Journal of Computational and Applied Mathematics 430 (2023) 115264

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Heavy-ball-based hard thresholding algorithms for sparse signal recovery

Zhong-Feng Sun [a], Jin-Chuan Zhou [a], Yun-Bin Zhao [b,*], Nan Meng [c,d]

[a] School of Mathematics and Statistics, Shandong University of Technology, Zibo, Shandong, China
[b] Shenzhen Research Institute of Big Data, Chinese University of Hong Kong, Shenzhen, Guangdong, China
[c] School of Mathematics, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom
[d] Department of Mathematical Sciences, University of Nottingham Ningbo China, Ningbo, Zhejiang, China

ARTICLE INFO

Article history:
Received 28 August 2022
Received in revised form 4 April 2023

Keywords:
Compressed sensing
Heavy-ball method
Sparse signal recovery
Hard thresholding algorithm
Restricted isometry property
Phase transition

ABSTRACT

The hard thresholding technique plays a vital role in the development of algorithms for sparse signal recovery. By merging this technique and heavy-ball acceleration method which is a multi-step extension of the traditional gradient descent method, we propose the so-called heavy-ball-based hard thresholding (HBHT) and heavy-ball-based hard thresholding pursuit (HBHTP) algorithms for signal recovery. It turns out that the HBHT and HBHTP can successfully recover a k-sparse signal if the restricted isometry constant of the measurement matrix satisfies $\delta_{3k} < 0.618$ and $\delta_{3k} < 0.577$, respectively. The guaranteed success of HBHT and HBHTP is also shown under the conditions $\delta_{2k} < 0.356$ and $\delta_{2k} < 0.377$, respectively. Moreover, the finite convergence of HBHTP and stability of the two algorithms are also established in this paper. Simulations on random problem instances are performed to compare the performance of the proposed algorithms and several existing ones. Empirical results indicate that the HBHTP performs very comparably to a few existing algorithms and it takes less average time to achieve the signal recovery than these existing methods.

1. Introduction

In compressed sensing scenarios, one needs to recover a sparse signal $x \in \mathbb{R}^n$ from linear measurements $y := Ax + v$, where $v \in \mathbb{R}^m$ are measurement errors and A is a known $m \times n$ measurement matrix with $m \ll n$. When $v = 0$, the measurements y are accurate. To recover the signal x in such an environment, one may use the optimization model

$$\min_z \{\|y - Az\|_2^2 : \ \|z\|_0 \le k\}, \tag{1.1}$$

where k (a given integer number) is an estimate of the sparsity level of x, and $\|z\|_0$ denotes the number of nonzero entries of $z \in \mathbb{R}^n$. In this paper, a vector z is said to be k-sparse if $\|z\|_0 \le k$. It is well known that when x is k-sparse and A satisfies certain assumptions, x will be the unique k-sparse solution to the problem (1.1) (see, e.g., [1–3]). Thus the recovery of x often amounts to solving (1.1), and the algorithms for such a problem are usually called compressed sensing algorithms or, in more general, sparse optimization algorithms.

* Corresponding author.
E-mail addresses: zfsun@sdut.edu.cn (Z.-F. Sun), jinchuanzhou@sdut.edu.cn (J.-C. Zhou), yunbinzhao@cuhk.edu.cn (Y.-B. Zhao), Nan.Meng@nottingham.edu.cn (N. Meng).

https://doi.org/10.1016/j.cam.2023.115264
0377-0427/

Table 1
Abbreviations.

ASM	Algorithm selection map
CoSaMP	Compressive sampling matching pursuit
GDS	Gradient descent with sparsification
HBHT	Heavy-ball-based hard thresholding
HBHTP	Heavy-ball-based hard thresholding pursuit
HTP	Hard thresholding pursuit
IHT	Iterative hard thresholding
OMP	Orthogonal matching pursuit
PSNR	Peak signal-to-noise ratio
PTC	Phase transition curve
RIC	Restricted isometry constant
RIP	Restricted isometry property
SP	Subspace pursuit

Let us first briefly review the thresholding algorithms for sparse signal recovery. Before doing so, we summarize the abbreviations in Table 1, which are frequently used in the paper. The thresholding technique was introduced by Donoho and Johnstone [4]. At present, there are three main classes of thresholding algorithms: hard thresholding [5–13], soft thresholding [14–16], and optimal thresholding [17–19]. A huge amount of work has been carried out for the class of hard thresholding algorithms. For instance, the IHT was studied early in [5], which is a combination of the gradient descent and hard thresholding technique. The IHT admits a few modifications including the GDS [20] and the normalized IHT with a fixed or adaptive steplength [7,10]. In addition, combining IHT [5,7,20] and orthogonal projection immediately leads to the HTP in [9]. The thresholding methods combined with Nesterov's acceleration technique were also studied (e.g., [8,11,13]). Recently, it was pointed out in [17] that performing hard thresholding is usually independent of the reduction of the residual $\|y - Az\|_2^2$ and thus the so-called optimal k-thresholding operator was proposed in [17,18]. Nevertheless, the optimal k-thresholding algorithm need to solve a quadratic convex optimization problem at every iteration which requires more computational time than the traditional hard thresholding methods.

The aim of this paper is to use the heavy ball method to accelerate the hard thresholding algorithms without increasing its computational complexity. Recall that the search direction at the iterate x^p in IHT and HTP is given by $A^T(y - Ax^p)$, which is the negative gradient of the residual at x^p. With the aid of the momentum term $x^p - x^{p-1}$, the search direction can be modified to

$$d^p = \alpha A^T(y - Ax^p) + \beta(x^p - x^{p-1}) \tag{1.2}$$

with two parameters $\alpha > 0$ and $\beta \geq 0$. This is the two-step heavy-ball method proposed for optimization problems by Polyak [21]. It has been shown that a fast local convergence of this method for optimization problems can be achieved provided that the parameters are properly chosen, and that the method can work even when the Hessian matrix of the objective function is ill-posed (see, e.g., Chapter 3 in [22]). The global convergence of the heavy ball method has also been discussed in the literature [23–26]. This method is widely used in such fields as distributed optimization [24,26], variational inequality [27], wireless network [28], nonconvex optimization [29,30], deep neural network [31], and image restoration [32]. For instance, it was found in [30] that the heavy ball momentum plays an important role in driving the iterates away from the saddle points of nonconvex optimization problems; It was also used in [32] to accelerate the Richardson–Lucy algorithm in image deconvolution without causing a remarkable increase of iteration complexity; Xin and Khan [26] observed that the distributed heavy ball method achieves a global $R-$linear rate for distributed optimization, and the momentum term can dramatically improves the convergence of the algorithm for ill-conditioned objective functions. These and other applications indicate that the heavy-ball method does admit certain advantage in enhancing the efficiency of an iterative method for optimization problems and may outperform the extra-point method and Nesterov acceleration method.

Motivated by the numerical advantage of heavy ball acceleration technique, we propose the HBHT and HBHTP algorithms for the recovery problem (1.1). The guaranteed performance of the two algorithms are shown under the assumption of RIP, which was originally introduced by Candès and Tao [33] and has now become a standard tool for the analysis of various compressed sensing algorithms. It is well known that the success of IHT for k-sparse signal recovery can be guaranteed under the RIP condition $\delta_{3k} < (\sqrt{5}-1)/2 \approx 0.618$ (see [34]) and that of HTP can be guaranteed under $\delta_{3k} < 1/\sqrt{3} \approx 0.577$ (see [9]). Under the same condition and proper choice of algorithmic parameters, we establish the guaranteed-performance results for the two algorithms HBHT and HBHTP. Roughly speaking, we show that HBHT is convergent under the condition $\delta_{3k} < (\sqrt{5}-1)/2$ and that HBHTP is convergent under the condition $\delta_{3k} < 1/\sqrt{3}$. By using an analysis method in [18], we further prove that the condition for theoretical performance of the two algorithms can be established in term of δ_{2k} as well. Specifically, the guaranteed success of HBHT and HBHTP can be ensured if $\delta_{2k} < (\sqrt{5}-1)/(2\sqrt{3}) \approx 0.356$ and $\delta_{2k} < 1/\sqrt{7} \approx 0.377$, respectively. Moreover, the finite convergence of HBHTP and recovery stability of the two methods are also shown in this paper.

A large amount of experiments on random problem instances of sparse signal recovery are performed to investigate the success rate and phase transition features of the proposed algorithms. We also compare the performances of the

proposed algorithms and several existing ones such as OMP [35,36], CoSaMP [37], SP [38], IHT and HTP. The empirical results show that incorporating heavy ball technique into IHT and HTP does remarkably improve the performance of IHT and HTP, respectively. The HBHTP not only admits robust signal recovery ability in both noisy and noiseless scenarios, but also takes relatively less average computational time to achieve the recovery success compared to several existing methods.

The paper is structured as follows. In Section 2, we described the HBHT and HBHTP algorithms and list some notations and useful inequalities. The theoretical analysis of the proposed algorithms is conducted in Sections 3 and 4. Numerical results are given in Section 5, and conclusions are drawn in the last section.

2. Preliminary and algorithms

2.1. Notation

Denote by $N := \{1, 2, \ldots, n\}$. For a subset $\Omega \subseteq N$, let $\overline{\Omega} := N \setminus \Omega$ and $|\Omega|$ denote the complement set and the cardinality of Ω, respectively. Given a vector $z \in \mathbb{R}^n$, the index set $supp(z) := \{i \in N : z_i \neq 0\}$ denotes the support of z, and $z_\Omega \in \mathbb{R}^n$ is the vector with entries

$$(z_\Omega)_i = \begin{cases} z_i, & i \in \Omega, \\ 0, & i \notin \Omega. \end{cases}$$

Let $\mathcal{L}_k(z)$ be the index set of the k largest absolute entries of z, and let $\mathcal{H}_k(\cdot)$ be the hard thresholding operator which retains the k largest entries in magnitude and zeroing out other entries of a vector. The k-sparse vector $\mathcal{H}_k(z)$ is the best k-term approximation of $z \in \mathbb{R}^n$. Denote by $\sigma_k(z)_q$, where $q > 0$ is an integer number, the residual of the best k-term approximation of z, i.e.,

$$\sigma_k(z)_q = \min_u \{\|z - u\|_q : \|u\|_0 \leq k\}.$$

For any $j \in \mathbb{R}$, $\lceil j \rceil$ represents the nearest integer greater than or equal to j, and $\lfloor j \rfloor$ represents the nearest integer less than or equal to j.

2.2. Basic inequalities

We first recall the RIC and RIP of a given measurement matrix.

Definition 2.1 ([33]). Let $A \in \mathbb{R}^{m\times n}$ with $m < n$ be a matrix. The restricted isometry constant (RIC) of order k, denoted δ_k, is the smallest number $\delta \geq 0$ such that

$$(1 - \delta)\|u\|_2^2 \leq \|Au\|_2^2 \leq (1 + \delta)\|u\|_2^2 \tag{2.1}$$

for all k-sparse vectors $u \in \mathbb{R}^n$(i.e., $\|u\|_0 \leq k$). If $\delta_k < 1$, then A is said to satisfy the restricted isometry property (RIP) of order k.

From the definition above, one can see that $\delta_t \leq \delta_s$ for any integer number $t \leq s$. The following properties of RIC have been frequently used in the analysis of compressed sensing algorithms.

Lemma 2.1 ([9,17]). *Let $u \in \mathbb{R}^m$, $v \in \mathbb{R}^n$ be two vectors, $t \in N$ be a positive integer number and $W \subseteq N$ be an index set.*

(i) If $|W \cup supp(v)| \leq t$, then $\left\|\left((I - A^TA)v\right)_W\right\|_2 \leq \delta_t\|v\|_2$.
(ii) If $|W| \leq t$, then $\left\|\left(A^Tu\right)_W\right\|_2 \leq \sqrt{1 + \delta_t}\|u\|_2$.

The next lemma is taken directly from [34], and it can also be implied from the result in [39].

Lemma 2.2 ([34]). *For any vector $z \in \mathbb{R}^n$ and for any k-sparse vector $x \in \mathbb{R}^n$, one has*

$$\|x - \mathcal{H}_k(z)\|_2 \leq \eta \left\|(x - z)_{W\cup W^*}\right\|_2, \tag{2.2}$$

where $\eta = (\sqrt{5} + 1)/2$, $W = supp(x)$ and $W^* = supp(\mathcal{H}_k(z))$.

2.3. Algorithm

We now describe the algorithms in this paper. The basic idea is to use the search direction d^p given by (1.2), resulted from the heavy-ball acceleration method, to generate a point $u^p = x^p + d^p$ where x^p denotes the current iterate of the algorithm. Then, perform a hard thresholding on u^p to produce the next iterate. This idea leads to Algorithm 1 for the recovery problem (1.1).

We may treat the point $\mathcal{H}_k(u^p)$ in HBHT as an intermediate point and perform a pursuit step (i.e., orthogonal projection) to generate the next iterate x^{p+1}. This leads to Algorithm 2 called HBHTP.

