Filomat 39:15 (2025), 5109–5120
https://doi.org/10.2298/FIL2515109X

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

Convergence analysis for support recovery with quasi-Newton hard thresholding-based algorithm

Lie-Jun Xie[a,*], **Jin-Ping Wang**[a,b]

[a] *School of Mathematics and Statistics, Ningbo University, Ningbo, Zhejiang, China*
[b] *School of Computer Science, Guangzhou College of Applied Science and Technology, Guangzhou, China*

Abstract. In compressed sensing, designing suitable algorithms for recovering sparse signals from an under-determined linear model is one of the important issues. Among these recovery algorithms, hard thresholding-based ones have attracted great attention in recent years. In this work, we propose a novel variant of hard thresholding-based algorithms called the quasi-Newton hard thresholding pursuit (QNHTP) algorithm by adopting the quasi-Newton search direction. We establish sufficient condition for support recovery guarantee in terms of the restricted isometry constant of the sensing matrix. In addition, we present a range of selectable stepsize parameters for applying the QNHTP algorithm to sparsity optimization problems that arise in compressive sensing. We demonstrate that by taking the stepsize parameter with a fixed constant of one, the optimal upper bound of restricted isometry constant can be achieved.

1. Introduction

Compressed sensing (CS) was first introduced by Donoho, Candès and Tao [10, 11, 19]. It breaks through the limitation of traditional Nyquist-Shannon sampling theory in signal processing. Its theories and applications have been extensively investigated over the past few decades. In the field of CS, a sparse or approximately sparse signal $\mathbf{x} \in \mathbb{R}^N$ can be sampled with a linear sampling operator $\mathbf{A} \in \mathbb{R}^{m\times N}(m \ll N)$. The sampling model can be formulated as an inverse problem with the form of

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

where $\mathbf{e} \in \mathbb{R}^m$ models the possible observation noise which equals the zero vector only in an idealized setting but for which a bound $\|\mathbf{e}\|_2 \le \epsilon$ is typically available, and $\mathbf{y} \in \mathbb{R}^m$ is the observed sample. The

2020 *Mathematics Subject Classification*. Primary 94A12; Secondary 65F10, 65J20, 15A29.
Keywords. Compressed sensing; sparse signal recovery; hard thresholding pursuit; quasi-Newton iterative method; restricted isometry constant.
Received: 17 May 2023; Revised: 30 March 2025; Accepted: 13 July 2025
Communicated by Predrag Stanimirović
Research supported by the National Natural Science Foundation of China under Grant 62071262 and the K.C. Wong Magna Fund in Ningbo University.
* Corresponding author: Lie-Jun Xie
Email addresses: xieliejun@nbu.edu.cn (Lie-Jun Xie), wangjinping@nbu.edu.cn (Jin-Ping Wang)
ORCID iDs: https://orcid.org/0000-0003-0482-1650 (Lie-Jun Xie), https://orcid.org/0000-0002-8569-830X (Jin-Ping Wang)

problem (1) is usually reformulated as an alternative sparsity constrained optimization problem, which can be expressed as follows:

$$\min_{\mathbf{x}\in\mathbb{R}^N} \mathbf{f}(\mathbf{x}) = \frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 \quad \text{subject to} \quad \|\mathbf{x}\|_0 \le s, \tag{2}$$

where $\|\cdot\|_2$ represents the Euclidean norm, $\|\mathbf{x}\|_0$ denotes the ℓ_0 norm of $\mathbf{x}$ that counts the number of non-zero entries in $\mathbf{x}$, and s is a given integer indicating the sparsity level of $\mathbf{x}$ (i.e., $\mathbf{x}$ is an s-sparse signal).

Unfortunately, the combinatorial nature of the sparsity constraint [27] makes (2) a computational NP-hard problem. Therefore, designing a suitable algorithm to recover the sparse signal $\mathbf{x}$ from the model (2) is an important topic in compressed sensing. In the past few decades, various sparse recovery algorithms have been proposed to address this issue. For instance, optimization methods such as ℓ_1-minimization [15], ℓ_p-minimization ($0 < p < 1$) [13, 14], ℓ_{1-2}-minimization[22, 40], and ℓ_{p-q}-minimization ($0 < p \le 1, 0 < q \le 2$) [29]; greedy methods including orthogonal matching pursuit (OMP) [31], compressive sampling matching pursuit (CoSaMP) [28], and subspace pursuit (SP) [16]; thresholding-based methods like hard thresholding algorithms [3–8, 23, 37–39], soft thresholding algorithm [17, 18, 21], firm thresholding algorithm [34], and optimal thresholding algorithm [41, 42]. Among these recovery algorithms, the hard thresholding ones have attracted great attention in recent years due to their simple structure and easy implementation. This family of algorithms aims to solve the original problem (2) by decreasing the objective function $\mathbf{f}(\mathbf{x})$ along a descent direction, with the goal of iteratively identifying the support of sparse signal $\mathbf{x}$ for recovery. Such algorithms can be roughly divided into two categories, which are termed as the iterative hard thresholding-type (IHT-type) and hard thresholding pursuit-type (HTP-type) algorithms in this paper. The IHT-type algorithm performs the iterative scheme

$$\mathbf{x}_{k+1} = \mathcal{H}_s(\mathbf{x}_k + \mu\mathbf{d}_k). \tag{3}$$

Adding a pursuit step to the IHT-type algorithm (3) results in an HTP-type algorithm where the iterative scheme is as follows:

$$\overline{\mathbf{x}_k} = \mathcal{H}_s(\mathbf{x}_k + \mu\mathbf{d}_k), \tag{4a}$$

$$\mathbf{x}_{k+1} = \arg\min_{\mathbf{x}\in\mathbb{R}^N} \left\{\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 : \operatorname{supp}(\mathbf{x}) \subseteq \operatorname{supp}(\overline{\mathbf{x}_k})\right\}. \tag{4b}$$

In (3) and (4a), $\mathcal{H}_s(\cdot)$ is the hard thresholding operator that retains only the s largest entries (in magnitude) of a signal, setting all others to zero. μ represents the stepsize, which can be fixed or updated iteratively, and $\mathbf{d}_k$ denotes the search direction. Notation $\operatorname{supp}(\mathbf{x})$ in (4b) denotes the support of $\mathbf{x}$. Different stepsize μ and search direction $\mathbf{d}_k$ have been adopted to develop a variety of effective IHT-type and HTP-type algorithms.

