Power Series Based Hard Thresholding Algorithms for Sparse Signal Recovery

Ketan Atul Bapat, *Graduate Student Member, IEEE*, and Mrityunjoy Chakraborty, *Senior Member, IEEE*

***Abstract*—This paper presents a unified treatment to hard thresholding based compressed sensing recovery algorithms. For this, it modifies the cost function by a power series in $\mathbf{AA}^H$ where $\mathbf{A}$ is the so-called sensing matrix. For appropriate choice of the power series coefficients, the proposed treatment not only leads to various existing hard thresholding based recovery algorithms, but, more importantly, it enables one to develop new algorithms belonging to this category. The paper also presents a convergence analysis of the proposed method and derives convergence guarantees in terms of the restricted isometry constant (RIC) of the sensing matrix. It is seen that in case of some of the well known hard thresholding based algorithms, the proposed convergence analysis results in wider ranges of algorithm parameters and faster convergence than suggested by existing analyses. Some new, power series based hard thresholding algorithms are also proposed and their recovery performance studied via simulation.**

***Index Terms*—Compressed sensing, hard thresholding, power series, restricted isometry constant (RICs).**

I. Introduction

In compressed sensing, the goal is to reconstruct a K-sparse $\mathbf{x} \in \mathbb{C}^N$ from the (noisy) linear measurements $\mathbf{y} = \mathbf{Ax} + \mathbf{e}$, where $\mathbf{A} \in \mathbb{C}^{m \times N}, m < N$, is the given sensing matrix. Majority of the greedy/thresholding based methods consider the following problem:

$$\min_{\mathbf{z} \in \mathbb{C}^N} \frac{1}{2} \|\mathbf{y} - \mathbf{Az}\|_2^2, \text{ subject to } \|\mathbf{z}\|_0 \leq K, \tag{1}$$

where $\|\mathbf{z}\|_0$ counts the nonzero entries in $\mathbf{z}$. Iterative hard thresholding (IHT) [1] and hard thresholding pursuit (HTP) [2] are two popular thresholding based methods that aim to solve the above problem. Both these methods employ gradient of the objective function used in (1), namely, $\mathbf{A}^H(\mathbf{y} - \mathbf{Az})$ for performing the update (where H in superscript denotes Hermitian transposition). In case of IHT, hard thresholding (HT) is performed on the gradient descent based update, whereas, in case of HTP, an additional least squares step is introduced following the hard thresholding step in the IHT algorithm. Sufficient conditions on the RICs to ensure convergence of the IHT and HTP methods are provided in [2], [3], [4].

In recent years, certain extensions of IHT and HTP have come up which do not use the standard gradient based update. For example, in [5], the authors propose methods which combine hard thresholding with feedback and null space tuning (NST) approach, requiring the computation of $\mathbf{A}^\dagger(\mathbf{y} - \mathbf{Ax})$ for carrying out an update, where $\mathbf{A}^\dagger = \mathbf{A}^H(\mathbf{AA}^H)^{-1}$ denotes the pseudo inverse of $\mathbf{A}$ (assuming $\mathbf{A}$ has full row rank). The NST+HT algorithm and its variants have been shown to be more efficient in error correction (denoising) than the standard IHT algorithm. In [6], a pseudo inverse based hard thresholding (PHT) algorithm is proposed which, as the name suggests, carries out an update involving $\mathbf{A}^\dagger$. In [7], the authors proposed two algorithms, namely Newton step based iterative hard thresholding (NSIHT) and Newton step based hard thresholding pursuit (NSHTP). These algorithms use inverse of the perturbed Hessian of the objective function in (1), namely, $(\mathbf{A}^H\mathbf{A} + \epsilon\mathbf{I})^{-1}$, for carrying out an update, where ϵ is a parameter. It was argued that NSHTP algorithm stabilizes the oscillatory behaviour that is observed in HTP algorithm. The update for NSIHT and NSHTP algorithms require calculation of $(\mathbf{A}^H\mathbf{A} + \epsilon\mathbf{I})^{-1}\mathbf{A}^H(\mathbf{y} - \mathbf{Az})$ for carrying out the update using hard thresholding.

In this paper, we provide a unified treatment to the gradient based recovery algorithms by means of a power series in $\mathbf{AA}^H$. We show that not only the above stated gradient based algorithms can be obtained as special cases of the proposed approach, but more importantly, one can derive new linearly convergent algorithms by properly tuning the power series coefficients. We propose a generalized convergence analysis, which results in more relaxed theoretical guarantees for certain existing methods. The proposed approach considers the following problem:

$$\min_{\mathbf{u} \in \mathbb{C}^N} \|\mathbf{Z}(\mathbf{y} - \mathbf{Au})\|_2^2 \text{ subject to } \|\mathbf{u}\|_0 \leq K, \tag{2}$$