Algorithm 1 Heavy-Ball-Based Hard Thresholding (HBHT)

Input (A, y, k) and two parameters $\alpha > 0$ and $\beta \geq 0$ and two initial points x^0 and x^1.

S1. At x^p, set

$$u^p = x^p + \alpha A^T(y - Ax^p) + \beta(x^p - x^{p-1}). \tag{2.3}$$

[Note: The term $\beta(x^p - x^{p-1})$ is called the momentum term in heavy ball method.]

S2. Let

$$x^{p+1} = \mathcal{H}_k(u^p).$$

Repeat the above steps until a certain stopping criterion is satisfied.

Algorithm 2 Heavy-Ball-Based Hard Thresholding Pursuit (HBHTP)

Input (A, y, k) and two parameters $\alpha > 0$ and $\beta \geq 0$ and two initial points x^0 and x^1.

S1. At x^p, set

$$u^p = x^p + \alpha A^T(y - Ax^p) + \beta(x^p - x^{p-1}).$$

S2. Let $S^{p+1} = \mathcal{L}_k(u^p)$, and

$$x^{p+1} = \arg\min_{z\in\mathbb{R}^n}\{\|y - Az\|_2^2 : supp(z) \subseteq S^{p+1}\}. \tag{2.4}$$

Repeat the above steps until a certain stopping criterion is satisfied.

Clearly, HBHT and HBHTP reduce to IHT and HTP, respectively, when $\alpha = 1$ and $\beta = 0$. The initial points x^0 and x^1 can be any vectors. The simplest choice is $x^1 = x^0 = 0$. To stop the algorithms, one can set the maximum number of iterations or use other stopping criteria such as $\|y - Ax^p\|_2 \leq \varepsilon$, where $\varepsilon > 0$ is a small tolerance. For instance, if the measurements y are accurate enough, then the measurement error $\|\nu\|_2 = \|y - Ax\|_2$ would be very small. In such a case, it makes sense to use the stopping criterion $\|y - Ax^p\|_2 \leq \varepsilon$. It is also worth mentioning that for simplicity, we treat α and β as fixed parameters throughout the paper. However, it should be pointed out that these parameters can be updated from step to step in order to get a better performance of the algorithms from both theoretical and practical viewpoints. This might be an interesting future work.

3. Convergence analysis

In this section, we analyze the performance of HBHT and HBHTP under the RIP of order $3k$ and $2k$, respectively. We also discussed the finite convergence of HBHTP under some conditions. Since the heavy-ball method is a two-step method in the sense that the next iterate x^{p+1} is generated based on the previous two iterates x^p and x^{p-1}, the analysis of HBHT and HBHTP is remarkably different from the traditional IHT and HTP. To this need, we first establish the following useful lemma.

Lemma 3.1. *Suppose that the nonnegative sequence* $\{a^p\} \subseteq \mathbb{R}$ $(p = 0, 1, \dots)$ *satisfies*

$$a^{p+1} \leq b_1 a^p + b_2 a^{p-1} + b_3, \quad p \geq 1, \tag{3.1}$$

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

$$a^p \leq \theta^{p-1}\left[a^1 + (\theta - b_1)a^0\right] + \frac{b_3}{1-\theta}, \tag{3.2}$$

with

$$0 \leq \theta := \frac{b_1 + \sqrt{b_1^2 + 4b_2}}{2} < 1.$$

Proof. Denote by $q_1 := \frac{-b_1 + \sqrt{b_1^2+4b_2}}{2}$. Note that $b_1, b_2 \geq 0$ and $b_1 + b_2 < 1$. It is straightforward to verify that $q_1 \geq 0$, $(b_1 + q_1)q_1 = b_2$ and

$$0 \leq \theta = \frac{b_1 + \sqrt{b_1^2 + 4b_2}}{2} = b_1 + q_1 < 1,$$

where $\theta < 1$ follows from the condition $b_1 + b_2 < 1$. Thus it follows from (3.1) that

$$a^{p+1} + q_1 a^p \le (b_1 + q_1)a^p + b_2 a^{p-1} + b_3 = \theta(a^p + q_1 a^{p-1}) + b_3,$$

which implies

$$\begin{aligned} a^{p+1} \le a^{p+1} + q_1 a^p \le & \theta^p (a^1 + q_1 a^0) + b_3(1 + \theta + \cdots + \theta^{p-1}) \\ \le & \theta^p \left[a^1 + (\theta - b_1)a^0\right] + \frac{b_3}{1-\theta}. \end{aligned}$$

Thus the relation (3.2) holds. ■

3.1. Guaranteed performance under RIP of order 3k

Denote by $\eta = (\sqrt{5}+1)/2$ throughout the remaining of this paper. We now prove the guaranteed performance of the proposed algorithms for signal recovery under some assumptions. We first consider the HBHT, to which the main result is stated as follows.

Theorem 3.1. *Suppose that the RIC, δ_{3k}, of the measurement matrix A and the parameters α and β obey the bounds*

$$\delta_{3k} < \frac{\sqrt{5}-1}{2} \approx 0.618, \quad 0 \le \beta < \frac{\eta}{1+\delta_{3k}} - 1, \quad \frac{2(1+\beta)-\eta}{1-\delta_{3k}} < \alpha < \frac{\eta}{1+\delta_{3k}}. \tag{3.3}$$

Let $y := Ax + \nu$ be the measurements of x with measurement errors ν. Then the iterates $\{x^p\}$ generated by HBHT satisfies

$$\|x_S - x^p\|_2 \le C_1 \tau^{p-1} + C_2 \|\nu'\|_2, \tag{3.4}$$

where $S = \mathcal{L}_k(x)$, $\nu' = \nu + Ax_{\bar{S}}$, and C_1, C_2 are the quantities given as

$$C_1 = \|x_S - x^1\|_2 + (\tau - b)\|x_S - x^0\|_2, \quad C_2 = \frac{\eta\alpha}{1-\tau}\sqrt{1+\delta_{2k}} \tag{3.5}$$

with $\tau := \frac{b+\sqrt{b^2+4\eta\beta}}{2}$. The fact $\tau < 1$ is ensured under (3.3) and together with

$$b = \eta(|1 - \alpha + \beta| + \alpha\delta_{3k}). \tag{3.6}$$

Proof. By (2.3), we have

$$u^p - x_S = (1 - \alpha + \beta)(x^p - x_S) + \alpha(I - A^T A)(x^p - x_S) - \beta(x^{p-1} - x_S) + \alpha A^T \nu', \tag{3.7}$$

where $\nu' = \nu + Ax_{\bar{S}}$. Denote $V^p := supp(\mathcal{H}_k(u^p))$. By using Lemma 2.2 and (3.7), we obtain

$$\begin{aligned} \|x^{p+1} - x_S\|_2 = & \|\mathcal{H}_k(u^p) - x_S\|_2 \\ \le & \eta \|(u^p - x_S)_{S\cup V^p}\|_2 \\ \le & \eta|1 - \alpha + \beta| \cdot \|x^p - x_S\|_2 + \eta\alpha \|[(I - A^T A)(x^p - x_S)]_{S\cup V^p}\|_2 \\ & + \eta\beta\|x^{p-1} - x_S\|_2 + \eta\alpha\|(A^T\nu')_{S\cup V^p}\|_2. \end{aligned} \tag{3.8}$$

Since $|S \cup V^p| \le 2k$ and $|supp(x^p - x_S) \cup S \cup V^p| \le 3k$, by using Lemma 2.1, we obtain

$$\|[(I - A^T A)(x^p - x_S)]_{S\cup V^p}\|_2 \le \delta_{3k}\|x^p - x_S\|_2 \tag{3.9}$$

and

$$\|(A^T\nu')_{S\cup V^p}\|_2 \le \sqrt{1+\delta_{2k}}\|\nu'\|_2. \tag{3.10}$$

Substituting (3.9) and (3.10) into (3.8) yields

$$\|x^{p+1} - x_S\|_2 \le b\|x^p - x_S\|_2 + \eta\beta\|x^{p-1} - x_S\|_2 + \eta\alpha\sqrt{1+\delta_{2k}}\|\nu'\|_2, \tag{3.11}$$

where b is given by (3.6). The recursive inequality (3.11) is of the form (3.1) in Lemma 3.1. We now point out that the coefficients of the right-hand side of (3.11) satisfy the condition of Lemma 3.1. In fact, suppose that $\delta_{3k} < (\sqrt{5}-1)/2 = \eta - 1$, which implies that $0 < \frac{\eta}{1+\delta_{3k}} - 1$. Thus the range for β in (3.3) is well defined, and hence $\frac{2(1+\beta)-\eta}{1-\delta_{3k}} < 1 + \beta < \frac{\eta}{1+\delta_{3k}}$. This implies that the range for α in (3.3) is also well defined. Merging (3.6) and (3.3) leads to

$$\begin{aligned} b = & \eta(|1 - \alpha + \beta| + \alpha\delta_{3k}) \\ = & \begin{cases} \eta[1 + \beta - \alpha(1 - \delta_{3k})], & \text{if } \frac{2(1+\beta)-\eta}{1-\delta_{3k}} < \alpha \le 1 + \beta, \\ \eta[-1 - \beta + \alpha(1 + \delta_{3k})], & \text{if } 1 + \beta < \alpha < \frac{\eta}{1+\delta_{3k}}, \end{cases} \end{aligned}$$

$$
\begin{aligned}
&< \begin{cases} \eta\left[1+\beta-\left(\dfrac{2(1+\beta)-\eta}{1-\delta_{3k}}\right)(1-\delta_{3k})\right], & \text{if } \frac{2(1+\beta)-\eta}{1-\delta_{3k}} < \alpha \le 1+\beta, \\ \eta\left[-1-\beta+\left(\dfrac{\eta}{1+\delta_{3k}}\right)(1+\delta_{3k})\right], & \text{if } 1+\beta < \alpha < \frac{\eta}{1+\delta_{3k}}, \end{cases} \\
&= \eta(\eta-1-\beta) \\
&= 1-\eta\beta,
\end{aligned}
$$

where the last equality follows from the fact that η is the root of the equation $t^2 - t = 1$. The above inequality means $b + \eta\beta < 1$ and hence the recursive formula (3.11) satisfies the condition of Lemma 3.1. Therefore, it follows from Lemma 3.1 that

$$\tau = \frac{b+\sqrt{b^2+4\eta\beta}}{2} < 1$$

and the bound (3.4) holds, where C_1, C_2 are given by (3.5). ■

If the signal x is k-sparse and the measurements are accurate, in which case $x_S = x$ and $\nu = 0$, then the above result implies that

$$\|x - x^p\|_2 \le \tau^{p-1} C_1 \to 0 \text{ as } p \to \infty,$$

which implies that the iterates generated by HBHT converges to the sparse signal.

We now establish the main performance result for HBHTP. We first recall a helpful lemma.

Lemma 3.2 ([9]). *Given the measurements $y := Ax + \nu$ of x and the index set S^{p+1}, the iterate x^{p+1} generated by the pursuit step (2.4) obeys*

$$\left\|x^{p+1} - x_S\right\|_2 \le \frac{1}{\sqrt{1-(\delta_{2k})^2}} \left\|(x^{p+1} - x_S)_{\overline{S^{p+1}}}\right\|_2 + \frac{\sqrt{1+\delta_k}}{1-\delta_{2k}} \left\|\nu'\right\|_2, \tag{3.12}$$

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

The main result concerning the guaranteed success of HBHTP is stated as follows.

Theorem 3.2. *Suppose that the RIC, δ_{3k}, of the matrix A and the parameters α and β obey*

$$\delta_{3k} < \frac{1}{\sqrt{3}} \approx 0.577, \quad 0 \le \beta < \frac{\frac{1}{\hat{\eta}}+1}{1+\delta_{3k}} - 1, \quad \frac{1+2\beta-\frac{1}{\hat{\eta}}}{1-\delta_{3k}} < \alpha < \frac{\frac{1}{\hat{\eta}}+1}{1+\delta_{3k}}, \tag{3.13}$$

where $\hat{\eta} = \frac{\sqrt{2}}{\sqrt{1-(\delta_{2k})^2}}$. Let $y := Ax + \nu$ be the measurements of x with errors ν. Then the iterates $\{x^p\}$ generated by HBHTP satisfies

$$\left\|x_S - x^p\right\|_2 \le C_3 \hat{\tau}^{p-1} + C_4 \left\|\nu'\right\|_2, \tag{3.14}$$

where $S = \mathcal{L}_k(x)$, $\nu' = \nu + Ax_{\overline{S}}$, and C_3, C_4 are given as

$$C_3 = \|x_S - x^1\|_2 + (\hat{\tau} - \hat{b})\|x_S - x^0\|_2, \quad C_4 = \frac{1}{1-\hat{\tau}}\left(\hat{\eta}\alpha\sqrt{1+\delta_{2k}} + \frac{\sqrt{1+\delta_k}}{1-\delta_{2k}}\right) \tag{3.15}$$

with constants $\hat{b}$, $\hat{\tau}$ being given by

$$\hat{b} = \hat{\eta}(|1-\alpha+\beta| + \alpha\delta_{3k}), \quad \hat{\tau} = \frac{\hat{b}+\sqrt{\hat{b}^2+4\hat{\eta}\beta}}{2} \tag{3.16}$$

and $\hat{\tau} < 1$ is guaranteed under the condition (3.13).