The iterative hard thresholding (IHT) algorithm was initially proposed by Blumensath in [5] to address the ℓ_0 regularized optimisation problem. It was formally introduced into compressed sensing for reconstructing sparse signals in [6], and represents the first IHT-type algorithm. By utilizing a unit stepsize and the steepest gradient method, IHT iteratively seeks solutions with

$$\mu = 1 \text{ and } \mathbf{d}_k = -\nabla\mathbf{f}(\mathbf{x}_k) = \mathbf{A}^T(\mathbf{y}-\mathbf{A}\mathbf{x}_k).$$

Although powerful, the performance of IHT heavily depends on the selection of stepsize [20]. A larger stepsize may cause divergence, while a smaller one may result in slow convergence or convergence to a local minima instead of the desired global minima [12]. In this regard, the normalized iterative hard thresholding (NIHT) algorithm [7] was developed by adopting the same search direction $\mathbf{d}_k$ as IHT and an adaptive stepsize

$$\mu = \frac{\|(-\nabla\mathbf{f}(\mathbf{x}_k))_{S_k}\|_2^2}{\|\mathbf{A}((-\nabla\mathbf{f}(\mathbf{x}_k))_{S_k})\|_2^2}, \tag{5}$$

where S_k is the support of $\mathbf{x}_k$. However, NIHT suffers from the slow asymptotic convergence rate of the steepest descent method [1]. One strategy to overcome this obstacle is to include the conjugate gradient method. To achieve this, the conjugate gradient iterative hard thresholding (CGIHT) algorithm [1, 2] was proposed by adopting an adaptive stepsize given in (5) and adjusting the search direction as

$$\mathbf{d}_0 = -\nabla\mathbf{f}(\mathbf{x}_0), \mathbf{d}_{k+1} = -\nabla\mathbf{f}(\mathbf{x}_{k+1}) + \alpha_k\mathbf{d}_k.$$

Recently, a new variant, named the heavy-ball-based hard thresholding (HBHT) algorithm [33], was successfully developed by using the unit stepsize and the new search direction

$$\mathbf{d}_k = \beta(\mathbf{x}_k - \mathbf{x}_{k-1}) - \alpha\nabla\mathbf{f}(\mathbf{x}_k) \tag{6}$$

with two suitable parameters $\alpha > 0$ and $\beta \geq 0$. More recently, Jin and Xie [24] proposed a momentum-based iterative hard thresholding (MIHT) algorithm by using the unit stepsize and introducing a new iterative search direction

$$\mathbf{d}_k = -\sum_{i=0}^{k}\mu\eta^i\nabla\mathbf{f}(\mathbf{x}_{k-i}) + \eta^{k+1}\mathbf{v}_0,$$

where $\mathbf{v}_0$ is an initial vector, and two parameters $\mu > 0$ and $0 < \eta < 1$ were restricted by the convergence condition of the MIHT algorithm. The new direction used in this variant was derived from the momentum method, which uses historical iteration information to refine the search direction and thereby accelerate convergence.

Another popular alternative is to add a pursuit step, as given in (4b), to the IHT-type algorithms, resulting in HTP-type algorithms. The first HTP-type algorithm is the hard thresholding pursuit (HTP) algorithm, developed by Foucart in [23], which uses an unit stepsize and negative gradient direction. After that, two improved HTP-type algorithms named HTP^{μ} and normalized HTP (NHTP) were also discussed in [8] to enhance the recovery performance of HTP. They adopt a constant stepsize and an adaptive stepsize with the form of (5), respectively. In [33], Sun et al. also introduced the heavy-ball-based hard thresholding pursuit (HBHTP) algorithm with a unit stepsize and the search direction shown in (6).

It is worth mentioning that the search direction of the above algorithms all adopts the gradient-based direction of the objective function $\mathbf{f}(\mathbf{x})$ in problem (2). However, as shown in classic optimization theory [9, 30], Newton-based methods are generally more efficient than gradient-based ones for solving nonlinear optimization problems. The main difficulty for Newton-based methods to deal with the problem (2) is the singularity of Hessian $\nabla^2\mathbf{f}(\mathbf{x}) = \mathbf{A}^T\mathbf{A}$ due to $m \ll N$. To overcome this obstacle, Meng and Zhao [26] modified the Hessian by introducing a suitable parameter $\epsilon > 0$ such that $\mathbf{A}^T\mathbf{A} + \epsilon\mathbf{I}$ is positive definite. They successfully proposed two new IHT-type and HTP-type algorithms, named the Newton-step-based iterative hard thresholding (NSIHT) and the Newton-step-based hard thresholding pursuit (NSHTP), respectively. The stepsizes in NSIHT and NSHTP are closely related to the singular values of sensing matrix $\mathbf{A}$, the given parameter ϵ, and the restricted isometry constant in theoretical analysis. Both algorithms adopt the search direction by

$$\mathbf{d}_k = (\mathbf{A}^T\mathbf{A} + \epsilon\mathbf{I})^{-1}\mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{x}_k). \tag{7}$$

In [44], the authors also considered the Newton-related HTP-type algorithm and rigorously established its global and quadratic convergence from an optimization perspective. Recently, Wen et al. [36] proposed the pseudo-inverse-based hard thresholding (PHT) algorithm by utilizing

$$\mathbf{d}_k = -(\nabla^2\mathbf{f}(\mathbf{x}_k))^{\dagger}\nabla\mathbf{f}(\mathbf{x}_k)$$

in (3) to adjust search direction and improve performance in sparse recovery.

Obviously, calculating the inverse of the corrected Hessian matrix is necessary for NSIHT, NSHTP and PHT, which incurs additional computational cost. However, instead of directly computing the Hessian and its inverse, quasi-Newton methods aim to introduce an approximation to the inverse Hessian in place of the true inverse. Based on this consideration, Jing et al. [25] utilized the approximation $2\mathbf{I} - \mathbf{A}^T\mathbf{A}$ for the inverse Hessian $(\nabla^2 \mathbf{f}(\mathbf{x}))^{-1} = (\mathbf{A}^T\mathbf{A})^{-1}$, which can be easily derived from the convergence of Neumann series $\sum_{i=0}^{\infty}(\mathbf{I}-\mathbf{A}^T\mathbf{A})^i$ under the condition $\|\mathbf{A}\|_2 < 1$. They developed a new IHT-type algorithm called quasi-Newton iterative projection (QNIP). The adaptive stepsize μ and search direction $\mathbf{d}_k$ in QNIP are calculated as follows:

$$\mu = \frac{\mathbf{d}_k^T \nabla \mathbf{f}(\mathbf{x}_k)}{\|\mathbf{A}\mathbf{d}_k\|_2^2} \text{ and } \mathbf{d}_k = -(2\mathbf{I}-\mathbf{A}^T\mathbf{A})\nabla \mathbf{f}(\mathbf{x}_k). \tag{8}$$

After that, Wang and Qu [35] proposed a novel HTP-type algorithm called quasi-Newton projection pursuit (QNPP) by adding a pursuit step to QNIP. The stepsize in QNPP is updated iteratively as given in (5), and the search direction is determined by

$$\mathbf{d}_k = \mathbf{A}^T(\mathbf{y}-\mathbf{A}\mathbf{x}_k) + (\mathbf{I}-\mathbf{A}^T\mathbf{A})\mathcal{H}_s\Big(\mathbf{A}^T(\mathbf{y}-\mathbf{A}\mathbf{x}_k)\Big). \tag{9}$$

Inspired by these works, we continue to study the performance of HTP-type algorithm (4) with arbitrary positive constant stepsize $\mu \in \mathbb{R}^+$ and quasi-Newton-based search direction given in (8). Specifically, for a given initial signal $\mathbf{x}_0 \in \mathbb{R}^N$, we consider a new variant of the HTP-type algorithms with iterative scheme

$$\overline{\mathbf{x}_k} = \mathcal{H}_s\Big(\mathbf{x}_k + \mu(2\mathbf{I}-\mathbf{A}^T\mathbf{A})\mathbf{A}^T(\mathbf{y}-\mathbf{A}\mathbf{x}_k)\Big), \tag{10a}$$

$$\mathbf{x}_{k+1} = \arg\min_{\mathbf{x}\in\mathbb{R}^n} \Big\{\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 : \text{supp}(\mathbf{x}) \subseteq \text{supp}(\overline{\mathbf{x}_k})\Big\}. \tag{10b}$$

The main findings of this study can be summarized as follows:

• We establish the sufficient condition for support recovery guarantee and the solution error bound in terms of the restricted isometry constant of sensing matrix $\mathbf{A}$ while applying QNHTP algorithm (10) to problem (2). (see Theorem 3.1)

Unlike the sufficient conditions in existing works, the condition presented in this work takes the form of $\delta_{3s} < \varphi(\mu)$, where μ is an arbitrary stepsize parameter. With this condition, we deduce a rang of selectable stepsize $\mu \in \Big((3-\sqrt{5})/2, (1+\sqrt{5})/2\Big)$ in (10a) to guarantee the convergence of QNHTP algorithm. Moreover, we show that adopting unit stepsize yields optimal upper bound for $\varphi(\mu)$.