where $\mathbf{Z}^H\mathbf{Z} = q(\mathbf{AA}^H)$, with $q(t)$ being a power series in t with coefficients $\alpha_i \in \mathbb{R}$, given by $q(t) = \sum_{i=c}^{\infty} \alpha_i t^i$ where c is an integer. Let $\lambda_i \geq 0, i = 1, \ldots, m$ be the eigenvalues of the positive semidefinite matrix $\mathbf{AA}^H$. Throughout the paper, we assume that $q(t)$ is a convergent power series for $t = \lambda_i, i = 1, \ldots, m$ and $q(\lambda_i) > 0$. Since an eigenvalue λ_i of $\mathbf{AA}^H$ gives rise to an eigenvalue $q(\lambda_i)$ for the matrix $q(\mathbf{AA}^H)$, we have $q(\mathbf{AA}^H)$ positive definite. Following the update strategies of IHT and HTP methods, two classes of algorithms: power series based iterative hard thresholding (PSIHT) and power series based hard thresholding pursuit (PSHTP) are proposed and sufficient conditions for reconstruction for both are derived in terms of RIC of the sensing matrix. Both PSIHT and PSHTP offer possibilities of developing new gradient based sparse recovery algorithms for which we present few examples here. The proposed analysis also establishes wider permissible range of algorithm parameters for convergence and also faster convergence rate for the NSIHT and NSHTP algorithms [7] than given in [7].

Received 1 May 2025; accepted 1 June 2025. Date of publication 6 June 2025; date of current version 30 June 2025. This work was supported by the Science and Engineering Research Board, the Govt. of India. The associate editor coordinating the review of this article and approving it for publication was Dr. Lu Gan. *(Corresponding author: Mrityunjoy Chakraborty.)*

The authors are with the Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur 721302, India (e-mail: ketan9318@gmail.com; mrityun@ece.iitkgp.ac.in).

This article has supplementary downloadable material available at https://doi.org/10.1109/LSP.2025.3577132, provided by the authors.

Digital Object Identifier 10.1109/LSP.2025.3577132

TABLE I
POWER SERIES BASED HARD THRESHOLDING ALGORITHMS

Input: $\mathbf{A}$, $\mathbf{y}, K, q(t)$.
Initialization: K-sparse $\mathbf{x}^0$ (typically $\mathbf{x}^0 = \mathbf{0}$).
while (stopping criteria not met)

- $\mathbf{z}^{n+1} = \mathbf{x}^n + \mathbf{A}^H q(\mathbf{A}\mathbf{A}^H)(\mathbf{y} - \mathbf{A}\mathbf{x}^n)$,
- $\mathbf{x}^{n+1} = \begin{cases} H_K[\mathbf{z}^{n+1}] & \text{// PSIHT,} \\ PS(\mathbf{y}, \mathbf{A}, \mathrm{supp}(H_K[\mathbf{z}^{n+1}])) & \text{// PSHTP.} \end{cases}$

end while
Output: K-sparse vector $\mathbf{x}^*$, $\widehat{S} = \mathrm{supp}(\mathbf{x}^*)$.

II. PROPOSED ALGORITHMS

A. Notations and Assumptions

The standard inner product in $\mathbb{C}^N$ is denoted by $\langle \mathbf{x}, \mathbf{y} \rangle := \mathbf{y}^H \mathbf{x}$ and the standard squared ℓ_2-norm is defined as $\|\mathbf{u}\|_2^2 := \sum_{i=1}^N |u_i|^2$. For a positive integer L, we use $[L]$ to denote the set $\{1, \ldots, L\}$. For a set $S \subseteq [N]$, $\overline{S} := [N] \setminus S$ and $|S|$ denote the complement and the cardinality of S respectively. The support of a vector $\mathbf{u}$, denoted by $\mathrm{supp}(\mathbf{u})$, is the index set of nonzero entries of $\mathbf{u}$. For a matrix $\mathbf{U}$, $\|\mathbf{U}\|_{2\to 2}$ denotes its spectral norm. For a set $\Omega \subseteq [N]$, $(\mathbf{u})_\Omega \in \mathbb{C}^N$ is a vector having entries $[(\mathbf{u})_\Omega]_i = u_i$ if $i \in \Omega$, and other entries set to zeros, and $\mathbf{u}_\Omega \in \mathbb{C}^{|\Omega|}$ is restriction of $\mathbf{u}$ to set Ω. The matrix $\mathbf{A}_\Omega$ is a submatrix of the matrix $\mathbf{A}$ having columns indexed by Ω. Throughout the paper we assume that the target signal $\mathbf{x}$ is indeed K sparse, and an upperbound on K is known. Let $\lambda_1, \ldots, \lambda_m$ be the eigenvalues of $\mathbf{A}\mathbf{A}^H$. We define $\beta_{\min} := \min_{\lambda_i} q(\lambda_i)$ and $\beta_{\max} := \max_{\lambda_i} q(\lambda_i), i \in [m]$. Since $q(\mathbf{A}\mathbf{A}^H)$ is positive definite as seen earlier, we have $\beta_{\min} > 0$ and $\beta_{\max} > 0$. By $PS(\mathbf{y}, \mathbf{A}, T)$, we denote the vector $\mathbf{u} = \arg\min_{\mathbf{z} \in \mathbb{C}^N} \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2^2$ subject to $\mathrm{supp}(\mathbf{z}) = T$, i.e., $\mathbf{A}\mathbf{u}$ is the orthogonal projection of $\mathbf{y}$ on the column space of $\mathbf{A}_T$. Lastly, for two sets A and B, $A\Delta B := (A \setminus B) \cup (B \setminus A)$.

