Applications of Mathematics

Jinyao Ma; Haibin Zhang; Shanshan Yang; Jiaojiao Jiang

A tight bound of modified iterative hard thresholding algorithm for compressed sensing

*Applications of Mathematics*, Vol. 68 (2023), No. 5, 623–642

Persistent URL: http://dml.cz/dmlcz/151836

# A TIGHT BOUND OF MODIFIED ITERATIVE HARD THRESHOLDING ALGORITHM FOR COMPRESSED SENSING

Jinyao Ma, Haibin Zhang, Shanshan Yang, Beijing,
Jiaojiao Jiang, Sydney

Received September 13, 2022. Published online March 21, 2023.

*Abstract.* We provide a theoretical study of the iterative hard thresholding with partially known support set (IHT-PKS) algorithm when used to solve the compressed sensing recovery problem. Recent work has shown that IHT-PKS performs better than the traditional IHT in reconstructing sparse or compressible signals. However, less work has been done on analyzing the performance guarantees of IHT-PKS. In this paper, we improve the current RIP-based bound of IHT-PKS algorithm from $\delta_{3s-2k} < \frac{1}{\sqrt{32}} \approx 0.1768$ to $\delta_{3s-2k} < \frac{\sqrt{5}-1}{4} \approx 0.309$, where $\delta_{3s-2k}$ is the restricted isometric constant of the measurement matrix. We also present the conditions for stable reconstruction using the IHT$^{\mu}$-PKS algorithm which is a general form of IHT-PKS. We further apply the algorithm on Least Squares Support Vector Machines (LS-SVM), which is one of the most popular tools for regression and classification learning but confronts the loss of sparsity problem. After the sparse representation of LS-SVM is presented by compressed sensing, we exploit the support of bias term in the LS-SVM model with the IHT$^{\mu}$-PKS algorithm. Experimental results on classification problems show that IHT$^{\mu}$-PKS outperforms other approaches to computing the sparse LS-SVM classifier.

*Keywords*: iterative hard thresholding; signal reconstruction; classification problem; least squares support vector machine

*MSC 2020*: 34B16, 34C25, 90C31

## 1. Introduction

According to the theory of compressive sensing (CS), one can effectively compress the data while acquiring the signal and the sampling frequency can be lower than the Nyquist frequency, which decreases the sample data and frees up the storage

The research has been supported by the National Natural Science Foundation of China (11771003).

DOI: 10.21136/AM.2023.0221-22 

space. It also provides enough information. Currently, compressed sensing is widely used in single-pixel cameras, medical imaging, radar, sampling theory, statistics, and machine learning. The standard compressed sensing problem is to reconstruct the sparse vector $x \in \mathbb{R}^n$ from the underdetermined measurement $y = Ax \in \mathbb{R}^m$, where $m < n$, $A \in \mathbb{R}^{m\times n}$ is the measurement matrix, and $y$ is the measurement vector. The vector $x$ is referred as $s$-sparse if $\|x\|_0 \leqslant s$, where the '$l_0$ norm' $\|x\|_0$ counts the number of nonzero entries in $x$ and $s$ is an integer number. The first algorithmic approach is $l_0$ minimization, which tries to reconstruct $x$ as a solution of the optimization problem

$$(\mathrm{P}_0) \qquad \min_{x\in\mathbb{R}^n} \|x\|_0 \quad \text{s.t.} \quad Ax = y.$$

However, it is a nonconvex problem and NP-hard in general. The second method solves a convex relaxed $(\mathrm{P}_0)$ and is known as $l_1$ minimization or basis pursuit [10],

$$(\mathrm{P}_1) \qquad \min_{x\in\mathbb{R}^n} \|x\|_1 \quad \text{s.t.} \quad Ax = y.$$

It has been shown that the sparse vector $x$ can be recovered stably provided that $x$ is sufficiently sparse and the matrix $A$ obeys a condition known as the restricted isometry property (RIP) [12]. That is to say the restricted isometry constant, denoted as $\delta_s$, of the matrix $A$ is smaller than a certain threshold value $\delta^*$, where $\delta^* \in (0, 1)$. The condition $\delta_s < \delta^*$, which ensures the success of signal recovery of a given algorithm, is called the restricted isometry property based (RIP-based) bound of the algorithm.

Compressed sensing algorithms can be classified into three categories: optimization methods, greedy methods and thresholding-based methods. The IHT algorithm belongs to the third category that provides near-optimal error guarantees but does not require matrix inversion [6]. There is a large amount of literature on the optimization properties of the IHT algorithm [1], [19], [29]. In this paper, we research a tighter RIP-based bound for an improved IHT algorithm. The IHT was first introduced by Blumensath et al. [4] and its main iteration step is

$$x^{(t+1)} = H_s(x^{(t)} + A^\top(y - Ax^{(t)})),$$

where the operator $H_s(z) \in \mathbb{R}^n$ called the hard thresholding operator retains the $s$ largest absolute entries of $z$ and sets other entries to zeros. Much work has been focusing on the RIP-based bound for guaranteed performance of the IHT algorithm. Blumensath et al. showed that under the condition of $\delta_{3s} < \frac{1}{\sqrt{32}} \approx 0.1768$, it can recover sparse signals stably [5]. Later, this result was improved to $\delta_{3s} < \frac{1}{2\sqrt{3}} \approx 0.2886$ by Foucart [13]. Recently, the result was improved to $\delta_{3s} < \frac{\sqrt{5}-1}{4} \approx 0.309$ by Zhao and Luo [31]. It is worth mentioning that Blumensath et al. developed

the RIP-based bound for the performance of IHT according to the geometric rate $\alpha \leqslant \frac{1}{2}$ instead of $\alpha < 1$. For the convenience of comparison, the results obtained by Foucart [13] and Zhao [31] are also provided in the case that the geometric rate $\alpha \leqslant \frac{1}{2}$ in this paper.

Researchers have found that the reconstruction of sparse or compressible signals with partially known support (PKS) in modified compressed sensing shows better performances than the traditional compressed sensing [25], [16], [3], [2], [11]. Different from the traditional basis pursuit (i.e., $l_1$ minimization) to solve the minimum $l_1$ norm of the overall sparse vector $x$, Vaswani et al. [24] considered minimizing $l_1$ norm of the sparse signal $x$ which does not contain the known partial support set. They proposed the modified basis pursuit

$$\min_{x\in\mathbb{R}^n} \|x_{T_0^c}\|_1 \quad \text{s.t.} \quad y = Ax \tag{1.1}$$

to recover the sparse signal, where $T_0$ is the known part of the support of $x$, $T_0^c$ is the complement of $T_0$, and $x_{T_0^c}$ denotes the vector in $\mathbb{R}^n$ equal to $x$ on $T_0^c$ and equal to zero on $T_0$. Jacques [15] later proved that this technique can also extend to the case of corrupted measurements and compressible signals. They proposed to solve the problem

$$\min_{x\in\mathbb{R}^n} \|x_{T_0^c}\|_1 \quad \text{s.t.} \quad \|y - Ax\|_2 \leqslant \varepsilon \tag{1.2}$$

and proved that the sparse signal can be recovered stably if $\delta_{2c}^2 + 2\delta_{k+2c} < 1$, where $c \in \mathbb{N}$, $k = |T_0|$. Differently from the traditional algorithm to solve the minimum $l_1$ norm of the overall sparse vector $x$, the above problem is solved by the minimum $l_1$ norm of the sparse signal $x$ which does not contain the known partial support set.

Carrillo et al. [9] further proposed the iterative hard thresholding with partially known support (IHT-PKS) algorithm to solve the compressed sensing problem. The main iteration step of IHT-PKS is

$$x^{(t+1)} = H_{s-k}^{T_0}(x^{(t)} + A^\top(y - Ax^{(t)})),$$

where $k = |T_0|$. The algorithm selects the first $s - k$ elements with the largest absolute value that are not in $T_0$ at each iteration and retains the elements in $T_0$. It was shown that when $\delta_{3s-2k} < \frac{1}{\sqrt{32}} \approx 0.1768$, the $s$-sparse signal can be stably recovered. Their experimental results demonstrated that IHT-PKS performs better than other algorithms such as the orthogonal matching pursuit (OMP) [17], compressed sampling matching pursuit (CoSaMP) [30], etc. Carrillo et al. also extended the ideas of modified compressed sensing to various greedy algorithms [8]. The benefits of known partial support on the performance of joint-sparse recovery algorithms were also studied by Besson et al. [3].