• We demonstrate that the iterative sequence $\{\mathbf{x}_k\}$ generated by the QNHTP algorithm (10) in an idealized setting converges to an s-sparse signal $\mathbf{x}$ at a geometric rate. Additionally, we prove that convergence is achieved within a finite number of iterations. (see Corollary 3.2)

The rest of the paper is organized as follows. Section 2 introduces the notations and some useful technical lemmas that are used to prove our results. Section 3 presents the theoretical results and compares our work with other existing ones. Section 4 provides a brief conclusion.

2. Preliminaries

2.1. Notations

Throughout this paper, we denote vectors by lowercase bold letters and matrices by uppercase bold letters. Let $\mathbb{R}^N$ be the N-dimension Euclidean space, and $\mathbb{R}^{m\times N}$ be the set of $m \times N$ real matrices. For $\mathbf{x} \in \mathbb{R}^N$ and $\mathbf{A} \in \mathbb{R}^{m\times N}$, $\|\mathbf{x}\|_2$ and $\|\mathbf{A}\|_2$ represent the Euclidean norm of $\mathbf{x}$ and the spectral norm of $\mathbf{A}$, respectively. $\mathbf{A}^T$ denotes the transform of matrix $\mathbf{A}$. supp($\mathbf{x}$) indicates the support of $\mathbf{x}$, namely, the set of indices of nonzero entries of $\mathbf{x}$. $\mathbf{I}$ and $\mathbf{0}$ refer to an identity matrix and a zero matrix or vector, respectively. Let $|S|$ be the cardinality of an index set $S \subseteq [N] = \{1, 2, \cdots, N\}$, and let $\overline{S}$ be its complement. Let $\mathbf{A}_S \in \mathbb{R}^{m\times|S|}$ be the sub-matrix of $\mathbf{A}$ that only contains those columns indexed by S, and $\mathbf{x}_S$ be the sub-vector of $\mathbf{x}$ that only contains those columns indexed by S. Let Σ_s be the set of all s-sparse vectors, namely, $\Sigma_s = \{\mathbf{x} \in \mathbb{R}^N, \|\mathbf{x}\|_0 \le s\}$.

2.2. Technical lemmas

In this subsection, we introduce some useful technical lemmas that will be frequently used in our proofs. We begin with the well-known restricted isometry constant (RIC), which is a quantity commonly used to measure the suitability of the sensing matrix $\mathbf{A}$. Its definition is as follows:

Definition 2.1. *([11]). The s-th RIC $\delta_s = \delta_s(\mathbf{A})$ of sensing matrix $\mathbf{A} \in \mathbb{R}^{m\times N}$ is the smallest $\delta_s \in (0,1)$ such that*

$$(1-\delta_s)\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1+\delta_s)\|\mathbf{x}\|_2^2$$

for all $\mathbf{x} \in \Sigma_s$. Equivalently, it is given by $\delta_s = \max_{S\subset[N],|S|\leq s} \left\|\mathbf{A}_S^T\mathbf{A}_S - \mathbf{I}\right\|_2$.

In the following, we will recall four useful technical results.

Lemma 2.2. ([23]). *Given any vector $\mathbf{v} \in \mathbb{R}^N$ and any index set $S \subset [N]$, if $|\mathrm{supp}(\mathbf{v}) \cup S| \leq t$ and $\mathbf{A}$ satisfies the RIC of order t, then*

$$\left\|\left((\mathbf{I}-\mathbf{A}^T\mathbf{A})\mathbf{v}\right)_S\right\|_2 \leq \delta_t\|\mathbf{v}\|_2.$$

Lemma 2.3. ([23]). *Given any vector $\mathbf{e} \in \mathbb{R}^m$ and an index set $S \subset [N]$ with $|S| \leq s$, if matrix $\mathbf{A}$ satisfies the RIC of order s, then*

$$\|(\mathbf{A}^T\mathbf{e})_S\|_2 \leq \sqrt{1+\delta_s}\|\mathbf{e}\|_2.$$

Lemma 2.4. ([41]). *Suppose $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ where $\mathbf{x} \in \mathbb{R}^N$ is an s-sparse vector, $\mathbf{e} \in \mathbb{R}^m$ is a possible observation noise, and $\mathbf{A} \in \mathbb{R}^{m\times N}$ is a sensing matrix satisfying the RIC of order s. For any s-sparse vector $\mathbf{v} \in \mathbb{R}^N$, if $\mathbf{x}^\natural = \arg\min_{\mathbf{z}\in\mathbb{R}^N}\{\|\mathbf{y}-\mathbf{A}\mathbf{z}\|_2^2 : \mathrm{supp}(\mathbf{z}) \subseteq \mathrm{supp}(\mathbf{v})\}$, then*

$$\|\mathbf{x}-\mathbf{x}^\natural\|_2 \leq \frac{1}{\sqrt{1-\delta_{2s}^2}}\|\mathbf{x}-\mathbf{v}\|_2 + \frac{\sqrt{1+\delta_s}}{1-\delta_{2s}}\|\mathbf{e}\|_2.$$

Lemma 2.5. ([43]). *For any vector $\mathbf{v} \in \mathbb{R}^N$ and any s-sparse vector $\mathbf{x} \in \mathbb{R}^N$, one has*

$$\|\mathcal{H}_s(\mathbf{v}) - \mathbf{x}\|_2 \leq \frac{\sqrt{5}+1}{2}\|(\mathbf{v}-\mathbf{x})_{S\cup S^*}\|_2,$$

where $S = \mathrm{supp}(\mathbf{x})$ and $S^* = \mathrm{supp}(\mathcal{H}_s(\mathbf{v}))$.

Remark 2.6. This is an important property of the hard thresholding operator $\mathcal{H}_s$. As pointed out in [43], this result can be easily deduced from the proof of [32, Theorem 1]. Additionally, the constant $(\sqrt{5}+1)/2$ in Lemma 2.4 has been replaced by $\sqrt{3}$ in [26], which dates back to the statement in [23].

We also require two important lemmas that play a crucial role in proving our main results.