B. Proposed Power Series Based Algorithms

In the proposed PSIHT algorithm, at the $(n+1)^{th}$ iteration, a gradient descent iteration is performed on the objective function to yield $\mathbf{z}^{n+1} = \mathbf{x}^n + \mathbf{A}^H q(\mathbf{A}\mathbf{A}^H)(\mathbf{y} - \mathbf{A}\mathbf{x}^n)$, and then $\mathbf{z}^{n+1}$ is hard thresholded to yield the estimate $\mathbf{x}^{n+1} = H_K[\mathbf{z}^{n+1}]$, where $H_K[\cdot]$ is a hard thresholding operator that retains the top K entries (magnitude-wise) of the argument vector and sets other entries to zero.

In the proposed PSHTP algorithm, the first step is the same as in the case of PSIHT algorithm. This yields $\widetilde{\mathbf{x}}^{n+1} = H_K[\mathbf{x}^n + \mathbf{A}^H q(\mathbf{A}\mathbf{A}^H)(\mathbf{y} - \mathbf{A}\mathbf{x}^n)]$. Next, a pursuit step is performed on the support identified by the first step. Mathematically, we solve the following problem: $\mathbf{x}^{n+1} = \arg\min_{\mathbf{u} \in \mathbb{C}^N} \|\mathbf{y} - \mathbf{A}\mathbf{u}\|_2^2$ subject to $\mathrm{supp}(\mathbf{u}) \subseteq \mathrm{supp}(\widetilde{\mathbf{x}}^{n+1})$. The algorithmic flow chart for the proposed algorithms is presented in Table I.

It is easily seen from Table I that for $q(t) = \mu > 0$, the PSIHT and PSHTP methods boil down to the standard IHT and HTP algorithms (with stepsize μ), while, if $q(t) = \mu(\epsilon + t)^{-1}, \epsilon > |t|, \mu > 0, \epsilon > 0$, the PSIHT and PSHTP algorithms result in the NSIHT and NSHTP algorithms respectively. Lastly, if $q(t) = \mu\, t^{-1}$, PSIHT algorithm becomes identical to the PHT algorithm with stepsize μ.

III. PRELIMINARIES

Definition 3.1: [8] The K-th order restricted isometry constant (RIC) δ_K of a matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ is the smallest number $\delta \geq 0$ such that

$$(1 - \delta)\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \delta)\|\mathbf{x}\|_2^2, \tag{3}$$

holds for all K-sparse vectors $\mathbf{x} \in \mathbb{C}^N$.

The above definition implies that eigenvalues of the matrix $\mathbf{A}_S^H \mathbf{A}_S \in [1 - \delta_s, 1 + \delta_s]$, where $S \subseteq [N]$ is any indexing set with cardinality $|S| \leq s$.

Lemma 3.1: [9], [10] Let $\mathbf{z} \in \mathbb{C}^N$ be a given vector and $\mathbf{x} \in \mathbb{C}^N$ be any K-sparse vector, then

$$\|(\mathbf{x} - H_K(\mathbf{z}))_{\overline{T}}\|_2 \leq \sqrt{2}\|(\mathbf{x} - \mathbf{z})_{S\Delta T}\|_2, \tag{4}$$

$$\|\mathbf{x} - H_K(\mathbf{z})\|_2 \leq \eta\|(\mathbf{x} - \mathbf{z})_{S\cup T}\|_2, \tag{5}$$

where $S = \mathrm{supp}(\mathbf{x}), T = \mathrm{supp}(H_K(\mathbf{z}))$ and $\eta = (\sqrt{5} + 1)/2$.

Lemma 3.2: Let $\mathbf{G} \in \mathbb{C}^{m \times m}$ be a Hermitian matrix and $\mathbf{W} \in \mathbb{C}^{m \times r}$ be any matrix. Then following relations hold:

$$\lambda_{\max}(\mathbf{W}^H\mathbf{G}\mathbf{W}) \leq \lambda_{\max}(\mathbf{G}) \cdot \lambda_{\max}(\mathbf{W}^H\mathbf{W}), \tag{6}$$

$$\text{and} \quad \lambda_{\min}(\mathbf{W}^H\mathbf{G}\mathbf{W}) \geq \lambda_{\min}(\mathbf{G}) \cdot \lambda_{\min}(\mathbf{W}^H\mathbf{W}). \tag{7}$$