In this paper, we prove that the result in [9] for IHT-PKS can be improved to $\delta_{3s-2k} < \frac{\sqrt{5}-1}{4} \approx 0.309$, which is also a tighter RIP-based bound than that required by IHT in [31] due to $\delta_{3s-2k} \leqslant \delta_{3s}$. Note that a general form of the IHT-PKS algorithm could allow stepsizes $\mu$ in front of $A^\top(y - Ax^{(t)})$, so we also present a sufficient condition for the stable recovery of the IHT$^\mu$-PKS algorithm. We further apply the IHT$^\mu$-PKS algorithm on the sparse Least Squares Support Vector Machine (LS-SVM) to obtain sparse classifiers. LS-SVM is one of the most popular tools for regression and classification of learning tasks. However, one of the major drawbacks of LS-SVM is the loss of sparseness, in which a great number of support vectors (SVs) are required in the model. The support vectors are typically a portion of training of samples, used to construct the decision function. Too many SVs might have a negative impact on the generalization ability of the trained model. The same holds for computational complexity scales with the increasing number of SVs. By using IHT$^\mu$-PKS, we can reduce the computation cost and achieve a more generalized and sparser LS-SVM model.

The rest of the paper is organized as follows. In Section 2, we introduce some related work on RIP-based sufficient conditions for basis pursuit, IHT and IHT-PKS. In Section 3, we present the conditions for stable recovery by the IHT$^\mu$-PKS algorithm and give a tighter RIP-based bound for guaranteed recovery via IHT-PKS. In Section 4, we discuss LS-SVM classifiers and present how the IHT$^\mu$-PKS algorithm can address the sparsity issue in LS-SVM. Numerical experiments are presented in Section 5, followed by the conclusions of this paper in Section 6.

## 2. Related work

In this section, some related work on RIP-based sufficient conditions for the basis pursuit, IHT and IHT-PKS are introduced. Table 1 gives the main notations used in the rest of this paper.

| Notation | Definition |
|---|---|
| $[n]$ | The set of natural numbers not exceeding $n$, i.e., $\{1, 2, \ldots, n\}$ |
| $S$; $S^c$ | A subset of $[n]$; the complement of a set $S$, i.e., $S^c = [n] \setminus S$ |
| $\mathrm{card}(S)$ | The cardinality of a set $S$ |
| $A_S$ | A submatrix of matrix $A$ with columns indexed by $S$ |
| $x_S$ | The vector in $\mathbb{R}^n$ equal to $x$ on $S$ and to zero on $S^c$ |
| $\mathrm{supp}(x)$ | $\{j \in [n] \colon x_j \neq 0\}$ |

Table 1. Notations used in this paper.

**2.1. Basis pursuit for sparse signal recovery.** For the optimization problem ($P_1$), lots of conditions have been investigated on the matrix $A$ for ensuring exact or approximate reconstruction of the sparse vector $x$. Foucart [12] proved that every $s$-sparse $x$ (i.e., $\|x\|_0 \leqslant s$) can be recovered from the problem if the restricted isometry constant $\delta_{2s}$ of the matrix $A$ of number $2s$ in a row is less than $\frac{3}{4+\sqrt{6}} \approx 0.4652$, where the restricted isometry constant is defined below.

**Definition 2.1** ([7])**.** The $s$th restricted isometry constant $\delta_s = \delta_s(A)$ of a matrix $A$ is the smallest $\delta \geqslant 0$ such that

$$(1-\delta)\|x\|_2^2 \leqslant \|Ax\|_2^2 \leqslant (1+\delta)\|x\|_2^2 \tag{2.1}$$

for all $s$-sparse vectors $x \in \mathbb{R}^n$. Equivalently,

$$\delta_s = \max_{S\subset[n],\mathrm{card}(S)\leqslant s} \|A_S^\top A_S - I\|_{2\to 2}, \tag{2.2}$$

where the operator norm $\|A\|_{2\to 2} := \sup\limits_{x\neq 0} \|Ax\|_2/\|x\|_2$.

If $\delta_s < 1$, $A$ is said to satisfy the restricted isometry property (RIP). The larger the restricted isometry constant, the smaller the number of measurements required for recovering sparse signal. As is shown in [14], for certain random matrix $A/\sqrt{m} \in \mathbb{R}^{m\times n}$ (such as Gaussian, sub-Gaussian, and Bernoulli random matrix), if we have

$$m \geqslant C\delta^{-2}[s(\ln(N/s)+1) + \ln(2\varepsilon^{-1})] \quad \forall\, \delta, \varepsilon \in (0,1), \quad C > 0$$

with probability at least $1-\varepsilon$, the restricted isometry constant of $A$ satisfies $\delta_s \leqslant \delta$. Hence, a great many scholars are motivated to find a tight RIP-based bound for the guaranteed performance of specific algorithms.

**2.2. Iterative hard thresholding (IHT).** In the section, we introduce the IHT algorithm, which is a thresholding-based method, used in the compressed sensing. To describe the IHT algorithm, it is necessary to introduce the hard thresholding operator first. For an arbitrary vector $v \in \mathbb{R}^n$ and a subset $\Omega$ of $[n]$, we define the operator

$$(P_\Omega(v))_i = \begin{cases} v_i & \text{if } i \in \Omega, \\ 0 & \text{otherwise.} \end{cases}$$

If $\tau$ is the support set corresponding to the positions of the top $k$ largest absolute values of the vector $v$, then we have the hard thresholding operator

$$H_k(v) := P_\tau(v).$$

The IHT algorithm is formally described as follows.

**Algorithm 1.** IHT

**Input:** The measurement matrix $A$, measurement vector $y$, sparsity level $s$.

1: Choose an initial $s$-sparse vector $x^{(0)}$, typically $x^{(0)} = \mathbf{0}$.

2: Repeat the following iteration until a stopping criterion is met:

$$x^{(t+1)} = H_s(x^{(t)} + A^\top(y - Ax^{(t)})).$$

**Output:** The $s$-sparse vector $\widehat{x}$.

Blumensath et al. [5] investigated the RIP-based bound for guaranteed performance of the IHT algorithm and derived that in the case of a given observation $y = Ax + e$ with error $e$, where $x$ is $s$-sparse, and if the restricted isometry constant of $A$ of number $3s$ in a row satisfies $\delta_{3s} < \frac{1}{\sqrt{32}}$, then at the $t$th iteration, the recovered sparse signal $x^{(t)}$ satisfies

$$\|x - x^{(t)}\|_2 \leqslant 2^{-t}\|x - x^{(0)}\|_2 + 5\|e\|_2.$$

A more generalized condition is given by [22]. Let $x \in \mathbb{R}^n$ be an $s$-sparse signal and $k \geqslant s$. If the restricted isometry constant of the measurement matrix $A$ satisfies $\delta_{2k+s} < 1/\sqrt{8v}$, where

$$v = 1 + \frac{\varrho + \sqrt{(4+\varrho)\varrho}}{2}, \quad \varrho = \frac{\min\{s, n-k\}}{k - s + \min\{s, n-k\}},$$

then at the $t$th iteration, the recovered sparse signal $x^{(t)}$ satisfies

$$\|x - x^{(t)}\|_2 \leqslant 2^{-t}\|x - x^{(0)}\|_2 + D\|e\|_2,$$

where $x^{(0)}$ is the initial point of IHT and $D$ is a constant.