Proof. Since $S^{p+1} = \mathcal{L}_k(u^p)$ in HBHTP and $S = \mathcal{L}_k(x)$, we have

$$\|(u^p)_{S^{p+1}}\|_2^2 \ge \|(u^p)_S\|_2^2.$$

Eliminating the entries indexed by $S \cap S^{p+1}$ from the above inequality and taking square root yields

$$\|(u^p)_{S^{p+1}\setminus S}\|_2 \ge \|(u^p)_{S\setminus S^{p+1}}\|_2.$$

Note that $(x_S)_{S^{p+1}\setminus S}=0$ and $(x^{p+1})_{S\setminus S^{p+1}}=0$. From the inequality above, we have

$$\begin{aligned}\|(u^p-x_S)_{S^{p+1}\setminus S}\|_2 \geq& \|(x_S-x^{p+1}+u^p-x_S)_{S\setminus S^{p+1}}\|_2\\ \geq& \|(x_S-x^{p+1})_{\overline{S^{p+1}}}\|_2-\|(u^p-x_S)_{S\setminus S^{p+1}}\|_2,\end{aligned}$$

where the second inequality follows from the triangular inequality and the fact $(x_S-x^{p+1})_{S\setminus S^{p+1}}=(x_S-x^{p+1})_{\overline{S^{p+1}}}$. It follows that

$$\begin{aligned}\|(x_S-x^{p+1})_{\overline{S^{p+1}}}\|_2 \leq& \|(u^p-x_S)_{S\setminus S^{p+1}}\|_2+\|(u^p-x_S)_{S^{p+1}\setminus S}\|_2\\ \leq& \sqrt{2\left(\|(u^p-x_S)_{S\setminus S^{p+1}}\|_2^2+\|(u^p-x_S)_{S^{p+1}\setminus S}\|_2^2\right)}\\ =& \sqrt{2}\|(u^p-x_S)_{S^{p+1}\triangle S}\|_2,\end{aligned}\tag{3.17}$$

where $S^{p+1}\triangle S:=(S^{p+1}\setminus S)\cup(S\setminus S^{p+1})$ is the symmetric difference of S^{p+1} and S. The last equality above follows from $(S^{p+1}\setminus S)\cap(S\setminus S^{p+1})=\emptyset$. Note that (3.7) remains valid for HBHTP. Merging (3.7) and (3.17) leads to

$$\begin{aligned}\|(x_S-x^{p+1})_{\overline{S^{p+1}}}\|_2 \leq& \sqrt{2}\{|1-\alpha+\beta|\cdot\|(x^p-x_S)_{S^{p+1}\triangle S}\|_2\\ &+\alpha\|(A^T\nu')_{S^{p+1}\triangle S}\|_2+\alpha\|[(I-A^TA)(x^p-x_S)]_{S^{p+1}\triangle S}\|_2\\ &+\beta\|(x^{p-1}-x_S)_{S^{p+1}\triangle S}\|_2\}.\end{aligned}\tag{3.18}$$

Since $|S^{p+1}\triangle S|\leq 2k$ and $|(S^{p+1}\triangle S)\cup supp(x^p-x_S)|\leq 3k$, by using Lemma 2.1, one has

$$\|[(I-A^TA)(x^p-x_S)]_{S^{p+1}\triangle S}\|_2\leq\delta_{3k}\left\|x^p-x_S\right\|_2\tag{3.19}$$

and

$$\|(A^T\nu')_{S^{p+1}\triangle S}\|_2\leq\sqrt{1+\delta_{2k}}\|\nu'\|_2.\tag{3.20}$$

Combining (3.18)–(3.20) leads to

$$\begin{aligned}\|(x_S-x^{p+1})_{\overline{S^{p+1}}}\|_2 \leq& \sqrt{2}\{(|1-\alpha+\beta|+\alpha\delta_{3k})\|x^p-x_S\|_2+\alpha\sqrt{1+\delta_{2k}}\|\nu'\|_2\\ &+\beta\|x^{p-1}-x_S\|_2\}.\end{aligned}$$

Merging the inequality above and (3.12) in Lemma 3.2, we obtain

$$\|x^{p+1}-x_S\|_2\leq\hat{b}\|x^p-x_S\|_2+\hat{\eta}\beta\|x^{p-1}-x_S\|_2+(1-\hat{\tau})C_4\|\nu'\|_2,\tag{3.21}$$

where $\hat{\eta},\hat{b},\hat{\tau},C_4$ are given exactly as in Theorem 3.2.

Since $\delta_{2k}\leq\delta_{3k}<\frac{1}{\sqrt{3}}$, we have $\sqrt{2}\delta_{3k}<\sqrt{1-(\delta_{3k})^2}\leq\sqrt{1-(\delta_{2k})^2}=\sqrt{2}/\hat{\eta}$, which implies that $0<\frac{\frac{1}{\hat{\eta}}+1}{1+\delta_{3k}}-1$. Therefore, the range for β in (3.13) is well defined, which also implies that

$$\frac{1+2\beta-\frac{1}{\hat{\eta}}}{1-\delta_{3k}}<1+\beta<\frac{\frac{1}{\hat{\eta}}+1}{1+\delta_{3k}},$$

and hence the range for α in (3.13) is also well defined. Thus it follows from (3.13) and (3.16) that

$$\begin{aligned}\hat{b}=&\hat{\eta}(|1-\alpha+\beta|+\alpha\delta_{3k})\\ =&\begin{cases}\hat{\eta}[1+\beta-\alpha(1-\delta_{3k})], & \frac{1+2\beta-\frac{1}{\hat{\eta}}}{1-\delta_{3k}}<\alpha\leq 1+\beta,\\ \hat{\eta}[-1-\beta+\alpha(1+\delta_{3k})], & 1+\beta<\alpha<\frac{\frac{1}{\hat{\eta}}+1}{1+\delta_{3k}},\end{cases}\\ <&1-\hat{\eta}\beta,\end{aligned}$$

i.e., $b+\hat{\eta}\beta<1$. Therefore, applying Lemma 3.1 to the recursive relation (3.21), we immediately conclude that $\hat{\tau}<1$ and the desired estimation (3.14) holds. ■

Remark 3.1. To our knowledge, the best known RIP-based bounds in terms of $3k$ for the IHT-type and HTP-type algorithms are $\delta_{3k}<\frac{\sqrt{5}-1}{2}$ and $\delta_{3k}<\frac{1}{\sqrt{3}}$, respectively. However, whether these bounds for the two algorithms are optimal or not are still not clear at present, although Zhao and Luo [34] conjectured that $\delta_{3k}<\frac{\sqrt{5}-1}{2}$ is optimal for IHT based on the tightness of (2.2). Similarly, it is not clear whether the bounds given in the main results of this section are optimal or not for the heavy-ball-based algorithms.

When the measurements are accurate and the signal is k-sparse, Theorem 3.2 implies that the sequence $\{x^p\}$ produced by the HBHTP must converge to the signal as $p\to\infty$. That is, the algorithm exactly recovers the signal in this case. It is also worth pointing out that computing the RIC of a matrix is generally difficult. Thus in practical applications, we do not

require that the parameters α and β be chosen to strictly meet the condition (3.3) or (3.13). These parameters can be set to roughly satisfy these conditions, for instance,

$$0 \le \beta < \eta - 1, \quad 2 + 2\beta - \eta < \alpha < \eta, \tag{3.22}$$

where $\eta = (\sqrt{5}+1)/2$. As examples, we may simply set $\alpha \in [0.4 + 2\beta, 1.6]$ and $\beta \in (0, 0.6]$ in HBHT for simplicity, and set $\alpha \in [0.3 + 2\beta, 1.7]$ and $\beta \in (0, 0.7]$ in HBHTP.

3.2. Guaranteed performance under RIP of order $2k$

Motivated by the idea of decomposition method in [18], we first establish a helpful inequality, based on which the guaranteed performance of the proposed algorithms can be characterized immediately in terms of RIP of order $2k$.

Lemma 3.3. *Let x and z be two k-sparse vectors, $S = supp(x)$ and $S^* \subseteq N$ be an index set. If $|S \cup S^*| \le 2k$, then*

$$\|[(I - A^TA)(x - z)]_{S\cup S^*}\|_2 \le \sqrt{3}\delta_{2k}\|x - z\|_2. \tag{3.23}$$

Proof. In this proof, we denote by $\mathbf{e} = (1, 1, \ldots, 1)^T$ the n-dimensional vector of ones and we use the symbol $u \otimes v := (u_1v_1, \ldots, u_nv_n)^T$ to denote the Hadamard product of two vectors $u, v \in \mathbb{R}^n$. Let x, z, S, S^* be specified as in this lemma. Let $\hat{\omega} \in \{0, 1\}^n$ be a $2k$-sparse binary vector such that $S \cup S^* \subseteq supp(\hat{\omega})$. We partition $\hat{\omega}$ into two k-sparse binary vectors ω' and ω'', i.e., $\hat{\omega} = \omega' + \omega''$, where $supp(\omega') \cap supp(\omega'') = \emptyset$. The following relation holds for any $u \in \mathbb{R}^n$:

$$\left\|u \otimes \hat{\omega}\right\|_2^2 = \left\|u \otimes \omega'\right\|_2^2 + \left\|u \otimes \omega''\right\|_2^2. \tag{3.24}$$

Note that $x - z$ can be decomposed into two sparse vectors $v^{(1)}$ and $v^{(2)}$, i.e., $x - z = v^{(1)} + v^{(2)}$, where $v^{(1)} = (x - z) \otimes \hat{\omega}$ is a $2k$-sparse vector and $v^{(2)} = (x - z) \otimes (\mathbf{e} - \hat{\omega})$ is a k-sparse vector since $S \subseteq supp(\hat{\omega})$ and z is k-sparse. It is easy to see that

$$\begin{aligned}
\|[(I - A^TA)(x - z)]_{S\cup S^*}\|_2 &\le \|[(I - A^TA)(x - z)]_{supp(\hat{\omega})}\|_2 \\
&= \|[(I - A^TA)(v^{(1)} + v^{(2)})] \otimes \hat{\omega}\|_2 \\
&\le \|[(I - A^TA)v^{(1)}] \otimes \hat{\omega}\|_2 + \|[(I - A^TA)v^{(2)}] \otimes \hat{\omega}\|_2.
\end{aligned} \tag{3.25}$$

Since $supp(v^{(1)}) \subseteq supp(\hat{\omega})$, we have $|supp(v^{(1)}) \cup supp(\hat{\omega})| \le 2k$. It follows from Lemma 2.1(i) that

$$\|[(I - A^TA)v^{(1)}] \otimes \hat{\omega}\|_2 = \|[(I - A^TA)v^{(1)}]_{supp(\hat{\omega})}\|_2 \le \delta_{2k}\|v^{(1)}\|_2. \tag{3.26}$$

Since $|supp(v^{(2)}) \cup supp(\omega')| \le 2k$ and $|supp(v^{(2)}) \cup supp(\omega'')| \le 2k$, by using (3.24) and Lemma 2.1(i), we obtain

$$\begin{aligned}
\|[(I - A^TA)v^{(2)}] \otimes \hat{\omega}\|_2^2 &= \|[(I - A^TA)v^{(2)}] \otimes \omega'\|_2^2 + \|[(I - A^TA)v^{(2)}] \otimes \omega''\|_2^2 \\
&\le 2(\delta_{2k})^2\|v^{(2)}\|_2^2,
\end{aligned}$$

i.e.,

$$\|[(I - A^TA)v^{(2)}] \otimes \hat{\omega}\|_2 \le \sqrt{2}\delta_{2k}\|v^{(2)}\|_2. \tag{3.27}$$

Combining (3.25), (3.26) and (3.27) yields

$$\begin{aligned}
\|[(I - A^TA)(x - z)]_{S\cup S^*}\|_2 &\le \delta_{2k}\left(\|v^{(1)}\|_2 + \sqrt{2}\|v^{(2)}\|_2\right) \\
&\le \sqrt{3}\delta_{2k}\sqrt{\|v^{(1)}\|_2^2 + \|v^{(2)}\|_2^2} = \sqrt{3}\delta_{2k}\|x - z\|_2,
\end{aligned}$$

where the second inequality follows from the fact $a + \sqrt{2}c \le \sqrt{3(a^2 + c^2)}$ for any $a, c \ge 0$. ■

According to Lemma 3.3, the term δ_{3k} in bounds (3.9) and (3.19) can be replaced with $\sqrt{3}\delta_{2k}$. Thus we immediately obtain the theoretical performance results for the proposed algorithms in terms of RIP of order $2k$.