Proof: Please see the supplementary material. ■

Lemma 3.3: Let $\mathbf{A} \in \mathbb{C}^{m \times N}, m < N$ and $\mathbf{e} \in \mathbb{C}^m$ be a vector. If $S \subseteq N$ be an indexing set with cardinality $|S| \leq s$, then following is satisfied:

$$\|(\mathbf{A}^H q(\mathbf{A}\mathbf{A}^H)\mathbf{e})_S\|_2 \leq \beta_{\max}\sqrt{1 + \delta_s}\|\mathbf{e}\|_2. \tag{8}$$

Proof: Please see the supplementary material. ■

Lemma 3.4: Let $\mathbf{A} \in \mathbb{C}^{m \times N}$ and $\mathbf{u}, \mathbf{v} \in \mathbb{C}^N$ be any two vectors. If $S := \mathrm{supp}(\mathbf{u}) \cup \mathrm{supp}(\mathbf{v})$, and $T \subseteq [N]$ are two indexing sets, then

$$|\langle \mathbf{u}, (\mathbf{I} - \mathbf{B})\mathbf{v} \rangle| \leq \rho_s\|\mathbf{u}\|_2\|\mathbf{v}\|_2, \tag{9}$$

$$\|[(\mathbf{I} - \mathbf{B})\mathbf{v}]_T\|_2 \leq \rho_{s'}\|\mathbf{v}\|_2, \tag{10}$$

where $\mathbf{B} = \mathbf{A}^H q(\mathbf{A}\mathbf{A}^H)\mathbf{A}, s' = |T \cup \mathrm{supp}(\mathbf{v})|, s = |S|$, and for any positive integer k,

$$\rho_k := \max\{1 - (1 - \delta_k)\beta_{\min}, (1 + \delta_k)\beta_{\max} - 1\}. \tag{11}$$

Proof: We have, $\langle \mathbf{u}, (\mathbf{I} - \mathbf{B})\mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle - \langle \mathbf{A}\mathbf{u}, q(\mathbf{A}\mathbf{A}^H)\mathbf{A}\mathbf{v} \rangle$

$$= \langle \mathbf{u}_S, \mathbf{v}_S \rangle - \langle \mathbf{A}_S\mathbf{u}_S, q(\mathbf{A}\mathbf{A}^H)\mathbf{A}_S\mathbf{v}_S \rangle = \langle \mathbf{u}_S, \mathbf{C}\mathbf{v}_S \rangle,$$

where $\mathbf{C} = \mathbf{I} - \mathbf{A}_S^H q(\mathbf{A}\mathbf{A}^H)\mathbf{A}_S$. Next, by applying Cauchy-Schwarz inequality, $|\langle \mathbf{u}, (\mathbf{I} - \mathbf{B})\mathbf{v} \rangle| \leq \|\mathbf{u}_S\|_2\|\mathbf{C}\mathbf{v}_S\|_2$

$$\leq \|\mathbf{u}_S\|_2\|\mathbf{C}\|_{2\to 2}\|\mathbf{v}_S\|_2. \tag{12}$$

Since $\mathbf{C}$ is a Hermitian matrix, one has

$$\|\mathbf{C}\|_{2\to 2} = \max\{|\lambda_{\max}(\mathbf{C})|, |\lambda_{\min}(\mathbf{C})|\}. \tag{13}$$

Using (7), $\lambda_{\max}(\mathbf{C}) = 1 - \lambda_{\min}(\mathbf{A}_S^H q(\mathbf{A}\mathbf{A}^H)\mathbf{A}_S) \leq 1 - \lambda_{\min}(\mathbf{A}_S^H\mathbf{A}_S) \cdot \lambda_{\min}(q(\mathbf{A}\mathbf{A}^H)) \leq 1 - (1 - \delta_s)\beta_{\min}$, where the last step is a consequence of Definition 3.1 and the fact that $\beta_{\min} > 0$. Similarly, it follows that $\lambda_{\min}(\mathbf{C}) \geq 1 - (1 + \delta_s)\beta_{\max}$. However, since $\lambda_{\min}(\mathbf{C}) \leq \lambda_{\max}(\mathbf{C})$, from (13) we have

$$\|\mathbf{C}\|_{2\to 2} \leq \max\{1 - (1 - \delta_s)\beta_{\min}, (1 + \delta_s)\beta_{\max} - 1\}.$$

Using the above results along with (12), (13), while noting that $\|\mathbf{v}_S\|_2 = \|\mathbf{v}\|_2$ and $\|\mathbf{u}_S\|_2 = \|\mathbf{u}\|_2$ proof of (9) is complete. If $\mathbf{v}' = [(\mathbf{I} - \mathbf{B})\mathbf{v}]_T$, then $\|\mathbf{v}'\|_2^2 = \langle \mathbf{v}', [(\mathbf{I} - \mathbf{B})\mathbf{v}] \rangle \leq \rho_{s'}\|\mathbf{v}'\|_2\|\mathbf{v}\|_2$. Simplifying, proof of (10) is complete. ■