Recently, a new improved RIP-based bound of the IHT algorithm was derived by Zhao and Luo [31]. They found that if the restricted isometry constant of $A$ of number $3s$ in a row satisfies $\delta_{3s} < \frac{\sqrt{5}-1}{4}$, then at the $t$th iteration, the recovered sparse signal $x^{(t)}$ satisfies

$$\|x - x^{(t)}\|_2 \leqslant 2^{-t}\|x - x^{(0)}\|_2 + D\|e\|_2.$$

This significantly improves the RIP-based bound from $\delta_{3s} < \frac{1}{\sqrt{32}}$ in [5] to $\delta_{3s} < \frac{\sqrt{5}-1}{4}$.

Some works have shown that the modified compressed sensing yields better results than the traditional compressed sensing in the reconstruction of the sparse signals if partial support of the sparse signal is known. Carrillo et al. [9] proposed a modified

iterative hard thresholding algorithm with partially known support (IHT-PKS) and proved that the sparse signal can be recovered stably, if the restricted isometry constant of the measurement matrix satisfies $\delta_{3s-2k} < \frac{1}{\sqrt{32}}$.

Assume the support set of $s$-sparse signal $x$ is partially determined, i.e., $T = \mathrm{supp}(x)$ and $T = T_0 \cup \Delta$, where $T_0 \subset \{1, 2, \ldots, n\}$ is the known part of the support set, $\Delta \subset \{1, 2, \ldots, n\}$ is the undetermined part of the support set. If the restricted isometry constant of the measurement matrix $A$ satisfies $\delta_{3s-2k} < \frac{1}{\sqrt{32}}$, where $k = |T_0|$, $\|A\|_2 < 1$ and $\|e\|_2 \leqslant \varepsilon$, then at the $t$th iteration, the recovered sparse signal $x^{(t)}$ satisfies

$$\|x - x^{(t)}\|_2 \leqslant \alpha^t \|x\|_2 + \beta\varepsilon,$$

where $\alpha = \sqrt{8}\delta_{3s-2k}$, $\beta = 2\sqrt{1+\delta_{2s-k}}(1-\alpha^t)/(1-\alpha)$.

In this paper, we improve the RIP-based bound in [9] and show that when the $(3s-2k)$th restricted isometric constant of a measurement matrix meets $\delta_{3s-2k} < \frac{\sqrt{5}-1}{4}$, the original sparse signal can be stably recovered. Note that the RIP-based bound $\delta_{3s-2k} < \frac{\sqrt{5}-1}{4}$ is tighter than $\delta_{3s-2k} < \frac{1}{\sqrt{32}}$ in [9]. This makes the recovery of sparse signal easier to implement. The details are described in Section 3.

## 3. Improved RIP bound for IHT-PKS

In this section we first introduce the IHT-PKS algorithm and then we show an improved RIP bound of IHT-PKS.

**3.1. Iterative hard thresholding with partially known support (IHT-PKS).** IHT-PKS algorithm was first proposed by Carrillo et al. [9]. It is a modified iterative hard thresholding algorithm that incorporates the known support in the recovery process. It has been verified that using the prior support information relaxes the conditions for stable reconstruction. To describe the IHT-PKS algorithm, we give the definition of another operator. For an arbitrary vector $\alpha \in \mathbb{R}^n$ and a subset $\varphi$ of $[n]$,

$$H_u^{\varphi}(\alpha) = \alpha_{\varphi} + H_u(\alpha_{\varphi^c}), \quad \text{where } u \in \mathbb{N}.$$

Assuming that the partial support of sparse signal is known a priori, the important step of the IHT-PKS algorithm is given as

$$x^{(t+1)} = H_{s-k}^{T_0}(x^{(t)} + A^{\top}(y - Ax^{(t)})),$$

where $k = |T_0|$ and $T_0$ represents the known part of the support set. The algorithm selects the first $s-k$ elements with the largest absolute value except $T_0$ at every iteration and retains the elements in $T_0$. The IHT-PKS algorithm reads as follows.

**Algorithm 2.** IHT-PKS

**Input:** The measurement matrix $A$, measurement vector $y$, sparsity level $s$, set $T_0$ of the partially known support, and $k = |T_0|$.

1: Choose an initial $s$-sparse vector $x^{(0)}$, typically $x^{(0)} = \mathbf{0}$.

2: Repeat the following iteration until a stopping criterion is met:

$$x^{(t+1)} = H_{s-k}^{T_0}(x^{(t)} + A^\top(y - Ax^{(t)})).$$

**Output:** The $s$-sparse vector $\widehat{x}$.

**Algorithm 3.** IHT$^\mu$-PKS

**Input:** The measurement matrix $A$, measurement vector $y$, sparsity level $s$, set $T_0$ of the partially known support, step size $\mu > 0$, and $k = |T_0|$.

1: Choose an initial $s$-sparse vector $x^{(0)}$, typically $x^{(0)} = \mathbf{0}$.

2: Repeat the following iteration until a stopping criterion is met:

$$x^{(t+1)} = H_{s-k}^{T_0}(x^{(t)} + \mu A^\top(y - Ax^{(t)})).$$

**Output:** The $s$-sparse vector $\widehat{x}$.

**3.2. Improved RIP-based bound.** We present a sufficient condition for the stable recovery of the IHT$^\mu$-PKS algorithm and give a tighter RIP-based bound for the guaranteed performance of the IHT-PKS algorithm in this section. The details of the IHT$^\mu$-PKS algorithm are as follows.

An inequality that will be used later is initially presented. Let $a^{(t+1)} = x^{(t)} + \mu A^\top(y - Ax^{(t)})$, $T = \mathrm{supp}(x)$, $T^{(t)} = \mathrm{supp}(x^{(t)})$ and $U^{(t)} = \mathrm{supp}(H_{s-k}(a^{(t)}_{T_0^c}))$. Put $B^{(t+1)} = T \cup T^{(t+1)} = T_0 \cup \Delta \cup U^{(t+1)}$, where $\Delta$ represents the undetermined part of the support set. Differently from the proof of Theorem 1 in [9] which shows that

$$\|x_{B^{(t+1)}} - x^{(t+1)}_{B^{(t+1)}}\|_2 \leqslant 2\|x_{B^{(t+1)}} - a^{(t+1)}_{B^{(t+1)}}\|_2,$$

when the step size $\mu = 1$. We refer to [22], however, this paper presents a new estimate

$$\|x_{B^{(t+1)}} - x^{(t+1)}_{B^{(t+1)}}\|_2 \leqslant \sqrt{\frac{3+\sqrt{5}}{2}}\|x_{B^{(t+1)}} - a^{(t+1)}_{B^{(t+1)}}\|_2,$$

where the step size $\mu > 0$, under the condition of partial support set, is known to derive an improved RIP-based bound. To facilitate the proof of the following theorem, we define a more extensive restricted isometry constant

$$\delta_{s,\mu} = \max_{S\subset[n],\mathrm{card}(S)\leqslant s} \left\| A_S^\top A_S - \frac{1}{\mu} I \right\|_{2\to 2}. \tag{3.1}$$

Notice that when $\mu = 1$, $\delta_s$ and $\delta_{s,\mu}$ are equivalent. The lemmas below also play a significant role when proving the conclusions of this paper.

**Lemma 3.1** ([22]). *Let $y \in \mathbb{R}^m$ be a vector. For all matrices $A \in \mathbb{R}^{m\times n}$ and any index set $\Gamma$ with $|\Gamma| = s$, the formula*

$$\|A_\Gamma^\top y\|_2 \leqslant \sqrt{1+\delta_s}\|y\|_2 \tag{3.2}$$

*holds.*

**Lemma 3.2** ([13]). *Let $v \in \mathbb{R}^n$ be a vector and $S$ be a subset of $[n]$. If $|S \cup \mathrm{supp}(v)| \leqslant t$ for all matrices $A \in \mathbb{R}^{m\times n}$, then the formula*

$$\|[(I - A^\top A)v]_S\|_2 \leqslant \delta_t\|v\|_2 \tag{3.3}$$

*holds.*

**Lemma 3.3.** *Let $v \in \mathbb{R}^n$ be a vector, $\mu$ be a constant, and $S$ be a subset of $[n]$. If $|S \cup \mathrm{supp}(v)| \leqslant t$ for all matrices $A \in \mathbb{R}^{m\times n}$, then the formula*

$$\Big\|\Big[\frac{1}{\mu}(I - A^\top A)v\Big]_S\Big\|_2 \leqslant \delta_{t,\mu}\|v\|_2 \tag{3.4}$$

*holds.*

P r o o f. The proof of this lemma is presented in Appendix. □

The following are the important conclusions given in this paper for guaranteeing stable the signal reconstruction using the IHT$^\mu$-PKS algorithm.