Corollary 3.1. *Let $y := Ax + \nu$ be the inaccurate measurements of x. If the RIC, δ_{2k}, of the matrix A and the parameters α and β in HBHT satisfy the conditions:*

$$\delta_{2k} < \frac{\sqrt{5}-1}{2\sqrt{3}} \approx 0.356, \quad 0 \le \beta < \frac{\eta}{1 + \sqrt{3}\delta_{2k}} - 1, \quad \frac{2(1+\beta) - \eta}{1 - \sqrt{3}\delta_{2k}} < \alpha < \frac{\eta}{1 + \sqrt{3}\delta_{2k}},$$

then the conclusion of Theorem 3.1 remains valid, with constants τ, C_1, C_2 being defined the same way therein except $b = \eta(|1 - \alpha + \beta| + \alpha\sqrt{3}\delta_{2k})$.

Table 2
RIP-based bounds.

Algorithms	IHT [34]	IHT$^{\mu}$/GDS [2,20]	HBHT	HTP [9]	HBHTP
$\delta_{3k} < \delta_*$	0.618		0.618	0.577	0.577
$\delta_{2k} < \delta_*$		0.333	0.356		0.377

Corollary 3.2. *Let $y := Ax + v$ be the inaccurate measurements of x. If the RIC, δ_{2k}, of the matrix A and the parameters α and β in HBHTP satisfy the conditions:*

$$\delta_{2k} < \frac{1}{\sqrt{7}} \approx 0.377, \quad 0 \le \beta < \frac{\frac{1}{\hat{\eta}} + 1}{1 + \sqrt{3}\delta_{2k}} - 1, \quad \frac{1 + 2\beta - \frac{1}{\hat{\eta}}}{1 - \sqrt{3}\delta_{2k}} < \alpha < \frac{\frac{1}{\hat{\eta}} + 1}{1 + \sqrt{3}\delta_{2k}},$$

then the conclusion of Theorem 3.2 remains valid, with constants $\hat{\tau}, C_3, C_4$ being defined the same way therein except $\hat{b} = \hat{\eta}(|1 - \alpha + \beta| + \alpha\sqrt{3}\delta_{2k})$.

Remark 3.2. According to Proposition 6.6 in [2], one has the relation $\delta_{3k} \le 3\delta_{2k}$. If we use this relation to derive an upper bound for the left-hand side of (3.23), then the resulting bound would be too loose. The bound (3.23) established here is much tighter, and thus it leads to a desired strong result. We summarize the best known conditions for guaranteed performance of several compressed sensing algorithms in Table 2. Our analysis indicates that the RIP-based bounds for the performance guarantee of HBHT and HBHTP can be the same as the best known bounds for IHT and HTP, respectively. Similar to the sufficient condition $\delta_{2k} < 0.333$ for the performance guarantee of IHT$^{\mu}$ in [2,20], the more relaxed condition $\delta_{2k} < 0.356$ is obtained for the algorithm HBHT in this paper. It is also interesting to observe that the sufficient condition $\delta_{2k} < 0.377$ for HBHTP is less restrictive than that of HBHT, while the conditions in terms of δ_{3k} for the two algorithms go other way round.

3.3. Finite convergence

From accurate measurements, the HBHTP can exactly recover a k-sparse signal in a finite number of iterations. The iteration complexity is given in the next result.

Theorem 3.3. *Suppose that the RIC, δ_{3k}, of the measurement matrix A and the algorithmic parameters α and β in HBHTP satisfy the condition (3.13). Then any k-sparse signal x with $\|x\|_0 = k$ can be exactly recovered by HBHTP from accurate measurements $y := Ax$ in at most*

$$p^* = \left\lceil \frac{\log\left(\frac{\sqrt{2}C_3}{\hat{\eta}\mu}\right)}{\log\left(1/\hat{\tau}\right)} \right\rceil \tag{3.28}$$

iterations, where $\mu = \min_{x_i \neq 0} |x_i|$, and $\hat{\eta}, \hat{\tau}, C_3$ are given in Theorem 3.2.

Proof. Denote by $S = \mathcal{L}_k(x)$. For any $t \in \overline{S}$, by using the definition of u^p in HBHTP, which is defined as (2.3), and noting that $x_t = 0$, we have

$$\begin{aligned} |(u^p)_t| &= \left|x_t + (1 - \alpha + \beta)(x^p - x)_t + \alpha\left[(I - A^TA)(x^p - x)\right]_t - \beta(x^{p-1} - x)_t\right| \\ &\le |1 - \alpha + \beta| \cdot |(x^p - x)_t| + \alpha\left|\left[(I - A^TA)(x^p - x)\right]_t\right| + \beta|(x^{p-1} - x)_t|, \end{aligned}$$

and for any $s \in S$, we have

$$\begin{aligned} |(u^p)_s| = &\left|x_s + (1 - \alpha + \beta)(x^p - x)_s + \alpha\left[(I - A^TA)(x^p - x)\right]_s - \beta(x^{p-1} - x)_s\right| \\ \ge &\ \mu - |1 - \alpha + \beta| \cdot |(x^p - x)_s| - \alpha\left|\left[(I - A^TA)(x^p - x)\right]_s\right| - \beta|(x^{p-1} - x)_s|, \end{aligned}$$

where $\mu = \min_{x_i \neq 0} |x_i|$. Combining the above two inequalities leads to

$$\begin{aligned} |(u^p)_t| - |(u^p)_s| + \mu \le &|1 - \alpha + \beta| \cdot [|(x^p - x)_t| + |(x^p - x)_s|] + \beta[|(x^{p-1} - x)_t| + |(x^{p-1} - x)_s|] \\ &+ \alpha\{|[(I - A^TA)(x^p - x)]_t| + |[(I - A^TA)(x^p - x)]_s|\} \\ \le &\sqrt{2}(|1 - \alpha + \beta| \cdot \|(x^p - x)_{\{s,t\}}\|_2 + \alpha\|[(I - A^TA)(x^p - x)]_{\{s,t\}}\|_2 \\ &+ \beta\|(x^{p-1} - x)_{\{s,t\}}\|_2). \end{aligned}$$

Since $s \in S = supp(x)$, we have $|supp(x^p - x) \cup \{s, t\}| \leq 2k + 1 \leq 3k$. Using (3.16) and Lemma 2.1(i), we obtain

$$\begin{aligned} |(u^p)_t| - |(u^p)_s| + \mu \leq & \sqrt{2}\left((|1-\alpha+\beta| + \alpha\delta_{3k})\|x^p - x\|_2 + \beta\|x^{p-1} - x\|_2\right) \\ = & \frac{\sqrt{2}}{\hat{\eta}}\left(\hat{b}\|x^p - x\|_2 + \hat{\tau}(\hat{\tau} - \hat{b})\|x^{p-1} - x\|_2\right) \\ \leq & \frac{\sqrt{2}}{\hat{\eta}}\hat{\tau}\left(\|x^p - x\|_2 + (\hat{\tau} - \hat{b})\|x^{p-1} - x\|_2\right), \end{aligned} \tag{3.29}$$

where $\hat{\eta}, \hat{\tau}, \hat{b}$ are given in Theorem 3.2, and the last inequality above follows from the fact $\hat{b} < \hat{\tau}$. Since x is a k-sparse vector and $\nu = 0$, then $\nu' = \nu + Ax_{\overline{S}} = 0$. Hence, (3.21) becomes

$$\left\|x^{p+1} - x\right\|_2 \leq \hat{b}\left\|x^p - x\right\|_2 + \hat{\eta}\beta\left\|x^{p-1} - x\right\|_2.$$

With the aid of (3.16) and note that $\hat{\tau}(\hat{\tau} - \hat{b}) = \hat{\eta}\hat{\beta}$, the inequality above can be further rewritten as

$$\|x^{p+1} - x\|_2 + (\hat{\tau} - \hat{b})\|x^p - x\|_2 \leq \hat{\tau}(\|x^p - x\|_2 + (\hat{\tau} - \hat{b})\|x^{p-1} - x\|_2).$$

Thus (3.29) reduces to

$$|(u^p)_t| - |(u^p)_s| + \mu \leq \frac{\sqrt{2}}{\hat{\eta}}C_3(\hat{\tau})^p,$$

where C_3 is given by (3.15). After p^* iterations, where p^* is given by (3.28), one must have that

$$|(u^{p^*})_t| - |(u^{p^*})_s| + \mu \leq \frac{\sqrt{2}}{\hat{\eta}}C_3(\hat{\tau})^{p^*} < \mu,$$

where the second inequality holds due to the definition of p^* in (3.28). It implies that $|(u^{p^*})_t| < |(u^{p^*})_s|$ for any $s \in S$ and $t \in \overline{S}$. This means $S = \mathcal{L}_k(u^{p^*})$. Note that $S^{p^*+1} = \mathcal{L}_k(u^{p^*})$ at the p^*-th iteration of HBHTP. Thus at the p^*-th iteration, one has $S = S^{p^*+1}$. Under the RIP condition which implies that any k columns of A are linearly independent, the system $y = Az$ has at most one k-sparse solution. Therefore, $x^{p^*+1} = x$, i.e., the HBHTP successfully recover the k-sparse signal x after finite number of iterations. ■

4. Stability analysis

An efficient compressed sensing algorithm should be able to recover signals in a stable manner in the sense that when the problem data (e.g., signal, measurement, noise level) admits a slight change, the quality of signal recovery can still be guaranteed and the recovery error is still under control. In this section, we establish a stability result for HBHT and HBHTP, respectively. Recall that for given two integer numbers s and q, the symbol $\sigma_s(x)_q$ denotes the error (in terms of ℓ_q-norm) of the best s-term approximation of the vector x, i.e., $\sigma_s(x)_q := \inf\{\|x - z\|_q : \|z\|_0 \leq s\}$. We first give the following inequalities taken from [2] (see, Theorem 2.5 and Lemma 6.10 therein).

Lemma 4.1. *(i) For any $z \in \mathbb{R}^n$, $\sigma_s(z)_2 \leq \frac{1}{2\sqrt{s}}\|z\|_1$. (ii) For any $u, v \in \mathbb{R}^n$ satisfying $\max\limits_{1\leq i\leq n}|u_i| \leq \min\limits_{1\leq i\leq n}|v_i|$, one has $\|u\|_2 \leq \frac{1}{\sqrt{n}}\|v\|_1$.*

We now establish a lemma which is an modification of Lemma 6.23 in [2] with ℓ_2-norm.

Lemma 4.2. *Given $x, x' \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$, $\nu \in \mathbb{R}^m$ and the scalars $\phi > 0$ and $\xi \geq 0$. Let x' be a k-sparse vector ($k \geq 2$) and $T := \mathcal{L}_k(x)$. If*

$$\|x_T - x'\|_2 \leq \phi\|Ax_{\overline{T}} + \nu\|_2 + \xi, \tag{4.1}$$

then

$$\|x - x'\|_2 \leq \frac{1 + 2\phi\sqrt{1+\delta_j}}{2\sqrt{j}}\sigma_j(x)_1 + \phi\|\nu\|_2 + \xi, \tag{4.2}$$

where $j = \lfloor\frac{k}{2}\rfloor$.

Proof. There are only two cases.