Lemma 3.5: [11] Let $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ be the noisy measurements of the K-sparse signal $\mathbf{x}$. If $T \subseteq [N]$ denotes an index set with $|T| = t$, then $\mathbf{z}^* = PS(\mathbf{y}, \mathbf{A}, T)$ satisfies the following:

$$\|(\mathbf{x} - \mathbf{z}^*)_T\|_2 \leq \delta_{K+t}\|\mathbf{x} - \mathbf{z}^*\|_2 + \|(\mathbf{A}^H\mathbf{e})_T\|_2. \tag{14}$$

Further, if $\delta_{K+t} < 1$, then

$$\|\mathbf{x} - \mathbf{z}^*\|_2 \leq C_1\|(\mathbf{x})_{\overline{T}}\|_2 + C_2\|(\mathbf{A}^H\mathbf{e})_T\|_2, \tag{15}$$

where $C_1 = \frac{1}{\sqrt{1-\delta_{K+t}^2}}$ and $C_2 = \frac{1}{1-\delta_{K+t}}$.

IV. Analysis of Proposed Algorithms

For both the proposed algorithms of Table I, we show that the estimate $\mathbf{x}^{n+1}$ at the $(n+1)^{th}$ iteration satisfies,

$$\|\mathbf{x} - \mathbf{x}^{n+1}\|_2 \leq \psi\|\mathbf{x} - \mathbf{x}^n\|_2 + \phi\|\mathbf{e}\|_2, \tag{16}$$

where ψ, ϕ depend only on $\beta_{\min}, \beta_{\max}$, and the RICs of $\mathbf{A}$ of various orders. Proceeding backwards recursively, (16) can be written as $\|\mathbf{x} - \mathbf{x}^{n+1}\|_2 \leq \psi^{n+1}\|\mathbf{x} - \mathbf{x}^0\|_2 + \frac{1-\psi^{n+1}}{1-\psi}\phi\|\mathbf{e}\|_2$. For convergence, it is necessary and sufficient to have $\psi < 1$, for which $\lim_{n\to\infty}\|\mathbf{x} - \mathbf{x}^{n+1}\|_2 \leq \frac{1}{1-\psi}\phi\|\mathbf{e}\|_2$. Our generalized analysis allows us to provide relaxed condition on the RICs for NSHTP algorithm, namely, the results are now provided in terms of $\delta_{3K} < 1/\sqrt{3}$ compared to $\delta_{3K} < 1/2$ in [7].

A. Analysis of the PSIHT

Theorem 4.1: If the matrix $\mathbf{A}$ satisfies $\delta_{3K} < 1/\eta \simeq 0.618$ and $q(t)$ is chosen such that $\beta_{\min}$ and $\beta_{\max}$ satisfy

$$\frac{1 - \frac{1}{\eta}}{1 - \delta_{3K}} < \beta_{\min} \leq \beta_{\max} < \frac{1 + \frac{1}{\eta}}{1 + \delta_{3K}}, \tag{17}$$

then the estimate produced by the PSIHT algorithm $\{\mathbf{x}^n\}_{n\geq 1}$, satisfies (16) with $\psi = \eta\rho_{3K} < 1$ and $\phi = \eta\beta_{\max}\sqrt{1 + \delta_{2K}}$.

Proof: We have, $\mathbf{x}^{n+1} = H_K[\mathbf{z}^{n+1}]$, where $\mathbf{z}^{n+1} = \mathbf{x}^n + \mathbf{A}^H q(\mathbf{A}\mathbf{A}^H)(\mathbf{y} - \mathbf{A}\mathbf{x}^n)$. Using Lemma 3.1 and defining $\mathbf{B} := \mathbf{A}^H q(\mathbf{A}\mathbf{A}^H)\mathbf{A}$, $T := S \cup S^{n+1}$, $S = \text{supp}(\mathbf{x})$, we have,

$$\|\mathbf{x} - \mathbf{x}^{n+1}\|_2 \leq \eta\|(\mathbf{z}^{n+1} - \mathbf{x})_T\|_2 \tag{18}$$

$$\begin{aligned}
&= \eta\|(\mathbf{x}^n + \mathbf{A}^H q(\mathbf{A}\mathbf{A}^H)(\mathbf{y} - \mathbf{A}\mathbf{x}^n) - \mathbf{x})_T\|_2 \\
&\leq \eta\|((\mathbf{I} - \mathbf{B})(\mathbf{x} - \mathbf{x}^n)))_T\|_2 + \eta\|(\mathbf{A}^H q(\mathbf{A}\mathbf{A}^H)\mathbf{e})_T\|_2 \\
&\leq \eta\rho_{3K}\|(\mathbf{x} - \mathbf{x}^n)\|_2 + \eta\beta_{\max}\sqrt{1 + \delta_{2K}}\|\mathbf{e}\|_2,
\end{aligned}$$

which results in

$$\|\mathbf{x} - \mathbf{x}^{n+1}\|_2 \leq \psi\|(\mathbf{x} - \mathbf{x}^n)\|_2 + \phi\|\mathbf{e}\|_2. \tag{19}$$

Here, we used the fact that $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ and applied triangle inequality in the first step. In the second step, we used (10) and (8), for the first and second terms respectively. To ensure that $\psi < 1$, one requires that $\beta_{\min}$ and $\beta_{\max}$ satisfy $\eta(1 - (1 - \delta_{3K})\beta_{\min}) < 1$, or equivalently, $\beta_{\min} > (1 - \frac{1}{\eta})/(1 - \delta_{3K})$, and $\eta(1 - (1 + \delta_{3K})\beta_{\max}) > -1$, or, equivalently, $\beta_{\max} < (1 + \frac{1}{\eta})/(1 + \delta_{3K})$.