**Theorem 3.1.** *Let $x \in \mathbb{R}^n$ be an $s$-sparse vector satisfying $y = Ax + e$, where $\|e\|_2 < \varepsilon$. Put $T = \mathrm{supp}(x)$ and $T = T_0 \cup \Delta$, $|T| = s$ and $|T_0| = k$. Suppose that the restricted isometric constant of the matrix $A \in \mathbb{R}^{m\times n}$ of number $(3s - 2k, \mu)$ in a row satisfies $\delta_{3s-2k,\mu} < (\sqrt{5}-1)/(4\mu)$, then through the* IHT$^\mu$*-PKS algorithm, the error in the $t$th iteration is*

$$\|x - x^{(t)}\|_2 \leqslant \alpha^t\|x\|_2 + \beta\varepsilon, \tag{3.5}$$

*where $\alpha = \sqrt{\frac{3+\sqrt{5}}{2}}\mu\delta_{3s-2k,\mu}$, $\beta = \mu\sqrt{\frac{3+\sqrt{5}}{2}(1+\delta_{2s-k})}(1-\alpha^t)/(1-\alpha)$.*

P r o o f. The proof of this theorem is presented in Appendix. □

Remark 3.1. According to Theorem 3.1, when $\delta_{3s-2k,1} < \frac{\sqrt{5}-1}{4} \approx 0.309$, we have $\|x - x^t\|_2 \leqslant 2^{-t}\|x\|_2 + \beta\varepsilon$. Compared to the restricted isometry constant $\delta_{3s-2k} < \frac{1}{\sqrt{32}} = 0.1768$ derived in [9], the condition given in this paper is weaker and easier to realize.

Remark 3.2. According to Theorem 3.1, when $\mu = 1$ we have $\|x - x^{(t)}\|_2 \leqslant 2^{-t}\|x\|_2 + \beta\varepsilon$, where $\beta = \sqrt{\frac{3+\sqrt{5}}{2}(1+\delta_{2s-k})(1-\alpha^t)/(1-\alpha)}$. Compared to $\beta = 2\sqrt{1+\delta_{2s-k}}(1-\alpha^t)/(1-\alpha)$ given in [9], we have a tighter error bound.

Remark 3.3. The RIP-based bound for the guaranteed performance of IHT-PKS is tighter than the very recent result for IHT in [31], i.e., that $\delta_{3s} < \frac{\sqrt{5}-1}{4}$ on account of $\delta_{3s-2k} \leqslant \delta_{3s}$. It should be noted that our proof method is different from that of [31]. A smaller order of RIP reduces the number of measurements required for the approximate reconstruction. In the worst case, when the cardinality of the partially known support is zero, we have the same condition required by IHT.

In summary, we not only present a theoretical analysis about the IHT$^\mu$-PKS algorithm but also give an improved RIP-based bound of the IHT-PKS algorithm.

## 4. Application on sparse LS-SVM

In this section, we first briefly review the well known theory about the sparse LS-SVM, and then we show how to use the IHT$^\mu$-PKS algorithm to perform model training and SVs selection simultaneously.

For a binary classification problem with the training set $\{x_i, y_i\}_{i=1}^N$, $x_i \in \mathbb{R}^d$ is the $i$th input sample and $y_i \in \{1, -1\}$ is the class label. The LS-SVM is trained by addressing the optimization problem

$$\min \mathcal{J}(w,b,e) = \frac{1}{2}w^\top w + \frac{\gamma}{2}\sum_{i=1}^N e_i^2 \text{ s.t. } y_i = w^\top\varphi(x_i) + b + e_i, \quad i = 1,2,\ldots,N, \tag{4.1}$$

where $\varphi(x_i)$ is a nonlinear function which maps the input space into a higher dimensional space, $e$ is the error term, $\gamma$ is a regularization parameter. Then, we can get its Lagrangian function

$$\mathcal{L}(w,b,e;\alpha) = \mathcal{J}(w,b,e) + \sum_{i=1}^N \alpha_i[y_i - w^\top\varphi(x_i) - b - e_i],$$

where $\alpha_i$ are the Lagrange multipliers, which can be either positive or negative due to the equality constraints. The Karush-Kuhn-Tucker conditions for the above problem

are reduced to a linear system by getting rid of $w$ and $e$,

$$\left[\begin{array}{c|c} Q+\gamma^{-1}I_N & \mathbf{1}_N \\ \hline \mathbf{1}_N^\top & 0 \end{array}\right]\left[\begin{array}{c} \boldsymbol{\alpha} \\ b \end{array}\right] = \left[\begin{array}{c} \mathbf{y} \\ 0 \end{array}\right], \tag{4.2}$$

where $Q_{ij} = \varphi(x_i)^\top \varphi(x_i)$ is a kernel function, and $\mathbf{1}_N$ represents the $N$ dimensional vector whose elements are equal to 1. The most used kernel functions are the linear kernel $Q_{ij} = x_i^\top x_j$, polynomial kernel $Q_{ij} = (x_i^\top x_j + 1)^d$ where $d \geqslant 1$, and Gaussian radial basis (RBF) kernel $Q_{ij} = \exp\{-\|x_i - x_j\|^2/\sigma^2\}$.

Notice that setting a particular element of $\alpha$ to zero is equivalent to eliminating the corresponding training sample or SVs. Hence, the goal of finding a sparse solution, within a given tolerance of accuracy, can be equated to solving (4.2) by minimizing the $l_0$ norm of the vector $[\alpha^\top, b]^\top$. The training algorithms for the sparse LS-SVM using compressive sampling are investigated in [27], [21], [28]. Let $x = \left[\begin{array}{c} \boldsymbol{\alpha} \\ b \end{array}\right]$, $\psi = \left[\begin{array}{c|c} Q+\gamma^{-1}I_N & \mathbf{1}_N \\ \hline \mathbf{1}_N^\top & 0 \end{array}\right]$, and $z = \left[\begin{array}{c} \mathbf{y} \\ 0 \end{array}\right]$. The sparse LS-SVM problem can be cast as

$$\min \|x\|_0 \quad \text{s.t.} \quad \psi x = z. \tag{4.3}$$

When problem (4.3) is minimized, all $N$ samples from the training set are taken into account, which can be expensive when working with large training sets. Yang [28] further designed the sparse LS-SVM training with fewer measurements. Given a measurement matrix $\mathcal{A} \in \mathbb{R}^{M\times N}$ $(M < N)$, the measurement vector is given by $\mathcal{Y} = \mathcal{A}z$ and the optimization problem (4.3) can be expressed as

$$\min \|x\|_0 \quad \text{s.t.} \quad \mathcal{D}x = \mathcal{Y}, \tag{4.4}$$

where $\mathcal{D} = \mathcal{A}\psi$ is the dictionary. As seen, the computational complexity is now only an $M \times (N+1)$-dimensional linear system that needs to be solved. Compared to the original model, which has the size $(N+1) \times (N+1)$, this is significantly smaller.

To train the sparse LS-SVM, a variety of optimization approaches have been applied to solve the LS-SVM model. For instance, Suykens et al. first proposed pruning training samples that have the smallest absolute support values [23]. However, this method might eliminate training samples near the decision boundary, which has a negative influence on the training performance. An improved method was proposed in [18], where only a reduced training set comprised of samples near the decision boundary is used to train the LS-SVM. Yang et al. proposed an OMP algorithm to regard the support vectors as a dictionary and select the important ones that minimize the residual output error iteratively [28]. Another sparsity enhanced method was proposed in [20], where the least important SVs are removed through regularization to accelerate the training process.