Lemma 2.7. *Given any vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^N$, if $|\mathrm{supp}(\mathbf{u})\cup\mathrm{supp}(\mathbf{v})| \leq t$ and $\mathbf{A}$ satisfies the RIC of order t and $\|\mathbf{A}\|_2 < 1$, then*

$$|\langle\mathbf{u}, (\mathbf{I}-\mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{A}\mathbf{v}\rangle| \leq \delta_t\|\mathbf{u}\|_2\|\mathbf{v}\|_2. \tag{11}$$

Furthermore, for any index set $\Omega \subset [N]$, if $|\Omega \cup \mathrm{supp}(\mathbf{v})| \leq t$, then

$$\left\|\left((\mathbf{I}-\mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{A}\mathbf{v}\right)_\Omega\right\|_2 \leq \delta_t\|\mathbf{v}\|_2. \tag{12}$$

Proof. Let $S = \text{supp}(\mathbf{u}) \cup \text{supp}(\mathbf{v})$, it reads

$$\begin{aligned}
\left|\langle \mathbf{u}, (\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{A}\mathbf{v}\rangle\right| &= \left|\langle \mathbf{A}\mathbf{u}, \mathbf{A}\mathbf{v}\rangle - \langle \mathbf{A}\mathbf{u}, \mathbf{A}\mathbf{A}^T\mathbf{A}\mathbf{v}\rangle\right| \\
&= \left|\langle \mathbf{A}_S\mathbf{u}, \mathbf{A}\mathbf{v}\rangle - \langle \mathbf{A}_S\mathbf{u}, \mathbf{A}_S\mathbf{A}_S^T\mathbf{A}\mathbf{v}\rangle\right| \\
&= \left|\langle \mathbf{A}_S\mathbf{u}, (\mathbf{I} - \mathbf{A}_S\mathbf{A}_S^T)\mathbf{A}\mathbf{v}\rangle\right| \\
&\le \|\mathbf{A}_S\mathbf{u}\|_2 \cdot \|(\mathbf{I} - \mathbf{A}_S\mathbf{A}_S^T)\mathbf{A}\mathbf{v}\|_2 \\
&\le \|\mathbf{A}_S\|_2 \cdot \|\mathbf{u}\|_2 \cdot \|\mathbf{I} - \mathbf{A}_S\mathbf{A}_S^T\|_2 \cdot \|\mathbf{A}\|_2 \cdot \|\mathbf{v}\|_2 \\
&\le \delta_t \|\mathbf{u}\|_2 \|\mathbf{v}\|_2,
\end{aligned}$$

where we use Definition 2.1 and the fact that spectral norm of a sub-matrix is not beyond the norm of entire matrix, i.e., $\|\mathbf{A}_S\|_2 \le \|\mathbf{A}\|_2$ in the last inequality. Thus, we finish the proof of (11).

Moreover, we notice that

$$\left\|\left((\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{A}\mathbf{v}\right)_\Omega\right\|_2^2 = \left|\left\langle\left((\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{A}\mathbf{v}\right)_\Omega, (\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{A}\mathbf{v}\right\rangle\right|.$$

Taking $\left((\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{A}\mathbf{v}\right)_\Omega$ as $\mathbf{u}$ in (11) yields

$$\left\|\left((\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{A}\mathbf{v}\right)_\Omega\right\|_2^2 \le \delta_t \left\|\left((\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{A}\mathbf{v}\right)_\Omega\right\|_2 \|v\|_2.$$

Therefore, by eliminating $\left\|\left((\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{A}\mathbf{v}\right)_\Omega\right\|_2$, the desired result (12) can be obtained. □

Lemma 2.8. *Given any vector* $\mathbf{e} \in \mathbb{R}^m$ *and any index set* $S \subset [N]$ *with* $|S| \le s$, *if matrix* $\mathbf{A}$ *satisfies the RIC of order* s *and* $\|\mathbf{A}\|_2 < 1$, *then*

$$\|((\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{e})_S\|_2 \le \sqrt{1 + \delta_s}\|\mathbf{e}\|_2.$$

Proof. Firstly, we will show that

$$\|\mathbf{I} - \mathbf{A}\mathbf{A}^T\|_2 \le 1. \tag{13}$$

In fact, consider the singular value decomposition $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where $\mathbf{U} \in \mathbb{R}^{m\times m}$ and $\mathbf{V} \in \mathbb{R}^{N\times N}$ are orthogonal matrices, and

$$\mathbf{\Sigma} = \left(\begin{array}{cccc|c} \sigma_1 & & & & \\ & \sigma_2 & & & \mathbf{0}_{m\times(N-m)} \\ & & \ddots & & \\ & & & \sigma_m & \end{array}\right)_{m\times N},$$

where σ_i $(i = 1, 2, \cdots, m)$ are the singular values of $\mathbf{A}$. They satisfy $0 \le \sigma_m \le \cdots \le \sigma_2 \le \sigma_1 < 1$. Therefore,

$$\|\mathbf{I} - \mathbf{A}^T\mathbf{A}\|_2 = \|\mathbf{V}(\mathbf{I} - \mathbf{\Sigma}^2)\mathbf{V}^T\|_2 = \|\mathbf{I} - \mathbf{\Sigma}^2\|_2 = 1 - \sigma_m^2 \le 1.$$

Furthermore, notice that

$$\left\|\left((\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{e}\right)_S\right\|_2^2 = \left|\left\langle(\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{e}, \left((\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{e}\right)_S\right\rangle\right|. \tag{14}$$

For the right of (14), we have

$$\begin{aligned}
&\left\langle(\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{e}, \left((\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{e}\right)_S\right\rangle \\
&= \left\langle\mathbf{e}, \mathbf{A}\left((\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{e}\right)_S\right\rangle - \left\langle\mathbf{A}\mathbf{A}^T\mathbf{e}, \mathbf{A}\left((\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{e}\right)_S\right\rangle \\
&= \left\langle(\mathbf{I} - \mathbf{A}\mathbf{A}^T)\mathbf{e}, \mathbf{A}\left((\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{e}\right)_S\right\rangle.
\end{aligned} \tag{15}$$

Substituting (15) into (14) and using Cauchy inequality, we get

$$\left\|\left((\mathbf{I}-\mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{e}\right)_S\right\|_2^2 \le \|(\mathbf{I}-\mathbf{A}\mathbf{A}^T)\mathbf{e}\|_2 \cdot \left\|\mathbf{A}\left((\mathbf{I}-\mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{e}\right)_S\right\|_2. \tag{16}$$

Notice that $\left\|\left((\mathbf{I}-\mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{e}\right)_S\right\|_0 \le s$, according to Definition 2.1, we have

$$\left\|\mathbf{A}\left((\mathbf{I}-\mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{e}\right)_S\right\|_2 \le \sqrt{1+\delta_s}\left\|\left((\mathbf{I}-\mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{e}\right)_S\right\|_2. \tag{17}$$

Moreover, (13) implies

$$\|(\mathbf{I}-\mathbf{A}\mathbf{A}^T)\mathbf{e}\|_2 \le \|\mathbf{I}-\mathbf{A}\mathbf{A}^T\|_2 \cdot \|\mathbf{e}\|_2 \le \|\mathbf{e}\|_2. \tag{18}$$

Therefore, by plugging (17) and (18) into (16) and eliminating $\|((\mathbf{I}-\mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{e})_S\|_2$, the desired result can be obtained. □

3. Main results

In this section, we will present our main result (Theorem 3.1) and provide a detailed proof using the technique lemmas introduced in Section 2. Moreover, Corollary 3.2 will be established to determine the number of iterations for an idealized setting (i.e., $\mathbf{y} = \mathbf{A}\mathbf{x}$) coupled with a fixed stepsize $\mu = 1$.