Together with the fact that $\beta_{\max} \geq \beta_{\min}$, these lead to (17). Note that for (17) to hold, we require, $(1 + \frac{1}{\eta})/(1 + \delta_{3K}) > (1 - \frac{1}{\eta})/(1 - \delta_{3K})$, which is satisfied if $\delta_{3K} < 1/\eta$. This completes the proof. ∎

B. Analysis of the PSHTP Algorithm

Theorem 4.2: If the sensing matrix satisfies $\delta_{3K} < 1/\sqrt{3}$ and $q(t)$ is chosen to satisfy

$$\frac{1 - \sqrt{\frac{1-\delta_{3K}^2}{2}}}{1 - \delta_{3K}} < \beta_{\min} \leq \beta_{\max} < \frac{1 + \sqrt{\frac{1-\delta_{3K}^2}{2}}}{1 + \delta_{3K}}, \tag{20}$$

then the estimate produced by PSHTP algorithm $\{\mathbf{x}^n\}_{n\geq 1}$ satisfies (16) with the coefficients ψ and ϕ given by

$$\psi = \sqrt{2}\rho_{3K}/\sqrt{1 - \delta_{2K}^2} < 1, \tag{21}$$

$$\phi = \frac{\sqrt{2}\beta_{\max}\sqrt{1 + \delta_{2K}}}{\sqrt{1 - \delta_{2K}^2}} + \frac{\sqrt{1 + \delta_K}}{1 - \delta_{2K}}. \tag{22}$$

Proof: Due to page limitations, we have provided the proof in the accompanying supplementary material. ∎

V. Implications of the Proposed Approach

Given the sensing matrix $\mathbf{A}$ which satisfies the conditions: $\delta_{3K} < 1/\eta \simeq 0.618$ (for the PSIHT algorithm) and $\delta_{3K} < 1/\sqrt{3}$ (for the PSHTP algorithm) with high probability, we first find out the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$ of $\mathbf{A}\mathbf{A}^H$, then design a $q(t)$ so that $B < q(\lambda_i) < C, i = 1, 2, \ldots, m$, where, for the the PSIHT algorithm, $B = (1 - \frac{1}{\eta})/(1 - \delta_{3K}), C = (1 + \frac{1}{\eta})/(1 + \delta_{3K})$ and for the PSHTP algorithm, $B = (1 - \sqrt{\frac{1-\delta_{3K}^2}{2}})/(1 - \delta_{3K}), C = (1 + \sqrt{\frac{1-\delta_{3K}^2}{2}})/(1 + \delta_{3K})$. A practical choice here will be to choose $q(t)$ that is monotonically increasing (resp. decreasing) in t, as, in such case, $\beta_{\min} := \min_{\lambda_i} q(\lambda_i) = q(\lambda_{\min})$ (resp. $\beta_{\max} = q(\lambda_{\min})$) and $\beta_{\max} := \max_{\lambda_i} q(\lambda_i) = q(\lambda_{\max})$ (resp. $\beta_{\min} = q(\lambda_{\max})$), where $\lambda_{\min}$ and $\lambda_{\max}$ are the minimum and maximum eigenvalues of $\mathbf{A}\mathbf{A}^H$ respectively. We have already seen some examples of such $q(t)$ at the conclusion of Section II. A few more examples of such monotonic $q(t)$ can be $q(t) = c_0 + c_1 t$, $q(t) = \mu e^{-\lambda t}$ etc.

Next we show that the proposed convergence analysis establishes wider ranges for algorithm parameters μ and ϵ, and reduced ψ for the NSIHT and NSHTP algorithms which implies faster convergence as compared to [7]. We consider the power series $q(t) = \mu(\epsilon + t)^{-1} = (\mu/\epsilon)\sum_{i=0}^{\infty}(-1)^i(t/\epsilon)^i$, where $\mu > 0, \epsilon > |t|$. In terms of the matrix power series,

$$q(\mathbf{A}\mathbf{A}^H) = \frac{\mu}{\epsilon}\sum_{i=0}^{\infty}(-1)^i(\mathbf{A}\mathbf{A}^H/\epsilon)^i, \epsilon > \lambda_{\max}(\mathbf{A}\mathbf{A}^H) = \sigma_h^2,$$

where σ_h denotes the highest singular value of $\mathbf{A}$. As seen earlier, both the NSIHT and NSHTP algorithms result from this choice of $q(t)$ in the proposed treatment.