In this paper, the IHT$^{\mu}$-PKS algorithm used to solve the sparse LS-SVM model does not ignore the bias term like other algorithms do. The support of the bias term in the LS-SVM is used to assist finding the important support vectors. It has also been proven in Theorem 1 that when the measurement matrix satisfies $\delta_{3s-2k,\mu} < \frac{\sqrt{5}-1}{4}$, it can stably recover the $s$-sparse signal. In consequence, it is feasible to apply the IHT$^{\mu}$-PKS algorithm to train the sparse LS-SVM model.

## 5. Numerical experiments

In this section, we verify the performance of IHT$^{\mu}$-PKS for real-world datasets. The task is to verify the advantages of the IHT$^{\mu}$-PKS algorithm in solving the sparsity issue of the LS-SVM. All the experiments were carried out in Matlab R2021a on a personal computer with 2.40 GHz Intel processor and 16.00 GB RAM.

**5.1. Data sets.** For the classification experiments, we use four data sets Ripley [26], Ionosphere[1], AlgerianForestFires[2], Madelon[3]. Each data set contains two subsets: training and testing. Table 2 gives the basic statistics of these data sets.

| Data set | Features | Size | |
|---|---|---|---|
| Ripley | 2 | Train | 250 |
| | | Test | 1000 |
| Ionosphere | 34 | Train | 202 |
| | | Test | 149 |
| AlgerianForestFires | 13 | Train | 122 |
| | | Test | 121 |
| Madelon | 500 | Train | 1000 |
| | | Test | 1600 |

Table 2. Statistics of the data sets from the UCI benchmark repository.

**5.2. Evaluation metrics.** To evaluate the effectiveness and efficiency of classification, we calculate the classification accuracy and the running time of different algorithms, where accuracy is the ratio of correct labels to recovered labels:

$$\text{accuracy} = \frac{|y_{\text{cor}}|}{|y_{\text{rec}}|} \times 100\%,$$

where $|y_{\text{cor}}|$ denotes the number of correct labels in recovered labels.

[1] `https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary.html`
[2] `http://archive.ics.uci.edu/ml/datasets.php`
[3] `https://jundongl.github.io/scikit-feature/datasets.html`

**5.3. Baseline methods.** We compare IHT$^{\mu}$-PKS with the three baselines commonly used in compressed sensing: orthogonal matching pursuit (OMP), iterative hard thresholding (IHT) and fast hard thresholding (FHTP) [13]. The stepsizes $\mu$ of $A^{\top}(y - Ax^{(t)})$ in IHT$^{\mu}$-PKS, IHT and FHTP are set to $\mu = 103/(1000\|\mathcal{D}\|_2^2)$. IHT is introduced in Section 2.2. FHTP is a fast version of the hard thresholding pursuit algorithm, where the orthogonal projection is replaced by a certain number of gradient descent iterations [13]. OMP is a greedy iterative method. In each iteration, it selects the columns of the matrix $A$ to maximize the correlation between the selected columns and the current redundant vector $y - Ax^{(t)}$; then, the relevant part is subtracted from the original signal vector. This is repeated until the number of iterations reaches the sparsity $s$. Yang et al. [28] adopted the OMP algorithm to train the LS-SVM and verified that compared to conventional training algorithm, the OMP algorithm achieves a significant improvement in terms of classification accuracy.

**5.4. Experimental results.**

**5.4.1. Classification.** We use the grid search to tune parameters $\gamma$, $d$ and $\sigma$ in the problem (4.4), where $d$ and $\sigma$ are parameters of a polynomial kernel $Q$ and RBF kernel $Q$, respectively. The Gaussian random matrix was selected as the measurement matrix and the zero vector was selected as the initial vector. For the classification experiment on the data Ripley, we employ a liner kernel function, on the Ionosphere data, we employ a polynomial kernel function, on the AlgerianForestFires and Madelon data, we use a RBF kernel function. For each data set, after we obtain the optimal parameters out of the grid search, we calculate the average accuracy and running time over 20 runs of the experiment. The experiment results of classifying the four real-world data sets are given in Tables 3, 4, 5 and 6 for the datasets Ripley, Ionosphere, AlgerianForestFires and Madelon, respectively. We particularly show the results when the sparsity $\mathcal{K}$ is 10%, 20% and 30%. The sparsity $\mathcal{K}$ is defined as

$$\mathcal{K} = \frac{K}{N+1} \times 100\%,$$

where $K$ is the number of selected SVs and $N$ is the number of training samples. Thus, a larger value of $\mathcal{K}$ means that more SVs are selected for training.

From Tables 3 to 6, we see that IHT$^{\mu}$-PKS obtains much higher accuracy in testing data sets than the baseline methods. In terms of time cost, the IHT$^{\mu}$-PKS algorithm takes less time than the other three algorithms, especially the OMP algorithm. On the Ripley data set, Ionosphere data set and Madelon data set, only when the sparsity is 10%, the time spent by the OMP algorithm is similar to the IHT and IHT$^{\mu}$-PKS

algorithms, but the time cost increases significantly with the number of selected support vectors increasing. When the sparsity $\mathcal{K}$ is 20% and 30%, the OMP algorithm takes almost ten times longer time than the IHT$^{\mu}$-PKS algorithm in this experiment.

| $\mathcal{K}$ | Data | OMP | IHT | FHTP | IHT$^{\mu}$-PKS |
|---|---|---|---|---|---|
| 10% | Train | 85.8200 | 84.9800 | 86.0000 | **86.0400** |
| | Test | 88.1000 | 89.2400 | 88.0900 | **89.2450** |
| 20% | Train | 85.8200 | 85.1200 | 86.0200 | **86.1000** |
| | Test | 88.1000 | 89.3000 | 87.4200 | **89.3100** |
| 30% | Train | 85.7600 | 85.2800 | **85.9600** | 85.8400 |
| | Test | 88.0500 | 89.1500 | 87.8200 | **89.5000** |
| 10% | Time | 0.00403 | 0.00411 | 0.00851 | **0.00397** |
| 20% | Time | 0.01320 | 0.00413 | 0.00872 | **0.00409** |
| 30% | Time | 0.03272 | 0.00412 | 0.00873 | **0.00407** |

Table 3. Classification accuracy (%) and time cost (in seconds) on the Ripley data set.

| $\mathcal{K}$ | Data | OMP | IHT | FHTP | IHT$^{\mu}$-PKS |
|---|---|---|---|---|---|
| 10% | Train | 83.9109 | 86.2376 | 84.5050 | **86.2871** |
| | Test | 93.3221 | 91.4765 | 84.832 | **95.0336** |
| 20% | Train | 85.1238 | 87.4505 | 86.287 | **87.6238** |
| | Test | 90.9732 | 94.6309 | 88.389 | **94.9664** |
| 30% | Train | 83.1683 | 88.3663 | 86.9060 | **88.3911** |
| | Test | 86.2752 | 92.7517 | 88.3560 | **93.8255** |
| 10% | Time | **0.00363** | 0.00464 | 0.00653 | 0.00399 |
| 20% | Time | 0.00789 | 0.00467 | 0.00662 | **0.00396** |
| 30% | Time | 0.01628 | 0.00471 | 0.00646 | **0.00395** |

Table 4. Classification accuracy (%) and time cost (in seconds) on the Ionosphere data set.

| $\mathcal{K}$ | Data | OMP | IHT | FHTP | IHT$^{\mu}$-PKS |
|---|---|---|---|---|---|
| 10% | Train | **92.7459** | 91.5164 | 91.7620 | 91.7623 |
| | Test | 93.1818 | 92.3554 | 93.5120 | **93.6777** |
| 20% | Train | **93.4426** | 92.4590 | 93.1150 | 92.7049 |
| | Test | 93.0579 | 91.6529 | 90.2070 | **93.7190** |
| 30% | Train | 93.6066 | 92.8279 | **94.2620** | 93.1967 |
| | Test | 91.9421 | 92.0248 | 90.7440 | **92.1901** |
| 10% | Time | **0.00101** | 0.00117 | 0.00152 | 0.00107 |
| 20% | Time | 0.00203 | 0.00112 | 0.00152 | **0.00106** |
| 30% | Time | 0.00420 | 0.00110 | 0.00152 | **0.00105** |

