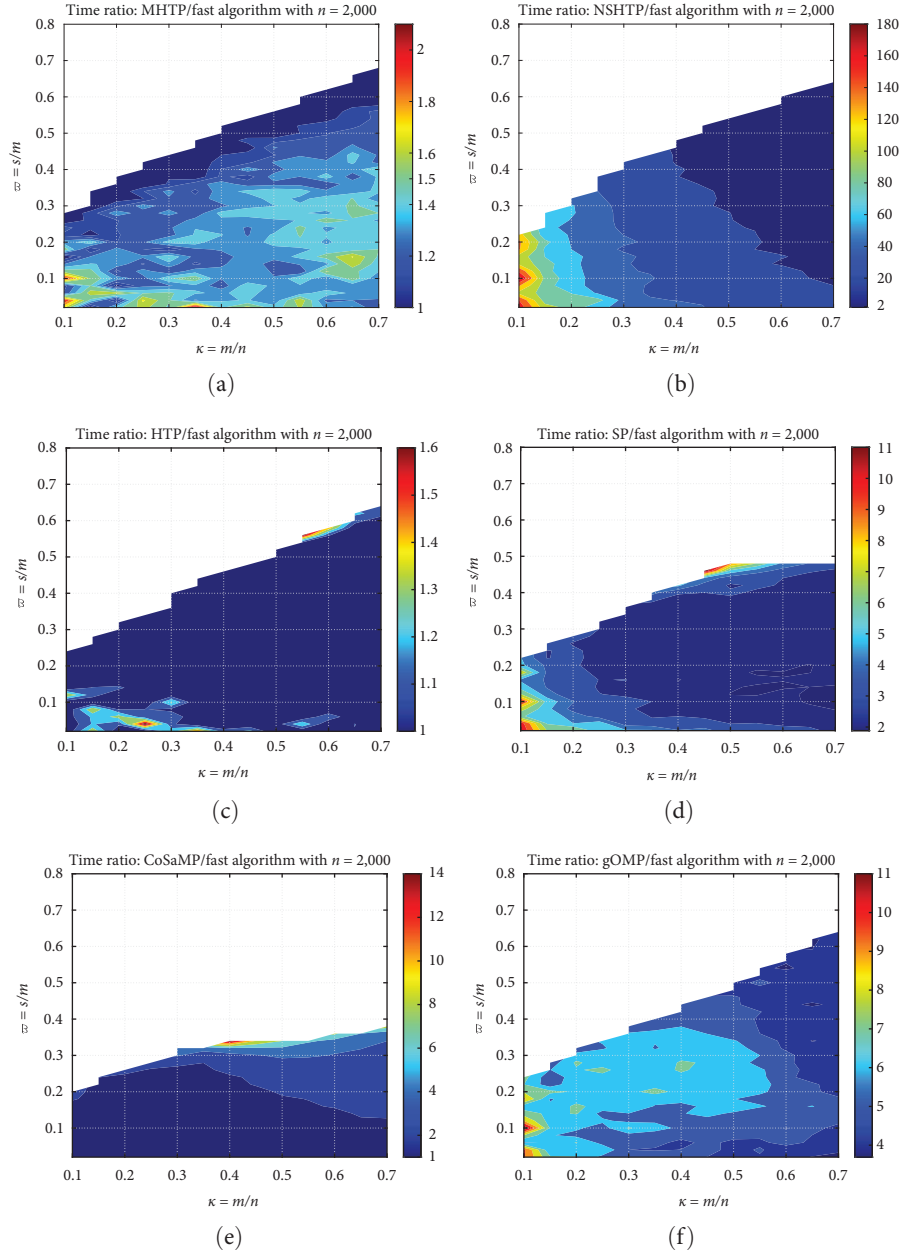


FIGURE 11: The average recover time of algorithms. (a) MHTP, (b) NSHTP, (c) HTP, (d) SP, (e) CoSaMP, and (f) gOMP.

to that of the fastest algorithm almost everywhere. Figure 11(e) shows that the ratio of CoSaMP is close to 1 in the regions with a small sparsity level, which corresponds to the case with small κ or large κ and small ϖ . However, the ratio lies in the interval $[2, 6]$ with the increase of κ and ϖ . From Figure 11(b), we see that NSHTP takes more time than other algorithms because its ratio is greater than 20 in many areas. Figures 11(d) and 11(f) show that the ratios corresponding to SP and gOMP are nearly 2–4 and 4–7, respectively. These phenomena indicate that MHTP and HTP are faster than NSHTP, SP, CoSaMP, and gOMP in most cases.