For comparison of the convergence conditions, we consider the noiseless setting for simplicity (though the same arguments apply to the noisy setting also). Here, we focus only on the NSIHT algorithm, though the arguments are equally valid with appropriate modifications for the NSHTP algorithm as well. For NSIHT (also for NSHTP), the analysis provided by us, and by the authors in [7] first derive the following expression:

$$\|\mathbf{x} - \mathbf{x}^{n+1}\|_2 \leq \psi\|\mathbf{x} - \mathbf{x}^n\|_2, \tag{23}$$

where ψ is dependent on $\delta_{3K}, \mu, \epsilon$ and then derive conditions to maintain $0 \leq \psi < 1$. In [7], ψ was obtained as $\psi = \sqrt{3}(\delta_{3K} + \sigma_h^2 - \frac{\mu\sigma_h^2}{\epsilon+\sigma_h^2})$ and to ensure $0 \leq \psi < 1$, the following ranges of ϵ, μ were worked out:

$\epsilon > \max\{\sigma_h^2, (\frac{\sigma_h^2-\sigma_l^2}{\frac{1}{\sqrt{3}}-\delta_{3K}} - 1)\sigma_h^2\}$ and $a < \mu \le b$, where $a = \epsilon + \sigma_h^2 - (\frac{1}{\sqrt{3}} - \delta_{3K})\frac{\epsilon+\sigma_h^2}{\sigma_h^2}, b = \epsilon + \sigma_l^2$ with σ_l being the lowest singular value of $\mathbf{A}$. Since $\sigma_h \ge \sigma_l$, we have, $\frac{\mu}{\epsilon+\sigma_h^2} \le \frac{\mu}{\epsilon+\sigma_l^2} \le 1$, where the last inequality follows from the above upper bound on μ. From the above expression of ψ, it then follows that to ensure $0 \le \psi < 1$, we need to satisfy $\delta_{3K} < 1/\sqrt{3}$.

The proposed analysis, however, shows that both the NSIHT and NSHTP algorithms enjoy parameter ranges that are more relaxed, and convergence rates that are faster than given in [7]. Using the fact that if $m < N$ then $\sigma_h^2 > 1 - \delta_{3K}$, (a formal proof for this is presented in the supplementary material), it establishes the following :

1) A wider range of μ for guaranteed convergence, namely,

$$a' < \frac{\mu}{\epsilon + \sigma_h^2} \leqslant \frac{\mu}{\epsilon + \sigma_l^2} < b', \tag{24}$$

where, $a' = \frac{1-\frac{1}{\sqrt{3}}}{1-\delta_{3K}}$ and $b' = \frac{1+\frac{1}{\sqrt{3}}}{1+\delta_{3K}}$, with $(\epsilon + \sigma_h^2)a' < a$ and $(\epsilon + \sigma_l^2)b' > b$.

2) A wider range for ϵ.

3) $\psi_{new} < \psi_{old}$, with the notations ψ_{old} and ψ_{new} to denote ψ as per [7] (also given above) and as per this paper, respectively.

Proof of the above claims is given in the accompanying supplementary material.

VI. Simulations

Due to page limitation, we restrict the simulations here to the noiseless setting only (the noisy case is considered in the supplementary material). For this, we consider the following choices of $q(t)$:

1) Exponential: $q(t) = \mu \exp(-\lambda t)$.

2) Unit degree polynomial: $q(t) = \mu(1 - t/\epsilon)$.

The support of the target K-sparse signal $\mathbf{x}$ is generated from $[N]$ following a uniform distribution without repetition and the nonzero entries of $\mathbf{x}$ are taken as i.i.d according to:

a) Normal distribution (zero mean, unit variance), and

b) Bernoulli ($x_i = \pm 1$, equi-probable).

A maximum of 100 iterations are permitted for all algorithms and if $\mathrm{MSD}(n) := \frac{\|\mathbf{x}^n - \mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2}$ falls below 10^{-4}, then the algorithms are halted and recovery is said to be successful. Ensemble averaging is carried out by taking 100 independent realizations of the problem. Entries of $\mathbf{A}$ are i.i.d. Gaussian with zero mean and variance $\frac{1}{m}$ following [2], [11].

The corresponding PSIHT (resp. PSHTP) algorithm for the cases in 1) and 2) above is henceforth referred to as Expo-IHT (resp. Expo-HTP) and Poly-IHT (resp. Poly-HTP) respectively. For comparative assessment, we also consider the following well known methods with best performing step-size selected: IHT ($\mu = 1/3$;[2]), HTP ($\mu = 1.6$ for Gaussian distribution and $\mu = 1$ for Bernoulli distribution;[2]), PHT ($\mu = 2$;[6]), CoSaMP [12], SP [13], NSIHT and NSHTP ($\mu = 5, \epsilon = \max\{\sigma_h^2 + 1, \mu - \sigma_l^2\}$ [7]). The parameters μ, ϵ and λ chosen for the functions $q(t)$ under consideration are:

- Expo-IHT algorithm: $\mu = 1.6, \lambda = 0.25$,
- Expo-HTP algorithm: $\mu = 2.75, \lambda = 0.15$,
- Poly-IHT algorithm: $\mu = 0.8, \epsilon = 16$,
- Poly-HTP algorithm: $\mu = 1.75, \epsilon = 10$.