Theorem 3.1. *Suppose* $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ *where* $\mathbf{x} \in \mathbb{R}^N$ *is an arbitrary signal and* $\mathbf{e} \in \mathbb{R}^m$ *is a possible observation noise. If S denotes the index set of s largest (in modulus) entries of* $\mathbf{x}$, *and the restricted isometry constant of sensing matrix* $\mathbf{A} \in \mathbb{R}^{m\times N}$ *obeys*

$$\delta_{3s} < \varphi(\mu) = \frac{\sqrt{(18-6\sqrt{5})\mu^2 + (12-4\sqrt{5})\mu + (8-4\sqrt{5})} - 4\mu|1-\mu|}{8\mu^2 + 3 - \sqrt{5}}, \tag{19}$$

then the iterative sequence $\{\mathbf{x}_k\}$ *generated by the QNHTP algorithm (10) satisfies*

$$\|\mathbf{x}_k - \mathbf{x}_S\|_2 \le \rho^k\|\mathbf{x}_0 - \mathbf{x}_S\|_2 + \tau\|\mathbf{A}\mathbf{x}_{\overline{S}} + \mathbf{e}\|_2, \tag{20}$$

where

$$\rho = \frac{\sqrt{5}+1}{2}\cdot\frac{|1-\mu| + 2\mu\delta_{3s}}{\sqrt{1-\delta_{2s}^2}} < 1 \text{ and } \tau = \frac{(\sqrt{5}+1)\mu\sqrt{1-\delta_{2s}} + \sqrt{1+\delta_s}}{(1-\delta_{2s})(1-\rho)}. \tag{21}$$

Proof. Rewrite $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ as $\mathbf{y} = \mathbf{A}\mathbf{x}_S + \mathbf{e}'$, where $\mathbf{e}' = \mathbf{A}\mathbf{x}_{\overline{S}} + \mathbf{e}$. Let $\overline{\mathbf{x}_k}$ and $\mathbf{x}_{k+1}$ as given in QNHTP algorithm (10). Denote $\mathbf{u}_k = \mathbf{x}_k + \mu(2\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{x}_k)$ and $S_{k+1} = \mathrm{supp}(\overline{\mathbf{x}_k}) = \mathrm{supp}(\mathcal{H}_s(\mathbf{u}_k))$.

On the one hand, by using Lemma 2.4, we have

$$\|\overline{\mathbf{x}_k} - \mathbf{x}_S\|_2 = \|\mathcal{H}_s(\mathbf{u}_k) - \mathbf{x}_S\|_2 \le \frac{\sqrt{5}+1}{2}\|(\mathbf{u}_k - \mathbf{x}_S)_{S\cup S_{k+1}}\|_2. \tag{22}$$

Substituting $\mathbf{y} = \mathbf{A}\mathbf{x}_S + \mathbf{e}'$ into $\mathbf{u}_k$ yields

$$\begin{aligned}\mathbf{u}_k - \mathbf{x}_S &= \mathbf{x}_k - \mathbf{x}_S + \mu\mathbf{A}^T(\mathbf{A}\mathbf{x}_S + \mathbf{e}' - \mathbf{A}\mathbf{x}_k) + \mu(\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T(\mathbf{A}\mathbf{x}_S + \mathbf{e}' - \mathbf{A}\mathbf{x}_k)\\ &= (1-\mu)(\mathbf{x}_k - \mathbf{x}_S) + \mu(\mathbf{I} - \mathbf{A}^T\mathbf{A})(\mathbf{x}_k - \mathbf{x}_S) + \mu\mathbf{A}^T\mathbf{e}' + \mu(\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{A}(\mathbf{x}_S - \mathbf{x}_k) + \mu(\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{e}'.\end{aligned}$$

Therefore,

$$\begin{aligned}\left\|(\mathbf{u}_k - \mathbf{x}_S)_{S\cup S_{k+1}}\right\|_2 \le{}& |1-\mu|\cdot\left\|(\mathbf{x}_k - \mathbf{x}_S)_{S\cup S_{k+1}}\right\|_2 + \mu\left\|\left((\mathbf{I} - \mathbf{A}^T\mathbf{A})(\mathbf{x}_k - \mathbf{x}_S)\right)_{S\cup S_{k+1}}\right\|_2 + \mu\|(\mathbf{A}^T\mathbf{e}')_{S\cup S_{k+1}}\|_2\\ &+ \mu\left\|\left((\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{A}(\mathbf{x}_S - \mathbf{x}_k)\right)_{S\cup S_{k+1}}\right\|_2 + \mu\left\|\left((\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{e}'\right)_{S\cup S_{k+1}}\right\|_2.\end{aligned} \tag{23}$$

In what follows, we analyze five terms on the right side of inequality (23). Firstly, it is an obvious fact that

$$\left\|(\mathbf{x}_k - \mathbf{x}_S)_{S\cup S_{k+1}}\right\|_2 \leq \|\mathbf{x}_k - \mathbf{x}_S\|_2. \tag{24}$$

Secondly, note that $|(S \cup S_{k+1}) \cup \mathrm{supp}(\mathbf{x}_k - \mathbf{x}_S)| \leq 3s$. Therefore, by using Lemma 2.1, we have

$$\left\|\left((\mathbf{I} - \mathbf{A}^T\mathbf{A})(\mathbf{x}_k - \mathbf{x}_S)\right)_{S\cup S_{k+1}}\right\|_2 \leq \delta_{3s}\|\mathbf{x}_k - \mathbf{x}_S\|_2. \tag{25}$$

Additionally, using (12) in Lemma 2.5 yields

$$\left\|\left((\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{A}(\mathbf{x}_S - \mathbf{x}_k)\right)_{S\cup S_{k+1}}\right\|_2 \leq \delta_{3s}\|\mathbf{x}_k - \mathbf{x}_S\|_2. \tag{26}$$

Similarly, by using Lemma 2.2 and Lemma 2.6 with the fact that $|\mathrm{supp}(\mathbf{x}_k - \mathbf{x}_S)| \leq 2s$, we obtain

$$\|(\mathbf{A}^T\mathbf{e}')_{S\cup S_{k+1}}\|_2 \leq \sqrt{1+\delta_{2s}}\|\mathbf{e}'\|_2 \tag{27}$$

and

$$\left\|\left((\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{e}'\right)_{S\cup S_{k+1}}\right\|_2 \leq \sqrt{1+\delta_{2s}}\|\mathbf{e}'\|_2, \tag{28}$$

respectively. Plugging (24)-(28) into (23) thus gives

$$\left\|(\mathbf{u}_k - \mathbf{x}_S)_{S\cup S_{k+1}}\right\|_2 \leq (|1-\mu| + 2\mu\delta_{3s})\|\mathbf{x}_k - \mathbf{x}_S\|_2 + 2\mu\sqrt{1+\delta_{2s}}\|\mathbf{e}'\|_2. \tag{29}$$

To conclude, by combining (22) and (29), we have

$$\|\overline{\mathbf{x}_k} - \mathbf{x}_S\|_2 \leq \frac{\sqrt{5}+1}{2}(|1-\mu| + 2\mu\delta_{3s})\|\mathbf{x}_k - \mathbf{x}_S\|_2 + (\sqrt{5}+1)\mu\sqrt{1+\delta_{2s}}\|\mathbf{e}'\|_2. \tag{30}$$