Table 5. Classification accuracy (%) and time cost (in seconds) on the AlgerianForestFires data set.

| $\mathcal{K}$ | Data | OMP | IHT | FHTP | IHT$^{\mu}$-PKS |
|---|---|---|---|---|---|
| 10% | Train | 55.3400 | 59.8650 | 58.3900 | **59.8800** |
| | Test | 50.0188 | 49.9750 | **50.0280** | 50.0188 |
| 20% | Train | 59.7250 | 67.0700 | 64.8150 | **67.5300** |
| | Test | 50.0250 | 50.0500 | 50.0250 | **50.0625** |
| 30% | Train | 62.8550 | 71.23 | 70.1250 | **71.6450** |
| | Test | 50.0563 | 50.0500 | 50.04375 | **50.0625** |
| 10% | Time | 0.14918 | 0.11860 | 0.55493 | **0.11399** |
| 20% | Time | 1.07840 | 0.17807 | 0.55575 | **0.11430** |
| 30% | Time | 3.49460 | 0.17787 | 0.55923 | **0.11404** |

Table 6. Classification accuracy (%) and time cost (in seconds) on the Madelon data set.

## 6. Conclusion and future work

In this paper, we give the conditions for signal stable recovery by the IHT$^{\mu}$-PKS algorithm. We also derive a tighter RIP-based bound of the IHT-PKS algorithm and prove that the recovered signal has a tighter error bound compared with existing results. The IHT$^{\mu}$-PKS algorithm is applied to solve the sparse LS-SVM problem. We provide a new insight into the solution through using the known support of the bias term and employing the IHT$^{\mu}$-PKS algorithm to train the sparse Least Squares Support Vector Machines. The experimental results show that compared with the traditional OMP, IHT and FHTP algorithms, the IHT$^{\mu}$-PKS algorithm is more effective in solving the sparse LS-SVM problem.

In the future, to verify if the upper bound $\delta_{3s-2k} < \frac{\sqrt{5}-1}{4}$ is sharp or not, we plan to construct a measurement matrix $A$ with the restricted isometry constant satisfying $\delta_{3s-2k} = \frac{\sqrt{5}-1}{4}$. If the measurement matrix $A$ is such that the IHT-PKS algorithm cannot recover $s$-sparse signal, we can say that $\delta_{3s-2k} < \frac{\sqrt{5}-1}{4}$ is a sharp bound. However, that is difficult because for a given matrix $A$ and a sparsity level $s$, computing $\delta_s$ is NP-hard.

## 7. Appendix

P r o o f of Lemma 3.3. Let $u \in \mathbb{R}^n$ be a vector and set $\zeta := \operatorname{supp}(u) \cup \operatorname{supp}(v)$, we have

$$\begin{aligned} (7.1) \qquad & \left\| \left\langle u, \left(\frac{1}{\mu} I - A^{\top} A\right) v \right\rangle \right\| \\ & = \left\| \frac{1}{\mu} \langle u, v \rangle - \langle Au, Av \rangle \right\| = \left\| \frac{1}{\mu} \langle u_{\zeta}, v_{\zeta} \rangle - \langle Au, Av \right\rangle \| \end{aligned}$$

$$= \Big\|\Big\langle u_\zeta, \Big(\frac{1}{\mu}I - A_\zeta^\top A_\zeta\Big)v_\zeta\Big\rangle\Big\| \leqslant \|u_\zeta\|_2\Big\|\Big(\frac{1}{\mu}I - A_\zeta^\top A_\zeta\Big)v_\zeta\Big\|_2$$
$$\leqslant \Big\|\Big\langle u_\zeta, \Big(\frac{1}{\mu}I - A_\zeta^\top A_\zeta\Big)v_\zeta\Big\rangle\Big\| \leqslant \|u_\zeta\|_2\Big\|\Big(\frac{1}{\mu}I - A_\zeta^\top A_\zeta\Big)\Big\|_{2\to 2}\|v_\zeta\|_2$$
$$\leqslant \|u_\zeta\|_2\delta_{t,\mu}\|v_\zeta\|_2 = \delta_{t,\mu}\|u\|_2\|v\|_2.$$

Indeed, using (7.1), we have

$$\Big\|\Big(\Big(\frac{1}{\mu}I - A^\top A\Big)v\Big)_S\Big\|_2^2 = \Big\langle\Big(\Big(\frac{1}{\mu}I - A^\top A\Big)v\Big)_S, \Big(\frac{1}{\mu}I - A^\top A\Big)v\Big\rangle \tag{7.2}$$
$$\leqslant \delta_{t,\mu}\Big\|\Big(\Big(\frac{1}{\mu}I - A^\top A\Big)v\Big)_S\Big\|_2\|v\|_2.$$

To get (3.4), the remaining step is to simplify by $\|((\frac{1}{\mu}I - A^\top A)v)_S\|_2$. □

P r o o f of Theorem 3.1. Since $|T| = s$ and $|T_0| = k$, we have $|\Delta| = s - k$. Let

$$a^{(t+1)} = x^{(t)} + \mu A^\top(y - Ax^{(t)}) = x^{(t)} + \mu(A^\top Ax - A^\top Ax^{(t)} + A^\top e). \tag{7.3}$$

At the $(t+1)$st iteration, $x^{(t+1)} = a_{T_0}^{(t+1)} + H_{s-k}(a_{T_0^c}^{(t+1)})$, the reconstruction error at the $t$th iteration is $r^{(t)} = x - x^{(t)}$. Let $T^{(t)} = \mathrm{supp}(x^{(t)})$, $U^{(t)} = \mathrm{supp}(H_{s-k}(a_{T_0^c}^{(t)}))$. For any $t$ there is $|\mathrm{supp}(a_{T_0}^{(t)})| \leqslant k$, $|U^{(t)}| \leqslant s - k$, $|T^{(t)}| \leqslant s$.

Let $B^{(t+1)} = T \cup T^{(t+1)} = T_0 \cup \Delta \cup U^{(t+1)}$, then

$$|B^{(t+1)}| \leqslant |T_0| + |\Delta| + |U^{(t+1)}| \leqslant 2s - k. \tag{7.4}$$

We first prove the formula

$$\|x_{B^{(t+1)}} - x_{B^{(t+1)}}^{(t+1)}\|_2 \leqslant \sqrt{\frac{3+\sqrt{5}}{2}}\|x_{B^{(t+1)}} - a_{B^{(t+1)}}^{(t+1)}\|_2. \tag{7.5}$$

Let $\Omega$ be the support of $x_{B^{(t+1)}}^{(t+1)}$ and $\Omega^c$ be the complement of the support of $x_{B^{(t+1)}}^{(t+1)}$. Let $\Omega'$ be the support of $x_{B^{(t+1)}}$ and put $a_1 = P_{\Omega\setminus\Omega'}(a_{B^{(t+1)}}^{(t+1)})$, $a_2 = P_{\Omega\cap\Omega'}(a_{B^{(t+1)}}^{(t+1)})$, $a_3 = P_{\Omega^c\setminus\Omega'}(a_{B^{(t+1)}}^{(t+1)})$, $a_4 = P_{\Omega^c\cap\Omega'}(a_{B^{(t+1)}}^{(t+1)})$. By the definition of $x^{(t+1)}$, $P_\Omega(x_{B^{(t+1)}}^{(t+1)}) = P_\Omega(a_{B^{(t+1)}}^{(t+1)})$. Similarly, for $x_{B^{(t+1)}}^{(t+1)}$ and $x_{B^{(t+1)}}$, define $\omega_i$ and $x_i$ $(1 \leqslant i \leqslant 4)$, respectively. We can get $\omega_1 = a_1$, $\omega_2 = a_2$, $\omega_3 = \omega_4 = x_1 = x_3 = 0$. Our aim is to find the maximum value of $\|x_{B^{(t+1)}} - x_{B^{(t+1)}}^{(t+1)}\|_2^2/\|x_{B^{(t+1)}} - a_{B^{(t+1)}}^{(t+1)}\|_2^2$, in fact,

$$\gamma = \frac{\|x_{B^{(t+1)}} - x_{B^{(t+1)}}^{(t+1)}\|_2^2}{\|x_{B^{(t+1)}} - a_{B^{(t+1)}}^{(t+1)}\|_2^2} = \frac{\|a_1\|_2^2 + \|a_2 - x_2\|_2^2 + \|x_4\|_2^2}{\|a_1\|_2^2 + \|a_2 - x_2\|_2^2 + \|a_3\|_2^2 + \|a_4 - x_4\|_2^2}. \tag{7.6}$$