4.2. Reconstruction of Real Images. In the following real-world experiments, we use standard test images, including

the Lena, Goldhill, Peppers, and Barbara images, to evaluate the performances of these algorithms. The size of each image is 512×512 pixels. For a given image $X \in \mathbb{R}^{n \times n}$ with $n = 512$, using the discrete wavelet transform, we obtain $DXD^T = \hat{X}$, where $\hat{X} \in \mathbb{R}^{n \times n}$ is a sparse or compressible matrix and $D \in \mathbb{R}^{n \times n}$ is an orthonormal wavelet matrix. Taking a normalized standard Gaussian matrices B in $\mathbb{R}^{m \times n}$ as the measurement matrix of X , where $m = \lceil \kappa \cdot n \rceil$ and κ is the sampling rate, we obtain the measurement $Y = BX \in \mathbb{R}^{m \times n}$. It follows that $\hat{Y} = A\hat{X}$, where $\hat{Y} = YD^T$ and $A = BD^T$ are m by n matrices. We now aim to recover \hat{X} from the linear measurements $\hat{Y} = A\hat{X}$, and reconstruct the original image X .

Comparisons of the peak signal-to-noise ratio (PSNR) and runtime of these algorithms with the sampling rates

TABLE 3: Comparison of PSNR (dB) and runtime (seconds) of algorithms with “sym8” wavelet.

	κ	MHTP		NSHTP		HTP		gOMP		SP		CoSaMP		ΔP	ΔT
		PSNR	Time	PSNR	Time	PSNR	Time	PSNR	Time	PSNR	Time	PSNR	Time		
Lena	0.3	32.10	15.55	31.10	98.83	31.83	15.40	31.40	11.10	28.76	63.19	28.59	47.26	0.00	4.45
	0.4	34.11	20.77	33.33	122.94	33.76	20.93	33.11	24.66	32.73	87.47	27.32	59.58	0.00	0.00
	0.5	35.52	45.92	35.21	153.44	35.37	45.84	34.46	47.20	35.16	95.77	34.04	96.64	0.00	0.07
Goldhill	0.3	29.73	15.40	30.14	98.23	29.70	15.88	29.70	11.55	29.15	63.56	28.52	46.82	0.41	3.85
	0.4	30.71	20.67	30.56	121.60	30.47	20.77	30.14	25.88	30.07	87.70	27.39	59.55	0.00	0.00
	0.5	31.64	45.75	31.39	152.65	31.50	45.92	30.83	48.67	31.32	95.77	30.29	96.08	0.00	0.00
Peppers	0.3	30.66	15.47	30.34	98.09	30.88	15.54	30.46	11.41	28.78	63.34	28.12	46.99	0.22	4.06
	0.4	32.74	20.52	31.94	121.98	32.44	20.87	31.79	25.35	31.41	87.72	27.51	59.77	0.00	0.00
	0.5	34.30	45.66	33.78	152.96	34.09	45.80	33.24	48.04	33.96	95.95	32.86	96.47	0.00	0.00
Barbara	0.3	30.45	15.65	30.35	97.91	30.41	15.40	30.24	11.38	28.89	63.28	28.39	46.71	0.00	4.27
	0.4	32.14	20.53	31.58	121.36	31.75	20.80	31.31	26.11	31.18	87.70	27.58	59.20	0.00	0.00
	0.5	33.48	45.88	33.14	154.26	33.32	45.71	32.57	47.80	33.16	96.25	32.25	96.39	0.00	0.18

Bold values represent the maximum value of PSNR among all algorithms under the same set of tests.

TABLE 4: Comparison of PSNR (dB) and runtime (seconds) of algorithms with “coif5” wavelet.