On the other hand, the following result can be obtained by using Lemma 2.3

$$\|\mathbf{x}_S - \mathbf{x}_{k+1}\|_2 \leq \frac{1}{\sqrt{1-\delta_{2s}^2}}\|\mathbf{x}_S - \overline{\mathbf{x}_k}\|_2 + \frac{\sqrt{1+\delta_s}}{1-\delta_{2s}}\|\mathbf{A}\mathbf{x}_{\overline{S}} + \mathbf{e}\|_2. \tag{31}$$

Finally, by substituting (30) into (31), we get

$$\begin{aligned}\|\mathbf{x}_S - \mathbf{x}_{k+1}\|_2 &\leq \frac{\sqrt{5}+1}{2}\cdot\frac{|1-\mu| + 2\mu\delta_{3s}}{\sqrt{1-\delta_{2s}^2}}\|\mathbf{x}_k - \mathbf{x}_S\|_2 + \tau'\|\mathbf{A}\mathbf{x}_{\overline{S}} + \mathbf{e}\|_2\\ &= \rho\|\mathbf{x}_k - \mathbf{x}_S\|_2 + \tau'\|\mathbf{A}\mathbf{x}_{\overline{S}} + \mathbf{e}\|_2,\end{aligned} \tag{32}$$

where ρ is as given in (21), and

$$\tau' = \frac{(\sqrt{5}+1)\mu\sqrt{1-\delta_{2s}} + \sqrt{1+\delta_s}}{1-\delta_{2s}}.$$

In the following, we analyze the RIC sufficient condition (19) such that

$$\rho = \frac{\sqrt{5}+1}{2}\cdot\frac{|1-\mu| + 2\mu\delta_{3s}}{\sqrt{1-\delta_{2s}^2}} < 1. \tag{33}$$

Notice that $\delta_{2s} \leq \delta_{3s}$, thus (33) holds as soon as the following inequality satisfies:

$$\frac{\sqrt{5}+1}{2}\cdot\frac{|1-\mu| + 2\mu\delta_{3s}}{\sqrt{1-\delta_{3s}^2}} < 1. \tag{34}$$

Rearranging (34) gives a quadratical polynomial inequality concerning variable $t = \delta_{3s}$

$$\phi(t) = \left(4\mu^2 + \frac{3-\sqrt{5}}{2}\right)t^2 + 4\mu|1-\mu|t + \left((1-\mu)^2 - \frac{3-\sqrt{5}}{2}\right) < 0.$$

Therefore, δ_{3s} must be smaller than the largest root of $\phi(t)$, namely,

$$\begin{aligned}\delta_{3s} &< \frac{-4\mu|1-\mu| + \sqrt{16\mu^2(1-\mu)^2 - 4(4\mu^2 + \frac{3-\sqrt{5}}{2})\left((1-\mu)^2 - \frac{3-\sqrt{5}}{2}\right)}}{8\mu^2 + 3 - \sqrt{5}} \\ &= \frac{-4\mu|1-\mu| + \sqrt{(18-6\sqrt{5})\mu^2 + (12-4\sqrt{5})\mu + (8-4\sqrt{5})}}{8\mu^2 + 3 - \sqrt{5}}.\end{aligned}$$

Thus, we prove that the desired constraint $\rho < 1$ holds as soon as (19) is satisfied. Besides, the expression for τ in Theorem 3.1 cn be immediately derived from (32), coupled with the fact that $\sum_{i=0}^{\infty} \rho^i = 1/(1-\rho)$ holds for any $\rho < 1$. □

Remark 3.2. Obviously, a meaningful RIC assumption requires that the upper bound function $\varphi(\mu)$ given in (19) satisfies $0 < \varphi(\mu) < 1$. This implies that the stepsize μ should be selected from the range of $\left((3-\sqrt{5})/2, (1+\sqrt{5})/2\right)$ for successful recovery via QNHTP algorithm (10). Moreover, as shown in [23], the larger the value of $\varphi(\mu)$, the better theoretical recovery performance is. In this regard, we plotted $\varphi(\mu)$ in Fig. 1 and found that $\mu = 1$ is an extreme point. This finding demonstrates that if a fixed constant stepsize parameter of one is used for QNHTP algorithm (10), then the optimal upper bound of RIC can be achieved.

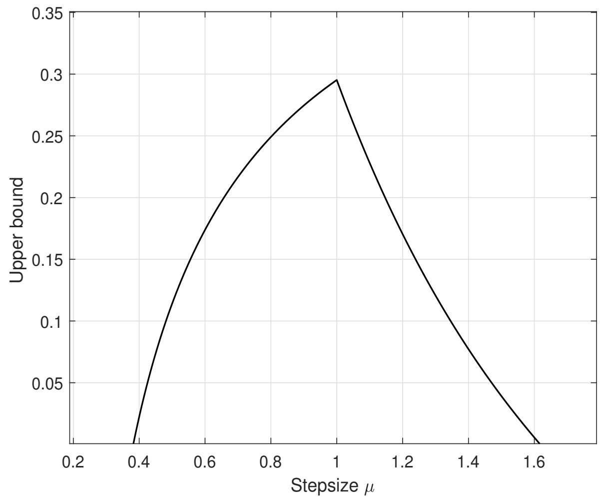

Figure 1: The curve of upper bound function $\varphi(\mu)$, which reaches its maximum at the unit stepsize.

Using this optimal upper bound $\varphi(1)$, the sufficient condition (19) in Theorem 3.1 is reduced to $\delta_{3s} < \sqrt{38-14\sqrt{5}}\big/(11-\sqrt{5}) \approx 0.2952$, which is an improvement over the QNNP algorithm [35] that also adopted the quasi-Newton direction and had a sufficient condition of $\delta_{3s} < \sqrt[3]{54+6\sqrt{87}}\big/3 - 2/\sqrt[3]{54+6\sqrt{87}} - 1 \approx 0.1795$. Moreover, taking into account an idealized setting (i.e., $\mathbf{y} = \mathbf{A}\mathbf{x}$) and s-sparse signal $\mathbf{x}$ (i.e., $\mathbf{x}_S = \mathbf{x}$) under this condition, the inequality (20) becomes

$$\|\mathbf{x} - \mathbf{x}_k\|_2 \le \rho_1^k \|\mathbf{x}_0 - \mathbf{x}\|_2 \quad \left(\rho_1 = (\sqrt{5}+1)\delta_{3s}/\sqrt{1-\delta_{2s}^2}\right), \tag{35}$$

from which we can see that the iterative sequence $\{\mathbf{x}_k\}$ generated by the QNHTP algorithm (10) converges to $\mathbf{x}$ at a geometric rate. Furthermore, it can be concluded that convergence under an idealized setting requires a finite number of iterations. We summarize this observation as following corollary:

Corollary 3.3. *If the RIC of order* $3s$ *of sensing matrix* $\mathbf{A} \in \mathbb{R}^{m\times N}$ *obeys*

$$\delta_{3s} < \frac{\sqrt{38 - 14\sqrt{5}}}{11 - \sqrt{5}} \approx 0.2952,$$

then any s-sparse signal $\mathbf{x} \in \mathbb{R}^N$ *is recovered by QNHTP algorithm (10) with* $\mathbf{y} = \mathbf{A}\mathbf{x}$ *in at most*

$$\left\lceil \frac{\ln\left(\sqrt{76 - 28\sqrt{5}}\|\mathbf{x}_0 - \mathbf{x}\|_2/((11 - \sqrt{5})\eta)\right)}{\ln(1/\rho_1)} \right\rceil \tag{36}$$

iterations, where $\eta = \min_{i\in\mathrm{supp}(\mathbf{x})} |\mathbf{x}_i|$*, and* $\rho_1 = (\sqrt{5} + 1)\delta_{3s}/\sqrt{1 - \delta_{2s}^2}$.