This problem will be discussed on two cases. First of all, if $\|x_{B^{(t+1)}} - a_{B^{(t+1)}}^{(t+1)}\|_2 = 0$, then we have $\|a_1\|_2 = \|a_3\|_2 = 0$, $a_2 = x_2$ and $a_4 = x_4$ due to the property that

the norm is greater than or equal to 0. And further, because $\|a_1\|_0 = s - \|a_2\|_0 = s - (s - \|a_4\|_0) = \|a_4\|_0$, we can get $\|x_4\|_2 = 0$. From (7.6), we then have $\|x_{B^{(t+1)}} - x^{(t+1)}_{B^{(t+1)}}\|_2^2 = 0$ and obviously the formula (7.5) is true.

Next, if $\|x_{B^{(t+1)}} - a^{(t+1)}_{B^{(t+1)}}\|_2 \neq 0$, then $\|a_3\|_2$ must be zero as it only appears at the denominator compared to other elements, therefore

$$\gamma = \frac{\|x_{B^{(t+1)}} - x^{(t+1)}_{B^{(t+1)}}\|_2^2}{\|x_{B^{(t+1)}} - a^{(t+1)}_{B^{(t+1)}}\|_2^2} = \frac{\|a_1\|_2^2 + \|a_2 - x_2\|_2^2 + \|x_4\|_2^2}{\|a_1\|_2^2 + \|a_2 - x_2\|_2^2 + \|a_4 - x_4\|_2^2} \tag{7.7}$$

and

$$(\gamma - 1)\|a_2 - x_2\|_2^2 + \gamma\|a_4 - x_4\|_2^2 - \|x_4\|_2^2 + (\gamma - 1)\|a_1\|_2^2 = 0. \tag{7.8}$$

Since we are interested in the maximum of $\gamma$, we assume $\gamma > 1$ below. In the case of $\gamma > 1$, $a_4$ is definitely not equal to 0 from (7.7). We can also have $a_1 \neq 0$ due to $\|a_1\|_0 = \|a_4\|_0$. Fixing $(a_1, a_2, a_4)$, we can consider the above formula as a function of $(x_2, x_4)$. It can be proven that the Hessian matrix

$$\begin{bmatrix} 2(\gamma - 1) & 0 \\ 0 & 2(\gamma - 1) \end{bmatrix}$$

of this function is positive definite, so the minimum value is obtained at a stable point

$$x_2^* = a_2, \quad x_4^* = \frac{\gamma}{\gamma - 1} a_4. \tag{7.9}$$

From (7.8), we know that the minimum value of the function defined as the left-hand side of (7.8) is less than or equal to 0. Substituting (7.9) into (7.8), we get

$$\|a_1\|_2^2\gamma^2 - (2\|a_1\|_2^2 + \|a_4\|_2^2)\gamma + \|a_1\|_2^2 \leqslant 0. \tag{7.10}$$

Solving the above inequality with respect to $\gamma$, due to $a_1 \neq 0$, we obtain

$$\gamma \leqslant 1 + (2\|a_1\|_2^2)^{-1}\Big(\|a_4\|_2^2 + \sqrt{(4\|a_1\|_2^2 + \|a_4\|_2^2)\|a_4\|_2^2}\Big). \tag{7.11}$$

As $a_1$ contains the first $s$ elements of the largest absolute value in $T$ except $T_0$, we have

$$\frac{\|a_4\|_2^2}{\|a_4\|_0} \leqslant \frac{\|a_1\|_2^2}{\|a_1\|_0}. \tag{7.12}$$

Because $\|a_1\|_0 = \|a_4\|_0$, we can get $\|a_4\|_2^2 \leqslant \|a_1\|_2^2$. Substituting this into (7.11), we obtain

$$\gamma \leqslant 1 + \frac{(1 + \sqrt{5})\|a_1\|_2^2}{2\|a_1\|_2^2} = \frac{3 + \sqrt{5}}{2}. \tag{7.13}$$

Then, we have

$$\|x_{B^{(t+1)}} - x^{(t+1)}_{B^{(t+1)}}\|_2 \leqslant \sqrt{\frac{3+\sqrt{5}}{2}}\|x_{B^{(t+1)}} - a^{(t+1)}_{B^{(t+1)}}\|_2. \tag{7.14}$$

The above equation can also be expressed as

$$\|(x - x^{(t+1)})_{B^{(t+1)}}\|_2 \leqslant \sqrt{\frac{3+\sqrt{5}}{2}}\|(x - a^{(t+1)})_{B^{(t+1)}}\|_2. \tag{7.15}$$

Given that $a^{(t+1)} = x^{(t)} + \mu A^\top Ax - \mu A^\top Ax^{(t)} + \mu A^\top e$, substituting this into (7.15), we have

(7.16)

$$\begin{aligned}\|(x - x^{(t+1)})_{B^{(t+1)}}\|_2 &\leqslant \sqrt{\frac{3+\sqrt{5}}{2}}\|(x - x^{(t)} - \mu A^\top Ax + \mu A^\top Ax^{(t)} - \mu A^\top e)_{B^{T+1}}\|_2 \\ &\leqslant \sqrt{\frac{3+\sqrt{5}}{2}}\|[(I - \mu A^\top A)(x - x^{(t)})]_{B^{(t+1)}}\|_2 \\ &\quad + \mu\sqrt{\frac{3+\sqrt{5}}{2}}\|(A^\top e)_{B^{(t+1)}}\|_2.\end{aligned}$$

By using Lemma 3.1 and Lemma 3.3 together with the fact that $|B^{(t+1)}| \leqslant 2s - k$ and $|B^{(t)} \cup B^{(t+1)}| = |T_0 \cup \Delta \cup U^{(t+1)} \cup U^{(t)}| \leqslant 3s - 2k$, we get

$$\begin{aligned}\|(x - x^{(t+1)})_{B^{(t+1)}}\|_2 &\leqslant \sqrt{\frac{3+\sqrt{5}}{2}}\mu\delta_{3s-2k,\mu}\|x - x^{(t)}\|_2 \\ &\quad + \sqrt{\frac{3+\sqrt{5}}{2}}\mu\|(A^\top e)_{B^{(t+1)}}\|_2 \\ &\leqslant \sqrt{\frac{3+\sqrt{5}}{2}}\mu\delta_{3s-2k,\mu}\|x - x^{(t)}\|_2 + \eta\|e\|_2,\end{aligned} \tag{7.17}$$

where $\eta = \mu\sqrt{\frac{3+\sqrt{5}}{2}(1 + \delta_{2s-k})}$. Hence, we obtain

$$\|x - x^{(t+1)}\|_2 \leqslant \sqrt{\frac{3+\sqrt{5}}{2}}\mu\delta_{3s-2k,\mu}\|x - x^{(t)}\|_2 + \eta\|e\|_2. \tag{7.18}$$

Putting $\alpha = \sqrt{\frac{3+\sqrt{5}}{2}}\mu\delta_{3s-2k,\mu}$ and assuming $x^0 = 0$, we have

$$\|x - x^{(t)}\|_2 \leqslant \alpha^t\|x\|_2 + \eta\|e\|_2\sum_{j=0}^{t-1}\alpha^j, \tag{7.19}$$

which can be rewritten as

$$\|x - x^{(t)}\|_2 \leqslant \alpha^t \|x\|_2 + \beta\varepsilon, \tag{7.20}$$

where $\beta = \mu\sqrt{\frac{3+\sqrt{5}}{2}(1+\delta_{2s-k})}(1-\alpha^t)/(1-\alpha)$.