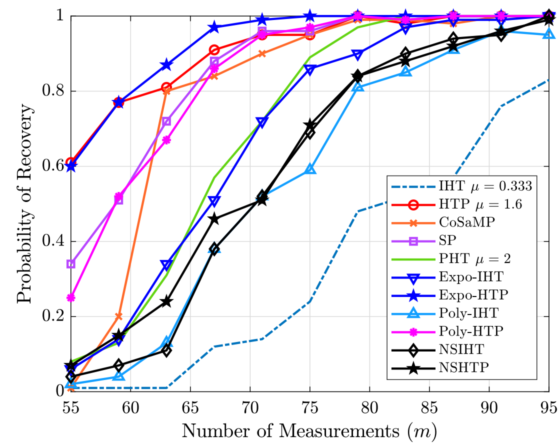

Fig. 1. Recovery performance of various algorithms for Gaussian sparse vectors with $N = 200, K = 20$.

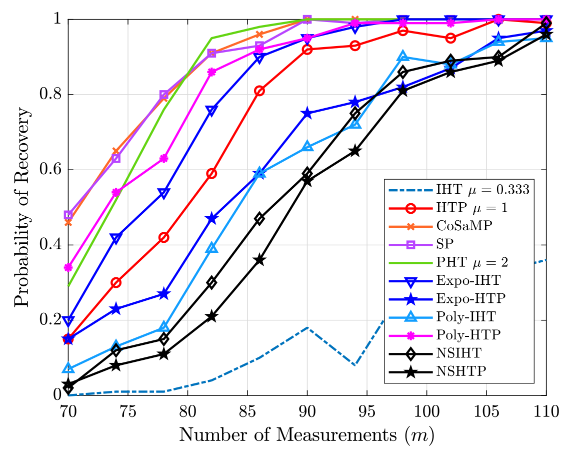

Fig. 2. Recovery performance of various algorithms for Bernoulli sparse vectors with $N = 200, K = 20$.

In Figs. 1 and 2, we plot the probability of recovery against number of measurements for the aforementioned algorithms, for Gaussian and Bernoulli sparse vectors respectively. It can be clearly seen that a single algorithm does not always outperform others. For example, Expo-HTP and HTP algorithms offer best performance for Gaussian distribution, whereas the PHT, SP, CoSaMP and Poly-HTP algorithms distinctly outperform the HTP and Expo-HTP algorithm for Bernoulli distribution. This clearly suggests that for a problem under consideration, one may vary the function $q(t)$ to yield varying performances and then select the best. For reproducing the figures presented in this paper, one may find the MATLAB codes at the link: https://drive.google.com/file/d/1k5e6GOmqC_QOI7GQMasCMGQwmJerp9AS/view?usp=drive_link.

Acknowledgment

The authors thank the anonymous reviewer-2 for suggesting an interesting way of showing that $\sigma_h^2 = 1 - \delta_{3K}$ is not only unlikely, but rather impossible to hold unless $m \ge N$.

REFERENCES

[1] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, 2009.

[2] S. Foucart, "Hard thresholding pursuit: An algorithm for compressive sensing," *SIAM J. Numer. Anal.*, vol. 49, no. 6, pp. 2543–2563, 2011.

[3] Y. Wang, Z. He, G. Zhang, and J. Wen, "Improved sufficient conditions based on RIC of order 2s for IHT and HTP algorithms," *IEEE Signal Process. Lett.*, vol. 30, pp. 668–672, 2023.

[4] L.-J. Xie, "Improved RIC bounds in terms of δ_{2s} for hard thresholding-based algorithms," *IEEE Signal Process. Lett.*, vol. 30, pp. 21–25, 2023.

[5] S. Li, Y. Liu, and T. Mi, "Fast thresholding algorithms with feedbacks for sparse signal recovery," *Appl. Comput. Harmon. Anal.*, vol. 37, no. 1, pp. 69–88, 2014.

[6] 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.*, vol. 24, no. 7, pp. 7621–7630, Jul. 2023.

[7] N. Meng and Y.-B. Zhao, "Newton-step-based hard thresholding algorithms for sparse signal recovery," *IEEE Trans. Signal Process.*, vol. 68, pp. 6594–6606, 2020.

[8] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. New York, NY, USA: Birkhäuser, 2013.

[9] J. Shen and P. Li, "A tight bound of hard thresholding," *J. Mach. Learn. Res.*, vol. 18, no. 208, pp. 1–42, 2018.

[10] Y. Zhao and Z. Luo, "Improved RIP-based bounds for guaranteed performance of two compressed sensing algorithms," *Sci. China Math.*, vol. 66, no. 5, pp. 1123–1140, 2023.

[11] J.-L. Bouchot, S. Foucart, and P. Hitczenko, "Hard thresholding pursuit algorithms: Number of iterations," *Appl. Comput. Harmon. Anal.*, vol. 41, no. 2, pp. 412–435, 2016.

[12] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.

[13] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.