Proof. Our proof follows the idea of [23, Corollary 3.6]. We need to show that there exists an integer k such that $S_k = S = \mathrm{supp}(\mathbf{x})$, that is to say, for all $p \in S$ and all $q \in \overline{S}$, it holds that

$$\left|\left(\mathbf{x}_k + (2\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{A}(\mathbf{x} - \mathbf{x}_k)\right)_p\right| > \left|\left(\mathbf{x}_k + (2\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{A}(\mathbf{x} - \mathbf{x}_k)\right)_q\right|. \tag{37}$$

For the left side of (37), since $\eta = \min_{\mathrm{supp}(\mathbf{x})} |\mathbf{x}_i| \le |\mathbf{x}_p|$, we have

$$\begin{aligned}
&\left|\left(\mathbf{x}_k + (2\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{A}(\mathbf{x} - \mathbf{x}_k)\right)_p\right| \\
&= \left|\left(\mathbf{x} + (\mathbf{I} - (2\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{A})(\mathbf{x}_k - \mathbf{x})\right)_p\right| \\
&\ge |\mathbf{x}_p| - \left|\left((\mathbf{I} - (2\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{A})(\mathbf{x}_k - \mathbf{x})\right)_p\right| \\
&\ge \eta - \left|\left((\mathbf{I} - (2\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{A})(\mathbf{x}_k - \mathbf{x})\right)_p\right|.
\end{aligned} \tag{38}$$

As for the right side of (37), since $\mathbf{x}_q = 0$, we have

$$\begin{aligned}
\left|\left(\mathbf{x}_k + (2\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{A}(\mathbf{x} - \mathbf{x}_k)\right)_q\right| &= \left|\left(\mathbf{x}_k - \mathbf{x} + (2\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{A}(\mathbf{x} - \mathbf{x}_k)\right)_q\right| \\
&= \left|\left((\mathbf{I} - (2\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{A})(\mathbf{x}_k - \mathbf{x})\right)_q\right|.
\end{aligned} \tag{39}$$

By combining (37)-(39), we need to find an integer k such that

$$\left|\left((\mathbf{I} - (2\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{A})(\mathbf{x}_k - \mathbf{x})\right)_p\right| + \left|\left((\mathbf{I} - (2\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{A})(\mathbf{x}_k - \mathbf{x})\right)_q\right| \le \eta. \tag{40}$$

Applying the symmetric difference inequality to the left side of (40) yields

$$\begin{aligned}
&\left|\left((\mathbf{I} - (2\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{A})(\mathbf{x}_k - \mathbf{x})\right)_p\right| + \left|\left((\mathbf{I} - (2\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{A})(\mathbf{x}_k - \mathbf{x})\right)_q\right| \\
&\le \sqrt{2}\left\|\left((\mathbf{I} - (2\mathbf{I} - \mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{A})(\mathbf{x}_k - \mathbf{x})\right)_{\{p,q\}}\right\|_2 \\
&= \sqrt{2}\left\|\left((\mathbf{I} - \mathbf{A}^T\mathbf{A})(\mathbf{I} - \mathbf{A}^T\mathbf{A})(\mathbf{x}_k - \mathbf{x})\right)_{\{p,q\}}\right\|_2 \\
&\le \sqrt{2}\|\mathbf{I} - \mathbf{A}^T\mathbf{A}\|_2 \cdot \left\|\left((\mathbf{I} - \mathbf{A}^T\mathbf{A})(\mathbf{x}_k - \mathbf{x})\right)_{\{p,q\}}\right\|_2 \\
&\le \sqrt{2}\delta_{3s}\|\mathbf{x}_k - \mathbf{x}\|_2 \\
&\le \frac{\sqrt{76 - 28\sqrt{5}}}{11 - \sqrt{5}}\rho_1^k\|\mathbf{x}_0 - \mathbf{x}\|_2,
\end{aligned}$$

where we use the fact (13) and Lemma 2.1 in the third inequality since $|\text{supp}(\mathbf{x}_k - \mathbf{x}) \cup \{p, q\}| \leq 3s$, and the last inequality is from (35) and $\delta_{3s} < \sqrt{38 - 14\sqrt{5}}/(11 - \sqrt{5})$. Thus, we see that (40) holds as soon as

$$\frac{\sqrt{76 - 28\sqrt{5}}}{11 - \sqrt{5}} \rho_1^k \|\mathbf{x}_0 - \mathbf{x}\|_2 \leq \eta,$$

from which it is easy to deduce the desired result (36). □

Remark 3.4. As mentioned in Section 1, the QNIP algorithm [25] is an IHT-type algorithms with quasi-Newton direction and adaptive stepsize, as given in (8). In the theoretical analysis of the QNIP algorithm, the authors performed an iterative scheme (10a) by choosing $\mu = 1/(1 - \delta_{2s}^2)$, and developed a RIC sufficient condition to guarantee successful recovery for s-sparse signals in an idealized setting. We analyze the upper bound using the same stepsize for comparison.

Actually, if we apply the proof idea of Theorem 3.1 to the QNIP algorithm, it is easy to derive that the expression for convergence rate ρ takes the form of

$$\rho = \frac{\sqrt{5} + 1}{2}(|1 - \mu| + 2\mu\delta_{3s}) \tag{41}$$

according to (30) derived from step (10a). Substituting $\mu = 1/(1 - \delta_{2s}^2)$ into (41) gives

$$\rho = \frac{\sqrt{5} + 1}{2} \cdot \frac{\delta_{2s}^2 + 2\delta_{3s}}{1 - \delta_{2s}^2} \leq \frac{\sqrt{5} + 1}{2} \cdot \frac{\delta_{3s}^2 + 2\delta_{3s}}{1 - \delta_{3s}^2},$$

from which we obtain the sufficient condition $\delta_{3s} < (\sqrt{2} - 1)(\sqrt{5} - 1)/2 \approx 0.2560$ such that $\rho < 1$. This upper bound is better than the one given in [25], which requires $\delta_{3s} < 3 - 2\sqrt{2} \approx 0.1716$. Furthermore, if we take the optimal stepsize $\mu = 1$ in (41), the upper bound can be further improved to $\delta_{3s} < (\sqrt{5} - 1)/4 \approx 0.3090$.

4. Conclusion

In this paper, we propose a new HTP-type sparse recovery algorithm by incorporating the quasi-Newton direction into the HTP algorithm. We demonstrate its convergence based on the RIC of the sensing matrix. It is worth mentioning that in this work, we only focus on the theoretical guarantees of the proposed algorithm. In future research, we will investigate its empirical performance and restriction to the sensing matrix **A** (its largest singular value must be below 1 for convergence to be guaranteed).

References

[1] J. D. Blanchard, J. Tanner, K. Wei, *CGIHT: conjugate gradient iterative hard thresholding for compressed sensing and matrix completion*, Information and Inference: A Journal of the IMA. **4** (2015), 289–327.

[2] J. D. Blanchard, J. Tanner, K. Wei, *Conjugate gradient iterative hard thresholding: observed noise stability for compressed sensing*, IEEE Trans. Signal Process. **63** (2015), 528–537.