We see that if $\alpha = \sqrt{\frac{3+\sqrt{5}}{2}}\mu\delta_{3s-2k,\mu} < 1$, the algorithm is guaranteed to converge. In order to achieve a quick converge, we can let $\sqrt{\frac{3+\sqrt{5}}{2}}\mu\delta_{3s-2k,\mu} < \frac{1}{2}$, this completes the proof of Theorem 3.1. □

*References*


[1] *K. Axiotis, M. Sviridenko*: Sparse convex optimization via adaptively regularized hard thresholding. J. Mach. Learn. Res. *22* (2021), Article ID 122, 47 pages. zbl MR

[2] *R. G. Baraniuk, V. Cevher, M. F. Duarte, C. Hegde*: Model-based compressive sensing. IEEE Trans. Inf. Theory *56* (2010), 1982–2001. zbl MR doi

[3] *A. Besson, D. Perdios, Y. Wiaux, J.-P. Thiran*: Joint sparsity with partially known support and application to ultrasound imaging. IEEE Signal Process. Lett. *26* (2019), 84–88. doi

[4] *T. Blumensath, M. E. Davies*: Iterative thresholding for sparse approximations. J. Fourier Anal. Appl. *14* (2008), 629–654. zbl MR doi

[5] *T. Blumensath, M. E. Davies*: Iterative hard thresholding for compressed sensing. Appl. Comput. Harmon. Anal. *27* (2009), 265–274. zbl MR doi

[6] *T. Blumensath, M. E. Davies*: Normalized iterative hard thresholding: Guaranteed stability and performance. IEEE J. Selected Topics Signal Process. *4* (2010), 298–309. doi

[7] *E. J. Candès, T. Tao*: Decoding by linear programming. IEEE Trans. Inf. Theory *51* (2005), 4203–4215. zbl MR doi

[8] *R. E. Carrillo, L. F. Polania, K. E. Barner*: Iterative algorithms for compressed sensing with partially known support. IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, Los Alamitos, 2010, pp. 3654–3657. doi

[9] *R. E. Carrillo, L. F. Polania, K. E. Barner*: Iterative hard thresholding for compressed sensing with partially known support. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, Los Alamitos, 2011, pp. 4028–4031. doi

[10] *S. S. Chen, D. L. Donoho, M. A. Saunders*: Atomic decomposition by basis pursuit. SIAM Rev. *43* (2001), 129–159. zbl MR doi

[11] *K. Cui, Z. Song, N. Han*: Fast thresholding algorithms with feedbacks and partially known support for compressed sensing. Asia-Pac. J. Oper. Res. *37* (2020), Article ID 2050013, 20 pages. zbl MR doi

[12] *S. Foucart*: A note on guaranteed sparse recovery via $l_1$-minimization. Appl. Comput. Harmon. Anal. *29* (2010), 97–103. zbl MR doi

[13] *S. Foucart*: Hard thresholding pursuit: An algorithm for compressive sensing. SIAM J. Numer. Anal. *49* (2011), 2543–2563. zbl MR doi

[14] *S. Foucart, H. Rauhut*: A Mathematical Introduction to Compressive Sensing. Applied and Numerical Harmonic Analysis. Birkhäuser, New York, 2013. zbl MR doi

[15] *L. Jacques*: A short note on compressed sensing with partially known signal support. Signal Process. *90* (2010), 3308–3312. zbl doi

[16] *M. A. Khajehnejad, W. Xu, A. S. Avestimehr, B. Hassibi*: Weighted $l_1$ minimization for sparse recovery with prior information. IEEE International Symposium on Information Theory. IEEE, Los Alamitos, 2009, pp. 483–487. doi

[17] *S. Khera, S. Singh*: Estimation of channel for millimeter-wave hybrid massive MIMO systems using Orthogonal Matching Pursuit (OMP). J. Phys., Conf. Ser. *2327* (2022), Article ID 012040, 9 pages. doi
[18] *Y. Li, C. Lin, W. Zhang*: Improved sparse least-squares support vector machine classifiers. Neurocomputing *69* (2006), 1655–1658. doi
[19] *H. Liu, R. Foygel Barber*: Between hard and soft thresholding: Optimal iterative thresholding algorithms. Inf. Inference *9* (2020), 899–933. zbl MR doi
[20] *R. Mall, J. A. K. Suykens*: Very sparse LSSVM reductions for large-scale data. IEEE Trans. Neural Netw. Learn. Syst. *26* (2015), 1086–1097. MR doi
[21] *Y.-H. Shao, C.-N. Li, M.-Z. Liu, Z. Wang, N.-Y. Deng*: Sparse $L_q$-norm least squares support vector machine with feature selection. Pattern Recognition *78* (2018), 167–181. doi
[22] *J. Shen, P. Li*: A tight bound of hard thresholding. J. Mach. Learn. Res. *18* (2018), Article ID 208, 42 pages. zbl MR
[23] *J. A. K. Suykens, L. Lukas, J. Vandewalle*: Sparse approximation using least squares support vector machines. IEEE International Symposium on Circuits and Systems (ISCAS). Vol. 2. IEEE, Los Alamitos, 2000, pp. 757–760. doi
[24] *N. Vaswani, W. Lu*: Modified-CS: Modifying compressive sensing for problems with partially known support. IEEE Trans. Signal Process. *58* (2010), 4595–4607. zbl MR doi
[25] *R. von Borries, C. J. Miosso, C. Potes*: Compressed sensing using prior information. International Workshop on Computational Advances in Multi-Sensor Adaptive Processing. IEEE, Los Alamitos, 2007, pp. 121–124. doi
[26] *X.-Z. Wang, H.-J. Xing, Y. Li, Q. Hua, C.-R. Dong, W. Pedrycz*: A study on relationship between generalization abilities and fuzziness of base classifiers in ensemble learning. IEEE Trans. Fuzzy Syst. *23* (2015), 1638–1654. doi
[27] *J. Yang, A. Bouzerdoum, S. L. Phung*: A training algorithm for sparse LS-SVM using compressive sampling. IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, Los Alamitos, 2010, pp. 2054–2057. doi
[28] *J. Yang, J. Ma*: A sparsity-based training algorithm for Least Squares SVM. IEEE Symposium on Computational Intelligence and Data Mining (CIDM). IEEE, Los Alamitos, 2014, pp. 345–350. doi
[29] *M. Yu, V. Gupta, M. Kolar*: Recovery of simultaneous low rank and two-way sparse coefficient matrices, a nonconvex approach. Electron. J. Stat. *14* (2020), 413–457. zbl MR doi
[30] *X. Zhang, W. Xu, Y. Cui, L. Lu, J. Lin*: On recovery of block sparse signals via block compressive sampling matching pursuit. IEEE Access *7* (2019), 175554–175563. doi
[31] *Y.-B. Zhao, Z.-Q. Luo*: Improved rip-based bounds for guaranteed performance of two compressed sensing algorithms. Available at `https://arxiv.org/abs/2007.01451` (2020), 10 pages.

*Authors' addresses*: *Jinyao Ma*, *Haibin Zhang*, *Shanshan Yang*, Beijing Institute for Scientific and Engineering Computing, Faculty of Science, Beijing University of Technology, No. 100 Ping Le Yuan, Chaoyang District, Beijing 100124, P. R. China, e-mail: `17377341661@163.com`, `zhanghaibin@bjut.edu.cn`, `17865814707@163.com`; *Jiaojiao Jiang* (corresponding author), School of Computer Science and Engineering, University of New South Wales, High St, Sydney NSW 2052, Australia, e-mail: `jiaojiao.jiang@unsw.edu.au`.