Case I: $k \geq 2$ is an odd integer number. In this case, $|T| = k = 2j + 1$. Denote $S_0 := \mathcal{L}_{j+1}(x) \subset T$ and $S_1 := T \setminus S_0$. It is not difficult to see that

$$\|x_{\overline{T}}\|_2 = \sigma_j(x_{\overline{S_0}})_2 \leq \frac{1}{2\sqrt{j}}\|x_{\overline{S_0}}\|_1 = \frac{1}{2\sqrt{j}}\sigma_{j+1}(x)_1, \tag{4.3}$$

where the first and final equalities follow from the definition of S_0, T and $\sigma_s(\cdot)_q$, and the inequality in between follows from Lemma 4.1(i). There exists an integer $r \geq 2$ such that $\overline{T}$ can be partitioned as $\overline{T} = \bigcup_{l=2}^{r} S_l$, where

$$S_2 = L_j(x_{\overline{T}}),\ S_3 = L_j(x_{\overline{T\cup S_2}}), \ldots, S_{r-1} = L_j(x_{\overline{T\cup S_2\cup\cdots\cup S_{r-2}}}),$$

$$S_r = \overline{T \cup S_2 \cup \cdots \cup S_{r-1}}$$

with cardinalities $|S_l| = j$ for $j = 2, \ldots, r-1$ and $|S_r| \leq j$. Using the triangular inequality together with (2.1), we obtain

$$\|Ax_{\overline{T}} + v\|_2 \leq \sum_{l=2}^{r} \|Ax_{S_l}\|_2 + \|v\|_2 \leq \sqrt{1+\delta_j} \sum_{l=2}^{r} \|x_{S_l}\|_2 + \|v\|_2. \tag{4.4}$$

Based on Lemma 4.1(ii), we observe that

$$\|x_{S_l}\|_2 \leq \frac{1}{\sqrt{j}} \|x_{S_{l-1}}\|_1, \quad 2 \leq l \leq r.$$

This together with (4.4) implies that

$$\begin{aligned} \|Ax_{\overline{T}} + v\|_2 \leq & \sqrt{\frac{1+\delta_j}{j}} \sum_{l=1}^{r-1} \|x_{S_l}\|_1 + \|v\|_2 \\ \leq & \sqrt{\frac{1+\delta_j}{j}} \|x_{\overline{S_0}}\|_1 + \|v\|_2 \\ = & \sqrt{\frac{1+\delta_j}{j}} \sigma_{j+1}(x)_1 + \|v\|_2. \end{aligned} \tag{4.5}$$

Combining (4.3), (4.1) with (4.5) leads to

$$\begin{aligned} \|x - x'\|_2 \leq & \|x_{\overline{T}}\|_2 + \|x_T - x'\|_2 \\ \leq & \frac{1}{2\sqrt{j}} \sigma_{j+1}(x)_1 + \phi \sqrt{\frac{1+\delta_j}{j}} \sigma_{j+1}(x)_1 + \phi \|v\|_2 + \xi. \end{aligned} \tag{4.6}$$

Case II: $k \geq 2$ is an even integer number. In this case, $|T| = k = 2j$. Denote $S_0 := \mathcal{L}_j(x) \subset T$. Repeating the argument in Case I and using the relation $\|x_{\overline{S_0}}\|_1 = \sigma_j(x)_1$ for this case, we obtain the following relation:

$$\|x - x'\|_2 \leq \frac{1}{2\sqrt{j}} \sigma_j(x)_1 + \phi \sqrt{\frac{1+\delta_j}{j}} \sigma_j(x)_1 + \phi \|v\|_2 + \xi. \tag{4.7}$$

Note that $\sigma_{j+1}(x)_1 \leq \sigma_j(x)_1$. Both (4.6) and (4.7) imply the desired relation (4.2) for any positive integer $k \geq 2$. ■

By using Lemma 4.2, the main result on the stability of HBHT can be stated as follows.

Theorem 4.1. *Suppose that the RIC, δ_{3k} ($k \geq 2$), of the matrix A and the parameters α and β satisfy the conditions in (3.3). Let $y := Ax + v$ be the measurements of x with measurement errors v. Then the sequence $\{x^p\}$, generated by HBHT with initial points $x^1 = x^0 = 0$, satisfies*

$$\left\|x - x^p\right\|_2 \leq \frac{1 + 2C_2\sqrt{1+\delta_j}}{2\sqrt{j}} \sigma_j(x)_1 + C_2\|v\|_2 + (\tau - b + 1)\tau^{p-1}\|x\|_2, \tag{4.8}$$

where $j = \lfloor \frac{k}{2} \rfloor$ and C_2, τ, b are given in Theorem 3.1.

Proof. According to (3.4), we know

$$\|x_S - x^p\|_2 \leq C_1\tau^{p-1} + C_2\|v'\|_2 = C_1\tau^{p-1} + C_2\|v + Ax_{\overline{S}}\|_2,$$

where $S = \mathcal{L}_k(x)$. This is the form of (4.1) in Lemma 4.2 with $x' = x^p$, $\phi = C_2$, $\xi = C_1\tau^{p-1}$ and $T = S$. Hence, it follows from (4.2) that

$$\|x - x^p\|_2 \leq \frac{1 + 2C_2\sqrt{1+\delta_j}}{2\sqrt{j}} \sigma_j(x)_1 + C_2\|v\|_2 + C_1\tau^{p-1}. \tag{4.9}$$

Substituting $x^1 = x^0 = 0$ into (3.5) yields

$$C_1 = \|x_S\|_2 + (\tau - b)\|x_S\|_2 \leq (\tau - b + 1)\|x\|_2. \tag{4.10}$$

Combining (4.9) and (4.10) leads to the desired estimation (4.8). ■

By a similar proof to the above, we obtain the stability result for HBHTP.

Theorem 4.2. *Suppose that the RIC, $\delta_{3k}(k \geq 2)$, of the matrix A and the parameters α and β satisfy the conditions in (3.13). Let $y := Ax + v$ be the measurements of x. Then the iterates $\{x^p\}$, generated by HBHTP with initial points $x^1 = x^0 = 0$, satisfies*

$$\|x - x^p\|_2 \leq \frac{1 + 2C_4\sqrt{1+\delta_j}}{2\sqrt{j}}\sigma_j(x)_1 + C_4\|v\|_2 + (\hat{\tau} - \hat{b} + 1)\hat{\tau}^{p-1}\|x\|_2, \tag{4.11}$$

where $j = \lfloor \frac{k}{2} \rfloor$ and $C_4, \hat{\tau}, \hat{b}$ are given in Theorem 3.2.

From (4.8) and (4.11), we see that the recovery error $\|x - x^p\|_2$ can be controlled and can be measured in terms of $\sigma_j(x)_1$, measurement errors, and the number of iterations performed. These results claims that a slight variance of these factors will not significantly affect the recovery error, and that if the signal is $j = \lfloor k/2 \rfloor$-compressible (i.e., $\sigma_j(x)_1$ is small) and if the measurements are accurate enough, then the signal will be recovered by the proposed algorithms provided that enough number of iterations are performed.

5. Numerical experiments

All mentioned experiments in this section were performed on a PC with the processor Intel(R) Core(TM) i7-10700 CPU @ 2.90 GHz and 16 GB memory. In these experiments, the measurement matrices $A \in \mathbb{R}^{m\times n}$ are Gaussian random matrices whose entries are independent and identically distributed (iid) and follow the standard normal distribution $\mathcal{N}(0, 1)$ in Section 5.1 and $\mathcal{N}(0, m^{-1})$ in Sections 5.2 and 5.3, respectively. All sparse vectors $x^* \in \mathbb{R}^n$ are also randomly generated, whose nonzero entries are iid and follow $\mathcal{N}(0, 1)$ and the position of nonzero entries follows the uniform distribution.

5.1. Comparison of performance

We first demonstrate some numerical results on recovery success rates of HBHT and HBHTP and the average number of iterations and CPU time required by these algorithms to achieve the recovery success of sparse signals. We compare their performances with the iterative algorithms OMP, SP, CoSaMP, HTP and IHT. We let HBHT and HBHTP start from $x^1 = x^0 = 0$ and other iterative algorithms start from $x^0 = 0$. The size of the matrices in this experiment is 400×800. All iterative algorithms are allowed to perform up to 50 iterations (which is set as the maximum number of iterations in our experiments), except for OMP which, by its structure, is performed exactly k iterations, equal to the sparsity level of the target signal x^*. For every given sparsity level k, 100 random examples of (A, x^*) are generated to estimate the success rates of algorithms. An individual recovery is called success if the solution produced by an algorithm satisfies the criterion

$$\|x^p - x^*\|_2 / \|x^*\|_2 \leq 10^{-3}. \tag{5.1}$$

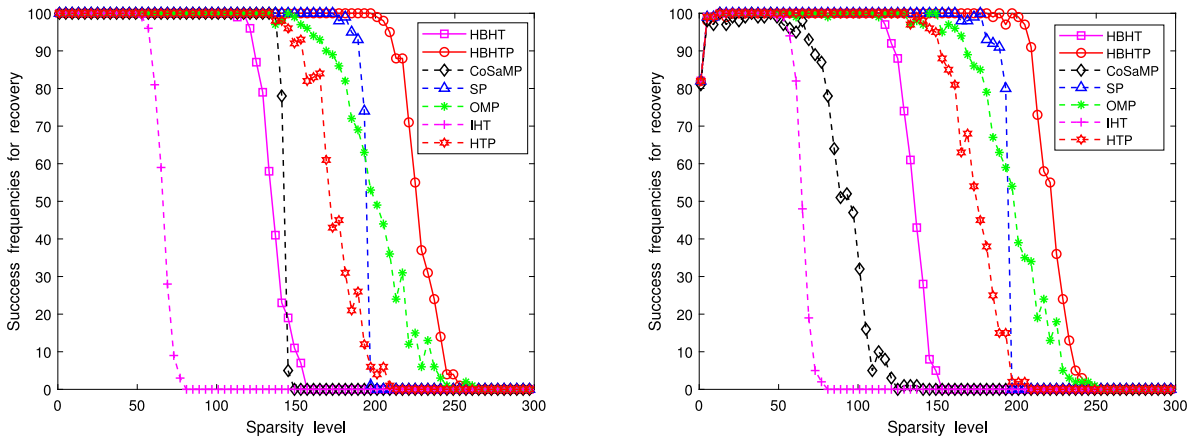
Let us first compare the algorithms in the case of A being un-normalized.

5.1.1. Performance with unnormalized matrices

The performance of iterative-type thresholding methods is closely related to the choice of stepsize in each step. When A is un-normalized/unscaled, initial simulations indicate that $\alpha = 10^{-3}$ is a proper choice for IHT and HTP, $\alpha \in [10^{-3}, 2 \times 10^{-3}]$ and $\beta \in (0, 0.6]$ are proper choices for HBHT, and $\alpha \in [10^{-3}, 8 \times 10^{-3}]$ and $\beta \in (0, 0.7]$ are suitable for HBHTP, where the range for β is implied from (3.22). We now start to compare algorithms using both accurate and inaccurate measurements. Given a random pair of (A, x^*), the accurate and inaccurate measurements are given respectively by $y := Ax^*$ and $y := Ax^* + \epsilon h$, where $\epsilon = 0.008$ and h is a standard Gaussian random noise vector. We use the fixed parameters $\alpha = 1.5 \times 10^{-3}$ and $\beta = 0.6$ in HBHT and $\alpha = 7 \times 10^{-3}$ and $\beta = 0.7$ in HBHTP. The estimated success rates of the algorithms are shown in Fig. 1 in which the sparsity level k is ranged from 1 to 297 with stepsize 4. It shows that the HBHTP generally outperforms the OMP, SP and HTP, and it might perform clearly better than CoSaMP, HBHT and IHT. We also observe from the experiments that the success rate of HBHT is slightly worse than that of CoSaMP in noiseless settings but it might be better than CoSaMP in noisy settings.

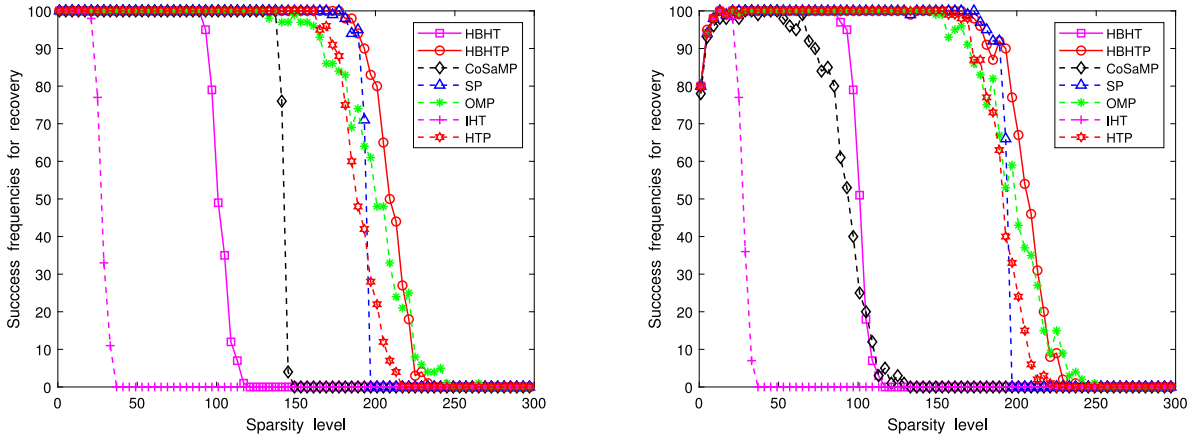
5.1.2. Performance with normalized matrices

The existing theory claims that the IHT and HTP with a larger stepsize such as $\alpha = 1$ remains convergent if the matrix satisfies the RIP, and it is well known that the normalized Gaussian matrix $\bar{A} := \frac{1}{\sqrt{m}}A$ may satisfy the RIP in high probability (see, e.g., Chapter 9 in [2] for details). In terms of a normalized matrix, the problem (1.1) is equivalent to $\underset{z}{\operatorname{argmin}}\{\|\bar{y} - \bar{A}z\|_2^2 : \|z\|_0 \leq k\}$, where $\bar{y} = \frac{1}{\sqrt{m}}y$. The entries of such a normalized Gaussian matrix follow the distribution $\mathcal{N}(0, m^{-1})$. By taking into account the theoretical results in previous sections and testing for the values of parameters (α, β), we found the choices $\alpha \in [0.4 + 2\beta, 1.6]$ and $\beta \in (0, 0.6]$ are suitable for HBHT and $\alpha \in [0.3 + 2\beta, 1.7]$ and $\beta \in (0, 0.7]$ are suitable for HBHTP to achieve a good performance. We repeated the experiments in Section 5.1.1 by setting the stepsize $\alpha = 1$ for IHT and HTP, the specific values $\alpha = 0.6$ and $\beta = 0.1$ for HBHT and $\alpha = 1.7$ and $\beta = 0.7$



(a) Accurate measurements (b) Inaccurate measurements

Fig. 1. Comparison of success frequencies (rates) of algorithms for signal recovery with accurate and inaccurate measurements, respectively. The parameters $\alpha = 1.5 \times 10^{-3}$ and $\beta = 0.6$ are set for HBHT and $\alpha = 7 \times 10^{-3}$ and $\beta = 0.7$ for HBHTP.