	κ	MHTP		NSHTP		HTP		gOMP		SP		CoSaMP		ΔP	ΔT
		PSNR	Time	PSNR	Time	PSNR	Time	PSNR	Time	PSNR	Time	PSNR	Time		
Lena	0.3	32.32	13.27	31.60	93.66	32.10	13.32	31.83	9.88	29.29	58.61	28.21	41.13	0.00	3.40
	0.4	34.33	18.72	33.78	117.44	33.94	18.80	33.22	22.00	33.14	80.42	27.53	53.35	0.00	0.00
	0.5	35.51	42.29	35.35	149.09	35.37	42.47	34.60	42.32	35.24	88.06	34.23	87.66	0.00	0.00
Goldhill	0.3	29.70	13.76	30.89	94.03	29.85	13.87	29.53	10.31	29.79	58.69	28.47	41.92	1.19	3.45
	0.4	30.62	18.27	30.82	116.52	30.46	18.33	30.02	23.09	30.21	79.88	27.64	52.98	0.20	0.00
	0.5	31.56	42.30	31.60	147.15	31.49	42.49	30.92	43.66	31.43	87.78	30.61	89.11	0.03	0.00
Peppers	0.3	31.37	13.35	31.17	94.04	31.16	13.54	30.96	10.02	29.65	58.46	28.28	41.33	0.00	3.34
	0.4	33.18	18.28	32.80	116.37	32.91	18.34	32.07	22.23	32.17	80.04	27.62	53.02	0.00	0.00
	0.5	34.51	42.38	34.18	147.33	34.44	42.40	33.61	43.19	34.28	88.03	33.07	88.27	0.00	0.00
Barbara	0.3	30.81	13.31	30.46	93.63	30.76	13.39	30.54	10.49	29.72	58.79	28.43	41.49	0.00	2.83
	0.4	32.40	18.17	32.09	117.88	32.11	18.23	31.52	22.85	31.68	80.10	27.63	53.53	0.00	0.00
	0.5	33.79	42.03	33.38	147.83	33.54	42.11	32.88	43.18	33.54	87.57	32.52	87.62	0.00	0.00

Bold values represent the maximum value of PSNR among all algorithms under the same set of tests.

$\kappa \in \{0.3, 0.4, 0.5\}$ are shown in Tables 3 and 4, wherein the “sym8” and “coif5” wavelet are used to acquire the sparse representation, respectively. In the experiments, the sparsity level is set as $s = \lceil 0.15 \cdot n \rceil$ and the parameter N in gOMP is selected as $N = 3$. Corresponding to $\kappa \in \{0.3, 0.4, 0.5\}$, the step sizes μ_2 (for HTP) and μ (for MHTP) are set to $\{2.3, 1.6, 1.4\}$ and $\{3.4, 2.8, 2.4\}$, respectively.

For a given image and a fixed κ , let HP be the largest value of the PSNR (see bold) and ST be the minimum runtime. Let MP , MT represent the PSNR and runtime of MHTP, respectively. We denote $\Delta P := HP - MP$ and $\Delta T := MT - ST$. Thus, the reconstruction capability of MHTP increases with the decrease of ΔP . Similarly, the speed of MHTP increases with a reduced ΔT ; in particular, MHTP is the fastest algorithm when $\Delta T = 0$. From the values of ΔP or PSNR (see bold), we observe that the PSNR of MHTP is the largest in most cases whenever using coiflet or symlet wavelet transform; that is, MHTP outperforms the other algorithms in terms of reconstruction quality (compare Tables 3 and 4).

In particular, when $\kappa = 0.3$, the MHTP consumes more time than the gOMP, however, the MHTP is faster than the gOMP in other situations. Moreover, the MHTP always requires less time than the NSHTP, SP and CoSaMP for any given κ , whereas the time consumed by the MHTP is close to that of the HTP. By comparing Tables 3 and 4, it is found that the source of HP is different for ‘Goldhill’. Specifically, in Table 3, HP only comes from NSHTP at $\kappa = 0.3$; in Table 4, for any κ , HP all comes from NSHTP. Hence the sparsity of the same image in the discrete wavelet transform domain is different with respect to different wavelets (such as ‘sym8’ and ‘coif5’), which in turn affects the reconstruct capability of the algorithm.

Corresponding to the first row in Table 3, the performance of these algorithms in visual quality via the image “Lena” is shown in Figures 12 and 13. From Figure 12, the reconstruction quality of the MHTP is improved remarkably with an increase in the sampling rate κ in the interval $[0.3, 0.5]$. A comparison of the images reconstructed using these



FIGURE 12: The performance of the MHTP with different sampling rates. (a) Original, (b) $\kappa = 0.3$, (c) $\kappa = 0.4$, and (d) $\kappa = 0.5$.



FIGURE 13: Continued.



FIGURE 13: Comparison of the reconstructed images of algorithms with $\kappa = 0.4$. (a) MHTP, (b) NSHTP, (c) HTP, (d) gOMP, (e) SP, and (f) CoSaMP.

algorithms with $\kappa = 0.4$ is shown in Figure 13, which indicates that the reconstruction quality of all algorithms is similar, except for CoSaMP; and one possible reason for CoSaMP failure is high sparsity level. Using the above mentioned experiments, we show that the MHTP is competitive with the other four mainstream algorithms in terms of reconstruction quality and runtime.