[3] T. Blumensath, *Accelerated iterative hard thresholding*, Signal Process. **92** (2012), 752–756.

[4] T. Blumensath, *Compressed sensing with nonlinear observations and related nonlinear optimization problems*, IEEE Trans. Inf. Theory. **59** (2013), 3466–3474.

[5] T. Blumensath, M. E. Davies, *Iterative threshloding for sparse approxiamtions*, J. Fourier Anal. Appl. **14** (2008), 629–654.

[6] T. Blumensath, M. E. Davies, *Iterative hard threshloding for compressed sensing*, Appl. Comput. Harmon. Anal. **27** (2009), 265–274.

[7] T. Blumensath, M. E. Davies, *Normalized iterative hard threshloding: guaranteed stability and performance*, IEEE J. Sel. Top. Signal Process. **4** (2010), 298–309.

[8] J. L. Bouchot, S. Foucart, P. Hitczenko, *Hard thresholding pursuit algorithms: Number of iterations*, Appl. Comput. Harmon. Anal. **41** (2016), 412–435.

[9] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge Univ. Press, 2004.

[10] E. J. Candès, J. Romberg, T. Tao, *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inf. Theory. **52** (2006), 489–509.

[11] E. J. Candès, T. Tao, *Decoding by linear programming*, IEEE Trans. Inf. Theory. **51** (2005), 4203–4215.
[12] C. Cartis, A. Thompson, *A new and improved quantitative recovery analysis for iterative hard thresholding algorithms in compressed sensing*, IEEE Trans. Inf. Theory. **61** (2015), 1–24.
[13] R. Chartrand, *Exact reconstruction of sparse signals via nonconvex minimization*, IEEE Signal Process. Lett. **14** (2007), 707–710.
[14] R. Chartrand, V. Staneva, *Restricted isometry properties and nonconvex compressive sensing*, Inverse Probl. **24** (2008), Art. no. 035020.
[15] S. S. Chen, D. L. Donoho, M. A. Saunders, *Atomic decomposition by basis pursuit*, SIAM J. Sci. Comput. **20** (1998), 33–61.
[16] W. Dai, O. Milenkovic, *Subspace pursuit for compressive sensing signal reconstruction*, IEEE Trans. Inf. Theory. **55** (2009), 2230–2249.
[17] I. Daubechies, M. Defrise, C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Appl. Math. **57** (2004), 1413–1457.
[18] D. L. Donoho, *De-noising by soft-thresholding*, IEEE Trans. Inf. Theory. **41** (1995), 613–627.
[19] D. L. Donoho, *Compressed sensing*, IEEE Trans. Inf. Theory. **52** (2006), 1289–1306.
[20] D. L. Donoho, A. Maleki, *Optimally tuned iterative thresholding algorithms for compressed sensing*, IEEE Sel. Top. Signal Process. **4** (2010), 330–341.
[21] M. Elad, *Why simple shrinkage is still relevant for redundant representations?*, IEEE Trans. Inf. Theory. **52** (2006), 5559–5569.
[22] E. Esser, Y. Lou, J. Xin, *A method for finding structured sparse solutions to nonnegative least squares problems with applications*, SIAM J. Imaging Sci. **6** (2013), 2010–2046.
[23] S. Foucart, *Hard thresholding pursuit: an algorithm for compressive sensing*, SIAM J. Numer. Anal. **49** (2011), 2543–2563.
[24] W. Jin, L.-J. Xie, *Momentum-based iterative hard thresholding algorithm for sparse signal recovery*, IEEE Signal Process. Lett. **32** (2025), 1346–1350.
[25] M.-L. Jing, X.-Q. Zhou, C. Qi, *Quasi-Newton iterative projection algorithm for sparse recovery*, Neurocomputing **144** (2011), 169–173.
[26] N. Meng, Y.-B. Zhao, *Newton-step-based hard thresholding algorithms for sparse signal recovery*, IEEE Trans. Signal Process. **68** (2020), 6594–6606.
[27] B. K. Natarajan, *Sparse approximate solutions to linear systems*, SIAM J. Comput. **24** (1995), 227–234.
[28] D. Needell, J. A. Tropp, *CoSaMP: Iterative signal recovery from incomplete and inaccurate samples*, Appl. Comput. Harmon. Anal. **26** (2009), 301–321.
[29] H. K. Nigam, S. Yadav, *Sparse recovery for compressive sensing via weighted L_{p-q} model*, Filomat **36** (2022), 4709–4716.
[30] J. Nocedal, S. Wright, *Numerical Optimization. Springer Science and Business Media*, Springer, New York, USA, 2006.
[31] Y. C. Pati, R. Rezaiifar, P. S.Krishnaprasad, *Orthogonal matching pursuits: recursive function approximation with applications to wavelet decomposition* in Proc. the 27th Annual Asilomar Conf. Signals Systems and Computers, Pacific Grove, CA, USA, (1993), 40–44.
[32] J. Shen, P. Li, *A tight bound of hard thresholding*, J. Mach. Learn. Res. **18** (2018), 1–42.
[33] Z.-F. Sun, J.-C. Zhou, Y.-B. Zhao, N. Meng, *Heavy-ball-based hard thresholding algorithms for sparse signal recovery*, J. Comput. Appl. Math. **430** (2023), Art. no. 115264.
[34] S. Voronin, H. J. Woerdeman, *A new iterative firm-thresholding algorithm for inverse problems with sparsity constraints*, Appl. Comput. Harmon. Anal. **35** (2013), 151–164.
[35] Q. Wang, G.-R. Qu, *A new greedy algorithm for sparse recovery*, Neurocomputing **275** (2018), 137–143.
[36] J. Wen, H. He, Z. He, and F. Zhu, *A pseudo-inverse-based hard thresholding algorithm for sparse signal recovery*, IEEE Trans. Intell. Transp. Syst. **24** (2023), 7621–7630.
[37] L.-J. Xie, *Improved RIC bounds in terms of δ_{2s} for hard thresholding-based algorithms*, IEEE Signal Process. Lett. **30** (2023), 21–25.
[38] L.-J. Xie, W. Jin, *From theoretical guarantee to practical performance: selectable and optimal step-lengths for IHT and HTP algorithms in compressed sensing*, Comput. Appl. Math. **44** (2023), Art. no. 1.
[39] L.-J. Xie, J.-P. Wang, *Sufficient condition based on nearly optimal order RIC for IHT algorithm*, Appl. Anal. (2024), doi:10.1080/00036811.2024.2311697.
[40] P. Yin, Y. Lou, Q. He, J. Xin, *Minimization of l_{1-2} for compressed sensing*, SIAM J. Sci. Comput. **37** (2015), 536–563.
[41] Y.-B. Zhao, *Optimal $k-$thresholding algorithms for sparse optimization problems*, SIAM J. Optim. **30** (2020), 31–55.
[42] Y.-B. Zhao, Z.-Q. Luo, *Analysis of optimal thresholding algorithms for compressed sensing*, Signal Process. **187** (2021), Art. no. 108148.
[43] Y.-B. Zhao, Z.-Q. Luo, *Improved RIP-based bounds for guaranteed performance of two compressed sensing algorithms*, Sci. China Math. **66** (2023), 1123–1140.
[44] S.-L. Zhou, N.-H. Xiu, H.-D. Qi, *Global and quadratic convergence of Newton hard-thresholding pursuit*, J. Mach. Learn. Res. **22** (2021), 1–45.