(a) Accurate measurements (b) Inaccurate measurements

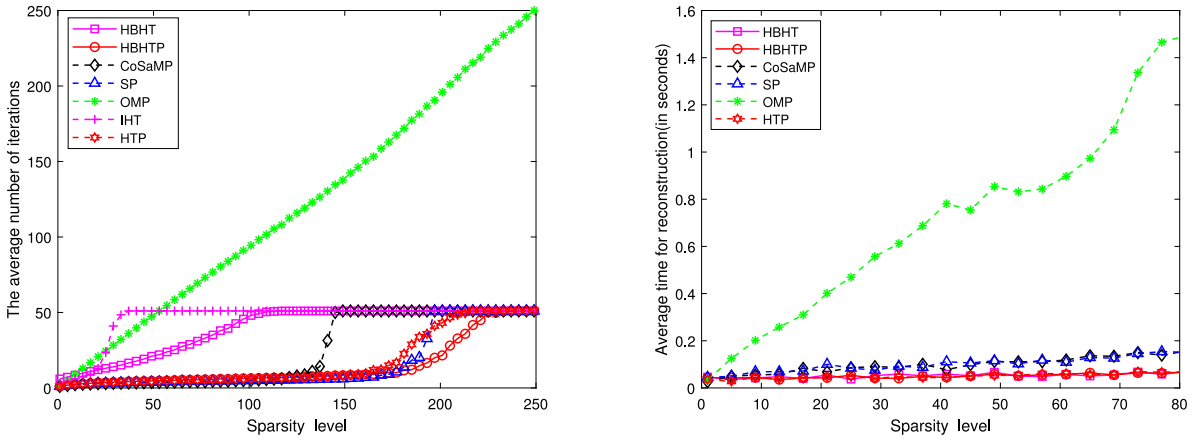
Fig. 2. Comparison of success frequencies of algorithms with accurate and inaccurate measurements, respectively. The parameters $\alpha = 0.6$ and $\beta = 0.1$ are set in HBHT, and $\alpha = 1.7$ and $\beta = 0.7$ are set in HBHTP.

for HBHTP. The results are demonstrated in Fig. 2 which appear to be similar to that of Fig. 1. However, one can observe that the normalization of the matrix, accordingly enlarged stepsize, and the choices of parameters do affect the recovery ability of HBHT, HBHTP, IHT and HTP to a certain degree. Again, it seems that the HBHTP performs generally better than other algorithms in noiseless and noisy settings, and the HBHT may perform better than CoSaMP and IHT in some noisy situations. Compared to IHT and HTP, the heavy-ball-based technique does play a vital role in speeding up and enhancing the performance of the traditional thresholding algorithms for sparse signal recovery.

5.1.3. Average number of iterations and time

We now compare the average number of iterations and CPU time required by several algorithms to meet the criterion (5.1) with accurate measurements. The testing environment is the same as Section 5.1.2. Within 50 iterations, if x^p satisfies criterion (5.1), the algorithm terminates and the number of iterations p is recorded. Otherwise, the number of iterations is recorded as 50. For OMP, the number of iterations is equal to the sparsity level of the input signal. Fig. 3(a) indicates that the average number of iterations required by HBHTP is lower than that of SP and HTP, and might be much lower than that of OMP, CoSaMP, HBHT and IHT especially when the sparsity level k is high.

Fig. 2(a) indicates that all k-sparse signals with $k \leq 80$ can be recovered by all mentioned algorithms except IHT. Thus we focus on the signals with sparsity levels $k \leq 80$ to compare the average time consumed by algorithms except



(a) Average number of iterations (b) Average CPU time

Fig. 3. Comparison of average number of iterations and time taken by algorithms to meet the recovery criterion (5.1) with accurate measurements. The parameters $\alpha = 0.6$ and $\beta = 0.1$ are set in HBHT and $\alpha = 1.7$ and $\beta = 0.7$ in HBHTP.

IHT to meet the criterion (5.1). The results are demonstrated in Fig. 3(b), from which one can see that OMP takes more time than other algorithms to recover the signal, and that the average time taken by SP, CoSaMP, HBHTP, HBHT and HTP increases slowly in a linear manner with respect to the sparsity level k, and the average time consumed by SP and CoSaMP is approximately twice of HBHT, HBHTP and HTP. This indicates that the proposed algorithms have some advantage in time saving for signal recovery.

5.2. Phase transition

We further investigate and compare the performances of algorithms through the empirical PTCs and ASM introduced in [12,40]. All $m \times n$ matrices in this subsection are Gaussian random matrices with fixed $n = 2^{12}$, whose entries are iid and follow the distribution $\mathcal{N}(0, m^{-1})$. The parameters (α, β) in HBHT and HBHTP are set the same exactly as in Section 5.1.2.

5.2.1. Phase transition curves

Denote by $\delta = m/n$ and $\rho = k/m$. The PTC of an algorithm separates the (δ, ρ) space into *success* and *failure* regions. The region below the curve, called recovery region, represents the problem instances with (δ, ρ) that can be exactly or approximately solved by the algorithm, while the region above the curve indicates the problem instances with (δ, ρ) to which the algorithm does not appear to find their correct solutions. The empirical PTCs demonstrated in this section are logistic regress curves identifying the 50% success rate for the given algorithm applying to a given problem class. This method was first introduced in [12,40].

We now briefly introduce the mechanism for generating such a curve. The interested readers may find more detailed information about this from the Refs. [12,40]. To generate the PTCs and ASM, we consider 25 different values of $m = \lceil \delta \cdot n \rceil$ where

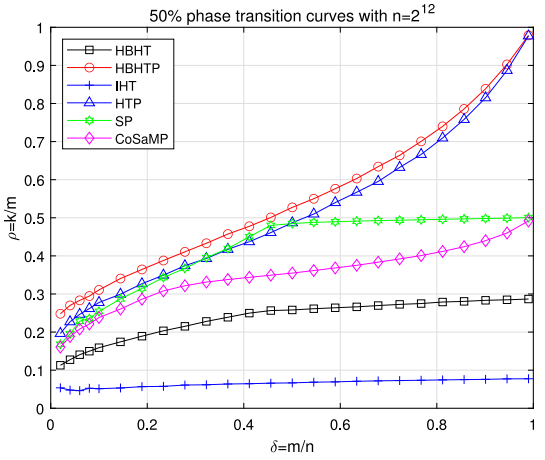
$$\delta \in \{0.02, 0.04, 0.06, 0.08\} \cup \{0.1, 0.1445, \ldots, 0.99\}, \tag{5.2}$$

where the interval $[0.1, 0.99]$ was equally divided into 20 parts. For every value of m, we collect 50 groups of sparsity levels $k = \lceil \rho \cdot m \rceil$ where ρ is ranged from 0.02 to 1 with stepsize 0.02. For a fixed m, the recovery phase transition region for each algorithm is estimated by the interval $[k_{\min}, k_{\max}]$, where $k_{\min}$ and $k_{\max}$ can be determined by a bisection method. They are the critical values to ensure that the recovery success rate is at least 90% for any $k < k_{\min}$ and at most 10% for any $k > k_{\max}$. For simplicity, we introduce the notations $k_j \overset{\Delta}{=} k_{\min} + \lceil j \cdot \Delta k \rceil (j = 0, 1, \ldots, J)$, where $\Delta k = (k_{\max} - k_{\min})/J$ and $J = k_{\max} - k_{\min}$ if $k_{\max} - k_{\min} < 50$; otherwise $J = 50$. When estimating the success rate of an algorithm, $Nb = 10$ problem instances are tested for each given (k, m, n), where $k = k_j, j = 0, 1, \ldots, J$. Based on the success rates, the PTCs can be obtained from the following logistic regression model [12,40]:

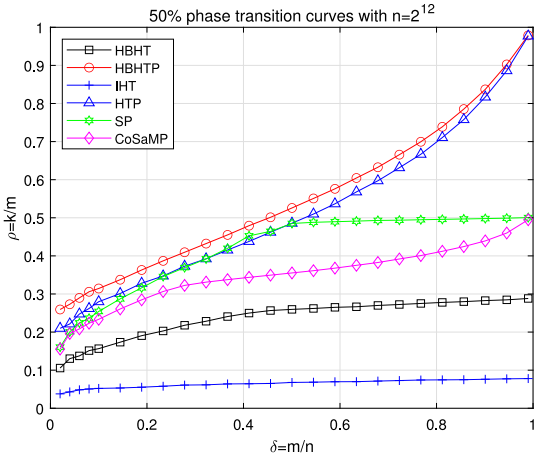
$$\min_{(\gamma_0, \gamma_1)} \sum_{j=0}^{J} \left| g(k_j/m) - \frac{suc(k_j, m, n)}{Nb} \right|,$$

where

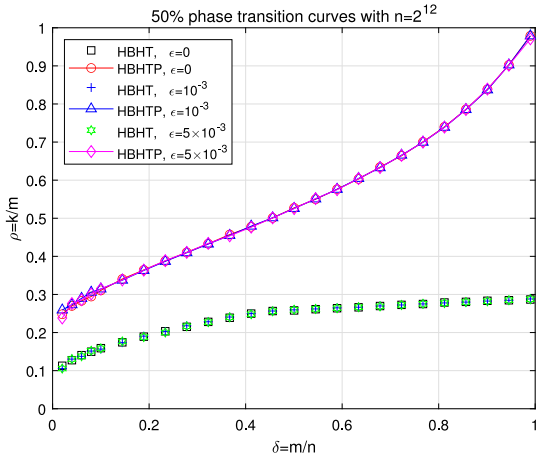
$$g(\rho) = \frac{1}{1 + \exp(-\gamma_0(1 - \gamma_1 \rho))},$$



(a) Accurate measurements



(b) Inaccurate measurements



(c) Different noises

Fig. 4. The 50% success rate PTCs for six algorithms.

and $suc(k_j, m, n)$ is the number of recovery success among Nb problem instances for each $(k_j, m, n), j = 0, 1, \ldots, J$. The 50% success recovery PTCs are defined by $g(\rho) = 0.5$.

The curves for the algorithms HBHT, HBHTP, IHT, HTP, CoSaMP and SP are summarized in Fig. 4. In this comparison, the parameters $\alpha = 0.6$ and $\beta = 0.1$ are used in HBHT and $\alpha = 1.7$ and $\beta = 0.7$ in HBHTP. The accurate and inaccurate measurements are given by $y = Ax^*$ and $y = Ax^* + \epsilon h$, respectively, where h is a standard Gaussian random vector and $\epsilon = 0.001$. From Fig. 4(a) and (b), we see that HBHTP has the highest PTC. This indicates that HBHTP may outperform the other five algorithms for sparse signal recovery in both noiseless and noisy environments. One can also see that the PTCs of SP, CoSaMP, HBHT and IHT are below the line $\rho = 0.5$ as $\delta \geq 0.5$. This implies that the recovery performance of these algorithms would not remarkably be improved even when the number of measurements is increased. By contrast, the PTCs of HBHTP and HTP are twice as high as those of SP and CoSaMP as $\delta \to 1$. To see the influence of noise levels on the performance of algorithms, the PTCs for HBHT and HBHTP with three different noise levels $\epsilon \in \{0, 10^{-3}, 5 \times 10^{-3}\}$ are demonstrated in Fig. 4(c), from which one can observe that the curves of HBHT and HBHTP do not significantly change with respect to the noise level when the noise level is relatively low. This sheds light on the stability of the two algorithms in signal recovery.