5. Conclusion

A new algorithm for solving the sparse optimization problem, called MHTP, is proposed using the information of the current point x^p and previous point x^{p-1} to produce the new iteration point x^{p+1} . Numerical experiments show that the recovery capability of MHTP is stronger than that of several mainstream algorithms, such as HTP, NSHTP, CoSaMP, SP, and gOMP. The runtime of MHTP is close to that of HTP and is often less than that of the other four algorithms. Two interesting but challenging research topics merit further investigation: one is to study other acceleration techniques such as the Nesterov-accelerated gradient method to improve the efficiency of the algorithms. In addition, MHTP is reduced to the standard HTP because the coefficient of the convex combination takes $\lambda = 1$. Hence, another research topic is whether the efficiency of MHTP can be further improved by dynamically adjusting the coefficients of the convex combinations in the iterative process.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Li-Ping Geng: Methodology, Writing—original draft. Jin-Chuan Zhou: Investigation, Supervision. Zhong-Feng Sun:

Investigation, Visualization. Jing-Yong Tang: Methodology, Software.

Acknowledgments

The second author's work is supported by National Natural Science Foundation of China (12371305 and 11771255), Shandong Province Natural Science Foundation (ZR2023MA020 and ZR2021MA066), and Young Innovation Teams of Shandong Province (2019KJ1013). The fourth author's work is supported by Natural Science Foundation of Henan Province (222300420520) and Key Scientific Research Projects of Higher Education of Henan Province (22A110020).

References

- [1] Q. Deng, H. Zeng, J. Zhang et al., "Compressed sensing for image reconstruction via back-off and rectification of greedy algorithm," *Signal Processing*, vol. 157, pp. 280–287, 2019.
- [2] Y. Wu, J. Sun, W. Chen, and J. Yin, "Improved image compressive sensing recovery with low-rank prior and deep image prior," *Signal Processing*, vol. 205, Article ID 108896, 2023.
- [3] M. Ke, Z. Gao, Y. Wu, X. Gao, and R. Schober, "Compressive sensing-based adaptive active user detection and channel estimation: Massive access meets massive mimo," *IEEE Transactions on Signal Processing*, vol. 68, pp. 764–779, 2020.
- [4] S. Ghosh, R. Agarwal, M. A. Rehan et al., "A compressed sensing approach to pooled RT-PCR testing for COVID-19 detection," *IEEE Open Journal of Signal Processing*, vol. 2, pp. 248–264, 2021.
- [5] J.-F. Cai, J. Li, X. Lu, and J. You, "Sparse signal recovery from phaseless measurements via hard thresholding pursuit," *Applied and Computational Harmonic Analysis*, vol. 56, pp. 367–390, 2022.
- [6] L. Stanković, M. Daković, and S. Vujović, "Adaptive variable step algorithm for missing samples recovery in sparse signals," *IET Signal Processing*, vol. 8, no. 3, pp. 246–256, 2014.
- [7] L. Stanković and M. Daković, "On a gradient-based algorithm for sparse signal reconstruction in the signal/measurements domain," *Mathematical Problems in Engineering*, vol. 2016, Article ID 6212674, 11 pages, 2016.