5.2.2. Algorithm selection map

The intersection of the recovery regions below the PTCs indicates that multiple algorithms are capable of signal recovery. To choose an algorithm, one might also consider the computational time for recovery. As a result, the so-called ASM was introduced in [12,40], which demonstrates the least average recovery time of the algorithms with accurate

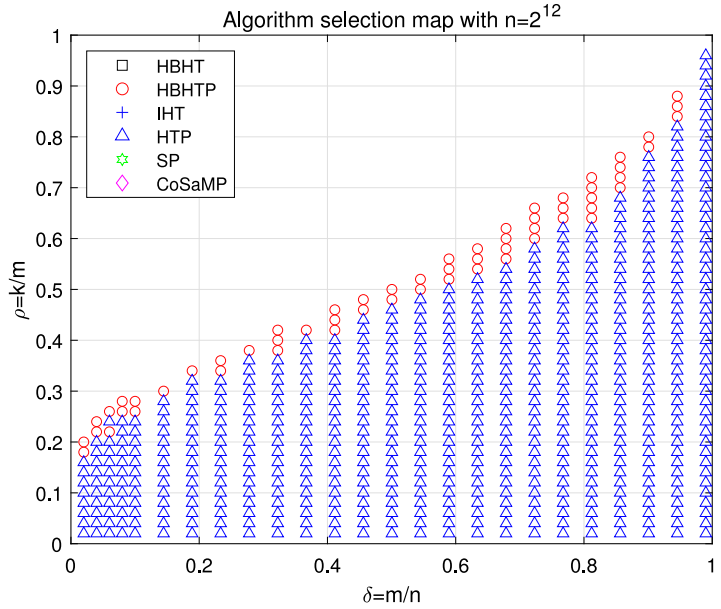


Fig. 5. Selection map with accurate measurements.

measurements. To draw an ASM, for each δ taking the values in (5.2), 10 problem instances are tested for every algorithm on the sampled phase space with the mesh (δ, ρ) with $\rho = \{j/50, j = 1, 2, \ldots, 50\}$ until the success rate is lower than 90%. The algorithm with least computational time will be identified on the map. The map is shown in Fig. 5, which clearly depicts two regions in the phase plane, wherein HBHTP is the fastest algorithm for solving problem instances with relatively large ρ, while the HTP reliably recovers the signal in least time in other cases.

After identifying the fastest algorithm, further information on the average recover time of algorithms are given in Fig. 6. The minimum average recovery time taken by the fastest algorithm is displayed in Fig. 6(a). When $\rho \leq 0.3$, the minimal average run time of algorithms is close to each other for any $\delta \in (0, 1)$. However, when $\rho > 0.3$, we see that the larger the value of ρ, the more average run time is required by the algorithm when $\delta > 0.6$. The ratios of the average recovery time for the algorithms HBHTP, HBHT, HTP, SP and CoSaMP against that of the fastest algorithm are displayed in Fig. 6(b)–(f), respectively. Fig. 6(b) shows that the larger the value of ρ, the smaller the ratio for a fixed δ, and the ratio for HBHTP is less than 1.5 when $\rho \geq 0.25$ or $\delta \leq 0.1$. By contrast, Fig. 6(c)–(f) show that the larger the value of ρ, the larger the ratios for those four algorithms. This phenomenon indicates that HBHTP might work better than other algorithms when the sparsity level k is relatively high. We also observe that HBHTP and HTP are comparable to each other, and that HBHT, SP and CoSaMP often consume more than twice of the minimal average time. One can also observe that the ratios for SP and CoSaMP can be three and five times higher, respectively, when ρ is large.

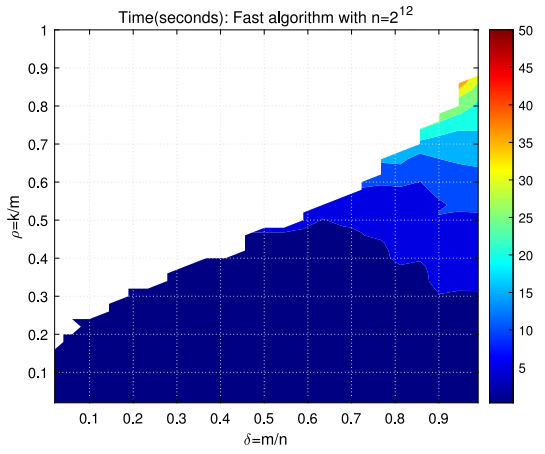
Finally, we demonstrate the change of average recovery time of algorithms against the factor ρ. The results for three different parameters

$$\delta \in \{0.2780, 0.5005, 0.7230\}$$

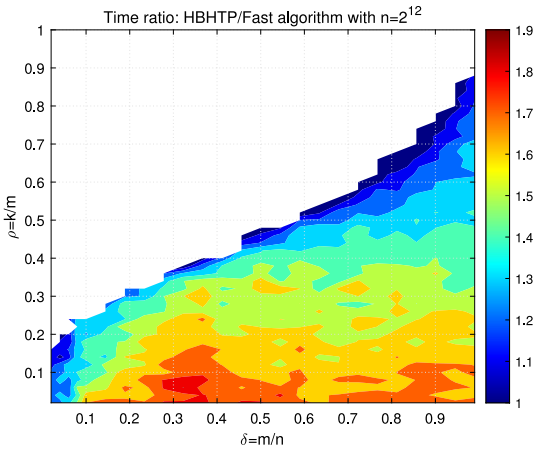
are given in Fig. 7. For $\rho \leq 0.2$, the average recover time of HBHTP, SP and CoSaMP are similar to each other. For $\rho \in [0.2, 0.5]$, the time consumed by HBHTP and HTP increases slowly compared to that of SP and CoSaMP as the sparsity level k increases. Moreover, the computational time of HTP approaches and surpasses that of HBHTP for $\rho \geq 0.4$ in Fig. 7(b) and for $\rho \geq 0.5$ in Fig. 7(c), respectively. Finally, we find that only HBHTP is typically able to recover the sparse signals fell into the region of the far right of Fig. 7(a)–(c). This provides some evidence to show that the HBHTP might admit a certain advantage in sparse signal recovery over several existing algorithms especially when ρ is relatively large.

5.3. Application to image reconstruction

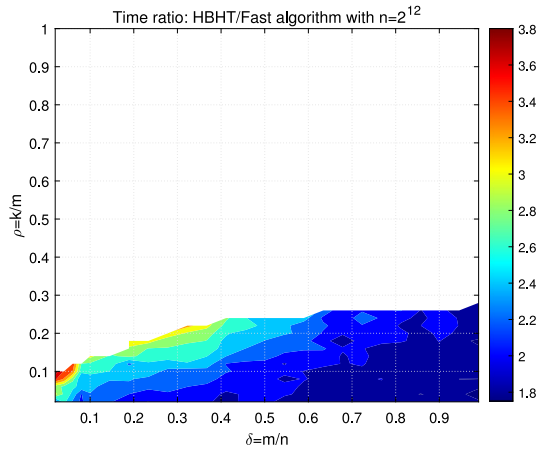
The performances of HBHT and HBHTP in image reconstruction are also verified via a few standard test images including Baboon, Barbara, Goldhill, Lena and Peppers. These images have the same size $n \times n$ with $n = 512$. The sparse representation of images are achieved by the discrete wavelet transform with *'sym8'* wavelet. In this experiment, all algorithmic parameters of HBHT and HBHTP are set exactly as in Sections 5.1.2 and 5.2 and the normalized Gaussian matrix is taken as the measurement matrix. The input sparsity-level in HBHT and HBHTP is set as $k = \lceil n/10 \rceil$.



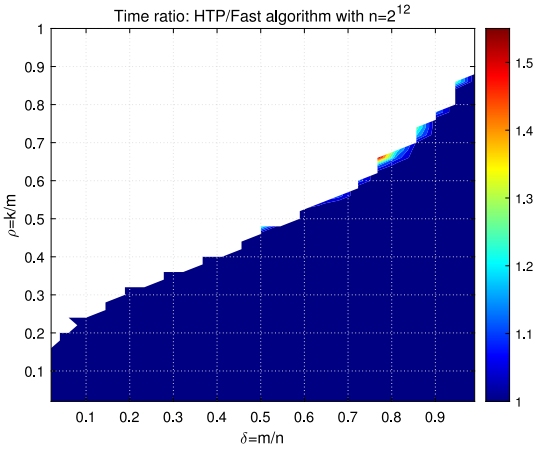
(a) Average time for the fastest algorithm



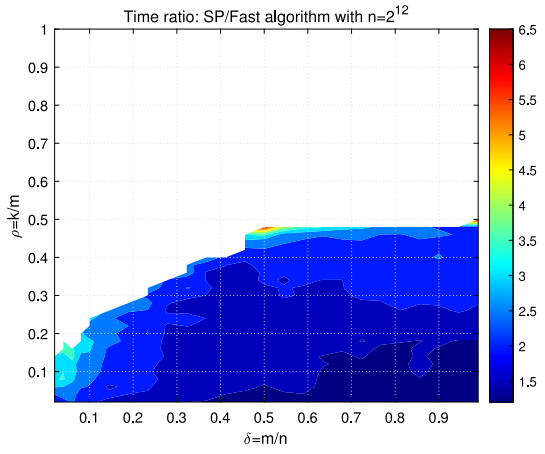
(b) HBHTP



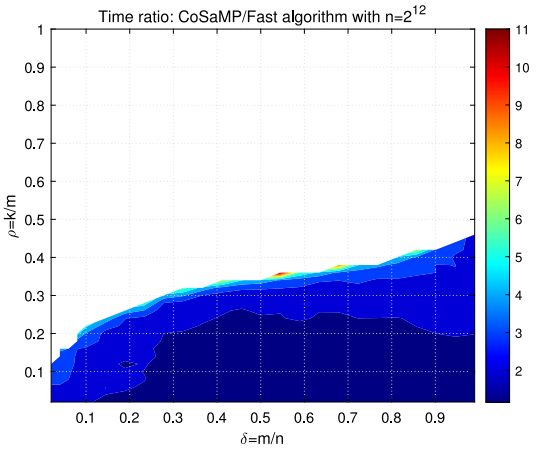
(c) HBHT



(d) HTP



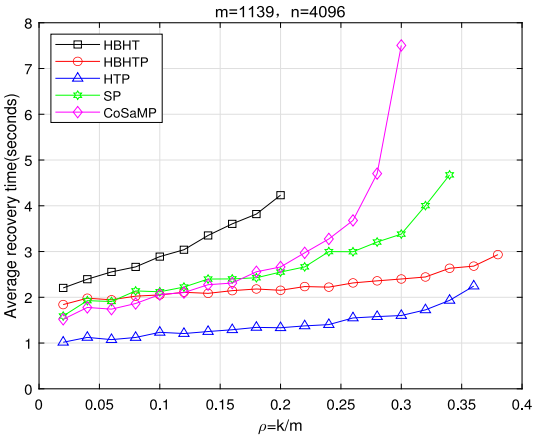
(e) SP



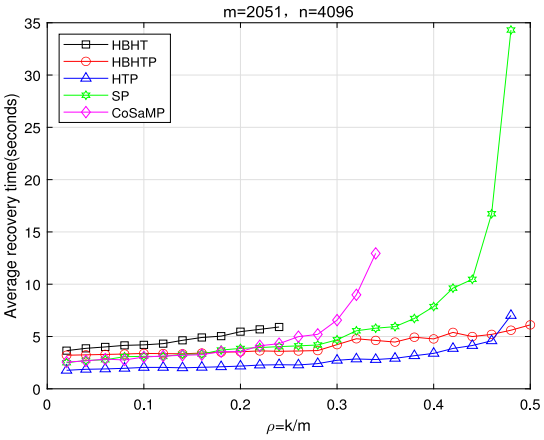
(f) CoSaMP

Fig. 6. (a) The minimum average time of algorithms; (b)–(f) The ratios of average time for several algorithms against the fastest one.

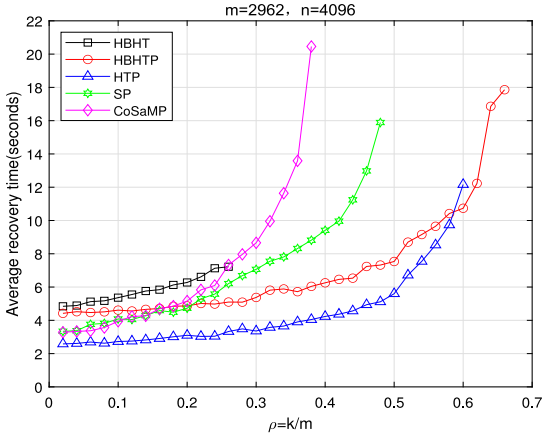
The reconstruction quality with HBHT and HBHTP is measured by the standard PSNR in three different sampling rates $\delta = m/n$, as shown in Tables 3 and 4, respectively. The first row of Tables 3 and 4 corresponds to the noiseless



(a) $\delta = 0.2780$



(b) $\delta = 0.5005$



(c) $\delta = 0.7230$

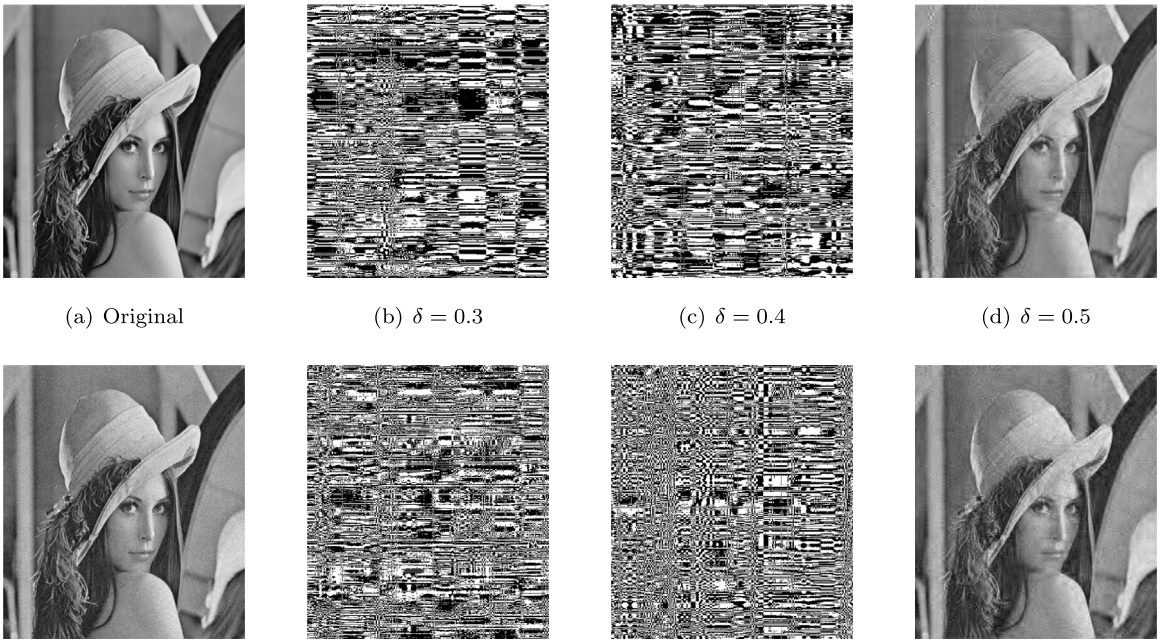
Fig. 7. Average recovery time with respect to the change of ρ under three different fixed values of δ.

Table 3
Comparison of PSNR (dB) for HBHT with different sampling rates.

	δ	Baboon	Barbara	Goldhill	Lena	Peppers
Noiseless cases	0.3	27.45	27.34	27.2	27.25	27.16
	0.4	27.62	27.08	27.05	27.36	27.12
	0.5	29.48	31.86	31.19	34.03	32.77
Noisy cases	0.3	27.49	27.27	27.43	27.19	27.12
	0.4	27.12	27.18	27.26	27.08	27.09
	0.5	29.2	29.55	30.03	30.67	30.49

situations and the second row corresponds to the noisy situations where the Gaussian white noise with mean 0 and variance 2.5×10^{-3} is added into the image by using the *'imnoise'* function in Matlab. From Table 3, we observe that the PSNR values of all images are less than 27.7 dB in both noiseless and noisy settings as $\delta = 0.3$ or 0.4. This is normal since the sampling is not enough to ensure a quality reconstruction. However, the PSNRs are greater than or equal to 29.2 dB as δ is increased to 0.5. Table 4 indicates that HBHTP is a robust reconstruction algorithm in both noiseless and noisy scenarios.

As an example, the reconstructed 'Lena' images by HBHT and HBHTP using different sampling rates are demonstrated in Figs. 8 and 9, respectively, in which the second rows are the results corresponding to noisy cases and 2–4 columns are the ones reconstructed by the algorithms. Fig. 8 indicates that HBHT fails to reconstruct images as $\delta = 0.3$ and 0.4, while it successfully reconstruct the image when $\delta = 0.5$ and in noisy situations. This is consistent with the result in Table 3.



(a) Original (b) $\delta = 0.3$ (c) $\delta = 0.4$ (d) $\delta = 0.5$

(e) Noisy case (f) $\delta = 0.3$ (g) $\delta = 0.4$ (h) $\delta = 0.5$

Fig. 8. Performance of HBHT for Lena with different sampling rates.

Table 4
Comparison of PSNR (dB) for HBHTP with different sampling rates.

	δ	Baboon	Barbara	Goldhill	Lena	Peppers
Noiseless cases	0.3	28.46	30.27	29.58	31.76	30.63
	0.4	28.76	31.78	30.43	33.67	32.44
	0.5	29.18	32.76	31.17	34.75	33.42
Noisy cases	0.3	28.24	28.93	28.78	29.27	29.14
	0.4	28.48	29.42	29.35	29.83	29.64
	0.5	28.78	29.84	29.66	30.28	30.19

The quality of reconstructed images is improved using HBHTP and relatively higher values of δ in both noiseless and noisy settings.

6. Conclusions

Incorporating the heavy-ball acceleration technique into the IHT and HTP methods leads to the HBHT and HBHTP algorithms for sparse signal recovery. The guaranteed performance of these algorithms has been established under the RIP assumption and suitable conditions for the choice of algorithmic parameters. The finite convergence of HBHTP and recovery stability of the two algorithms were also shown in this paper. The numerical performance of algorithms has been investigated from several difference perspectives including the recovery success rate, average number of iterations and computational times. Comparison of the proposed algorithms with a few existing ones is also made through the phase transition analysis including the PTC and ASM. Simulations on random problem instances indicate that under proper choices of parameters, HBHTP is an efficient algorithm for sparse signal recovery and it may outperform several existing algorithms in many cases.

(a) Original (b) $\delta = 0.3$ (c) $\delta = 0.4$ (d) $\delta = 0.5$

(e) Noisy case (f) $\delta = 0.3$ (g) $\delta = 0.4$ (h) $\delta = 0.5$

Fig. 9. Performance of HBHTP for Lena with different sampling rates.

CRediT authorship contribution statement

Zhong-Feng Sun: Software, Visualization, Writing – original draft, Investigation. **Jin-Chuan Zhou:** Investigation, Methodology. **Yun-Bin Zhao:** Supervision, Conceptualization, Methodology, Investigation, Writing – review & editing. **Nan Meng:** Experiment, Software, Validation.

Data availability

Data will be made available on request.

Acknowledgments

The work was founded by the National Natural Science Foundation of China (NSFC#12071307 and 11771255), Young Innovation Teams of Shandong Province (#2019KJI013), and Domestic and Oversea Visiting Program for the Middle-aged and Young Key Teachers of Shandong University of Technology. The authors thank Prof. Dr. André A. Keller and two anonymous reviewers for their helpful comments and suggestions that help improve the paper. The authors also thank Dr. Yingmei Wang for her discussion on the experiments in Section 5.3.

References

[1] M. Elad, Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing, Springer, New York, 2010.
[2] S. Foucart, H. Rauhut, A Mathematical Introduction to Compressive Sensing, Springer, New York, 2013.
[3] Y.B. Zhao, Sparse Optimization Theory and Methods, CRC Press, Boca Raton, FL, 2018.
[4] D.L. Donoho, I.M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, Biometrika 81 (3) (1994) 425–455.
[5] T. Blumensath, M.E. Davies, Iterative thresholding for sparse approximations, J. Fourier Anal. Appl. 14 (2008) 629–654.
[6] T. Blumensath, M.E. Davies, Iterative hard thresholding for compressed sensing, Appl. Comput. Harmon. Anal. 27 (3) (2009) 265–274.
[7] T. Blumensath, M.E. Davies, Normalized iterative hard thresholding: Guaranteed stability and performance, IEEE J. Sel. Top. Signal Process. 4 (2) (2010) 298–309.
[8] V. Cevher, On accelerated hard thresholding methods for sparse approximation, in: Proc. SPIE Wavelets Sparsity XIV, 2011, 813811.
[9] S. Foucart, Hard thresholding pursuit: An algorithm for compressive sensing, SIAM J. Numer. Anal. 49 (6) (2011) 2543–2563.
[10] T. Blumensath, Accelerated iterative hard thresholding, Signal Process. 92 (3) (2012) 752–756.
[11] A. Kyrillidis, V. Cevher, Matrix recipes for hard thresholding methods, J. Math. Imaging Vision 48 (2014) 235–265.
[12] J.D. Blanchard, J. Tanner, K. Wei, CGIHT: Conjugate gradient iterative hard thresholding for compressed sensing and matrix completion, IMA J. Inf. Inference 4 (4) (2015) 289–327.

[13] R. Khanna, A. Kyrillidis, IHT dies hard: Provable accelerated iterative hard thresholding, in: Proc. 21st Int. Conf. Artif. Intell. Stat, Vol. 84, 2018, pp. 188–198.
[14] I. Daubechies, M. Defrise, C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Comm. Pure Appl. Math. 57 (11) (2004) 1413–1457.
[15] D.L. Donoho, De-noising by soft-thresholding, IEEE Trans. Inform. Theory 41 (3) (1995) 613–627.
[16] M. Elad, Why simple shrinkage is still relevant for redundant representations, IEEE Trans. Inform. Theory 52 (12) (2006) 5559–5569.
[17] Y.B. Zhao, Optimal *k*-thresholding algorithms for sparse optimization problems, SIAM J. Optim. 30 (1) (2020) 31–55.
[18] Y.B. Zhao, Z.Q. Luo, Analysis of optimal thresholding algorithms for compressed sensing, Signal Process. 187 (2021) 108148.
[19] N. Meng, Y.B. Zhao, M. Kočvara, Z. Sun, Partial gradient optimal thresholding algorithms for a class of sparse optimization problems, J. Global Optim. 84 (2022) 393–413.
[20] R. Garg, R. Khandekar, Gradient descent with sparsification: An iterative algorithm for sparse recovery with restricted isometry property, in: Proc. 26th Annu. Int. Conf. Mach. Learn., 2009, pp. 337–344.
[21] B.T. Polyak, Some methods of speeding up the convergence of iteration methods, USSR Comput. Math. Math. Phys. 4 (5) (1964) 1–17.
[22] B.T. Polyak, Introduction To Optimization, Optimization Software, Inc. Publications Division, New York, 1987.
[23] E. Ghadimi, H.R. Feyzmahdavian, M. Johansson, Global convergence of the heavy-ball method for convex optimization, in: 2015 Eur. Control Conf., 2015, pp. 310–315.
[24] M. Gürbüzbalaban, A. Ozdaglar, P.A. Parrilo, On the convergence rate of incremental aggregated gradient algorithms, SIAM J. Optim. 27 (2) (2017) 1035–1048.
[25] L. Lessard, B. Recht, A. Packard, Analysis and design of optimization algorithms via integral quadratic constraints, SIAM J. Optim. 26 (1) (2016) 57–95.
[26] R. Xin, U.A. Khan, Distributed heavy-ball: A generalization and acceleration of first-order methods with gradient tracking, IEEE Trans. Automat. Control 65 (6) (2020) 2627–2633.
[27] K. Huang, S. Zhang, A unifying framework of accelerated first-order approach to strongly monotone variational inequalities, 2021, arXiv preprint arXiv:2103.15270v1.
[28] J. Liu, A. Eryilmaz, N.B. Shroff, E.S. Bentley, Heavy-ball: A new approach to tame delay and convergence in wireless network optimization, in: 35th Annu. IEEE Int. Conf. Comput. Commun., 2016, pp. 1–9.
[29] P. Ochs, Y. Chen, T. Brox, T. Pock, Ipiano: Inertial proximal algorithm for nonconvex optimization, SIAM J. Imaging Sci. 7 (2) (2014) 1388–1419.
[30] T. Sun, D. Li, Z. Quan, H. Jiang, S. Li, Y. Dou, Heavy-ball algorithms always escape saddle points, 2019, arXiv preprint arXiv:1907.09697v1.
[31] W. Tao, S. Long, G. Wu, Q. Tao, The role of momentum parameters in the optimal convergence of adaptive Polyak's heavy-ball methods, 2021, arXiv preprint arXiv:2102.07314v1.
[32] H. Wang, P.C. Miller, Scaled heavy-ball acceleration of the Richardson–Lucy algorithm for 3D microscopy image restoration, IEEE Trans. Image Process. 23 (2) (2014) 848–854.
[33] E.J. Candès, T. Tao, Decoding by linear programming, IEEE Trans. Inform. Theory 51 (12) (2005) 4203–4215.
[34] Y.B. Zhao, Z.Q. Luo, Improved RIP-based bounds for guaranteed performance of two compressed sensing algorithms, Sci. China Math. (2022) http://dx.doi.org/10.1007/s11425-021-1987-2.
[35] G. Davis, S. Mallat, Z. Zhang, Adaptive time-frequency decompositions, Opt. Eng. 33 (7) (1994) 2183–2191.
[36] J.A. Tropp, A.C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, IEEE Trans. Inform. Theory 53 (12) (2007) 4655–4666.
[37] D. Needell, J.A. Tropp, CoSaMP: Iterative signal recovery from incomplete and inaccurate samples, Appl. Comput. Harmon. Anal. 26 (3) (2009) 301–321.
[38] W. Dai, O. Milenkovic, Subspace pursuit for compressive sensing signal reconstruction, IEEE Trans. Inform. Theory 55 (5) (2009) 2230–2249.
[39] J. Shen, P. Li, A tight bound of hard thresholding, J. Mach. Learn. Res. 18 (208) (2018) 1–42.
[40] J.D. Blanchard, J. Tanner, Performance comparisons of greedy algorithms in compressed sensing, Numer. Linear Algebra Appl. 22 (2) (2015) 254–282.