- [8] H. Q. Cai, J.-F. Cai, T. Wang, and G. Yin, "Accelerated structured alternating projections for robust spectrally sparse signal recovery," *IEEE Transactions on Signal Processing*, vol. 69, pp. 809–821, 2021.
- [9] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [10] D. L. Donoho and Y. Tsaig, "Fast solution of ℓ_1 -norm minimization problems when the solution may be sparse," *IEEE Transactions on Information Theory*, vol. 54, no. 11, pp. 4789–4812, 2008.
- [11] H. Ge, W. Chen, and M. K. Ng, "New RIP bounds for recovery of sparse signals with partial support information via weighted ℓ_p -minimization," *IEEE Transactions on Information Theory*, vol. 66, no. 6, pp. 3914–3928, 2020.
- [12] A. Wan, "Uniform RIP conditions for recovery of sparse signals by ℓ_p ($0 < p \leq 1$) minimization," *IEEE Transactions on Signal Processing*, vol. 68, pp. 5379–5394, 2020.
- [13] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.
- [14] N. Han, S. Li, and J. Lu, "Orthogonal subspace based fast iterative thresholding algorithms for joint sparsity recovery," *IEEE Signal Processing Letters*, vol. 28, pp. 1320–1324, 2021.
- [15] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.
- [16] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [17] J. Wen, R. Zhang, and W. Yu, "Signal-dependent performance analysis of orthogonal matching pursuit for exact sparse recovery," *IEEE Transactions on Signal Processing*, vol. 68, pp. 5031–5046, 2020.
- [18] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, 2009.
- [19] M. Fornasier and H. Rauhut, "Iterative thresholding algorithms," *Applied and Computational Harmonic Analysis*, vol. 25, no. 2, pp. 187–208, 2008.
- [20] S. Foucart and G. Lecue, "An IHT algorithm for sparse recovery from subexponential measurements," *IEEE Signal Processing Letters*, vol. 24, no. 9, pp. 1280–1283, 2017.
- [21] J.-L. Bouchot, S. Foucart, and P. Hitczenko, "Hard thresholding pursuit algorithms: number of iterations," *Applied and Computational Harmonic Analysis*, vol. 41, no. 2, pp. 412–435, 2016.
- [22] S. Foucart, "Hard thresholding pursuit: An algorithm for compressive sensing," *SIAM Journal on Numerical Analysis*, vol. 49, no. 6, pp. 2543–2563, 2011.
- [23] X.-T. Yuan, P. Li, and T. Zhang, "Gradient hard thresholding pursuit," *Journal of Machine Learning Research*, vol. 18, pp. 1–43, 2018.
- [24] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Springer, New York, 2013.
- [25] Y.-B. Zhao, "Optimal k -thresholding algorithms for sparse optimization problems," *SIAM Journal on Optimization*, vol. 30, no. 1, pp. 31–55, 2020.
- [26] Y.-B. Zhao and Z.-Q. Luo, "Analysis of optimal thresholding algorithms for compressed sensing," *Signal Processing*, vol. 187, p. 108148, 2021.
- [27] Y.-B. Zhao and Z.-Q. Luo, "Natural thresholding algorithms for signal recovery with sparsity," *IEEE Open Journal of Signal Processing*, vol. 3, pp. 417–431, 2022.
- [28] H. Liu and R. F. Barber, "Between hard and soft thresholding: optimal iterative thresholding algorithms," *Information and Inference: A Journal of the IMA*, vol. 9, no. 4, pp. 899–933, 2020.
- [29] J. D. Blanchard, J. Tanner, and K. Wei, "CGIHT: conjugate gradient iterative hard thresholding for compressed sensing and matrix completion," *Information and Inference: A Journal of the IMA*, vol. 4, no. 4, pp. 289–327, 2015.
- [30] N. Meng and Y.-B. Zhao, "Newton-step-based hard thresholding algorithms for sparse signal recovery," *IEEE Transactions on Signal Processing*, vol. 68, pp. 6594–6606, 2020.
- [31] N. Meng, Y.-B. Zhao, M. Kočvara, and Z. Sun, "Partial gradient optimal thresholding algorithms for a class of sparse optimization problems," *Journal of Global Optimization*, vol. 84, pp. 393–413, 2022.
- [32] Z.-F. Sun, J.-C. Zhou, Y.-B. Zhao, and N. Meng, "Heavy-ball-based hard thresholding algorithms for sparse signal recovery," *Journal of Computational and Applied Mathematics*, vol. 430, Article ID 115264, 2023.
- [33] A. Esmaeili, E. A. Kangarshahi, and F. Marvasti, "Iterative null space projection method with adaptive thresholding in sparse signal recovery," *IET Signal Processing*, vol. 12, no. 5, pp. 605–612, 2018.
- [34] H. Li and G. Liu, "Perturbation analysis of signal space fast iterative hard thresholding with redundant dictionaries," *IET Signal Processing*, vol. 11, no. 4, pp. 462–468, 2017.
- [35] Y. Wang, W. Lu, and H. Lian, "Best subset selection for high-dimensional non-smooth models using iterative hard thresholding," *Information Sciences*, vol. 625, pp. 36–48, 2023.
- [36] J. Wang, S. Kwon, P. Li, and B. Shim, "Recovery of sparse signals via generalized orthogonal matching pursuit: a new analysis," *IEEE Transactions on Signal Processing*, vol. 64, no. 4, pp. 1076–1089, 2016.
- [37] J. D. Blanchard and J. Tanner, "Performance comparisons of greedy algorithms in compressed sensing," *Numerical Linear Algebra with Applications*, vol. 22, no. 2, pp. 254–282, 2015.