Consecutive Support Matching Induced Parameter Tuning Accelerates Momentum Iterative Hard Thresholding

Samrat Mukhopadhyay*, Debasmita Mukherjee†
*Department of Electronics Engineering, IIT (ISM) Dhanbad
Email: samrat@iitism.ac.in
†Department of Computer Science and Information Technology, Techno Bengal Institute of Technology
Email: debasmitam97@gmail.com

***Abstract*—Momentum-based acceleration of iterative hard thresholding (IHT) can dramatically speed up sparse signal recovery, but its effectiveness hinges on careful parameter tuning which is complicated by frequent support changes. We propose `CosMIHT` (Consecutive Support Matching Induced Momentum IHT): whenever two consecutive iterates share the same support, it estimates the extreme eigenvalues of the support-restricted Gram matrix via a lightweight power method and switches to the corresponding optimal heavy-ball parameters, automatically interpolating between cautious behavior during support discovery and near-optimal acceleration after support identification. Under standard RIP assumptions, we develop a two-phase convergence theory establishing linear error contraction and a finite lock-in time bound in the wandering phase, and near-optimal accelerated convergence with high probability in the lock-in phase. Numerical experiments confirm that `CosMIHT` achieves faster convergence than state-of-the-art iterative sparse recovery methods without compromising recovery performance.**

***Index Terms*—Iterative hard thresholding; consecutive support matching; heavy-ball; Restricted Isometry Property (RIP); support identification; compressed sensing.**

I. Introduction

We consider the sparse recovery problem: given a measurement vector $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}^\star + \boldsymbol{e} \in \mathbb{R}^M$, where $\boldsymbol{A} \in \mathbb{R}^{M\times N}$ is a known sensing matrix and $\boldsymbol{e}$ is measurement noise, recover the K-sparse vector $\mathbf{x}^\star \in \mathbb{R}^N$. This amounts to solving

$$\min_{\boldsymbol{x}\in\mathbb{R}^N} \ \tfrac{1}{2}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{x}\|_0 \le K. \tag{1}$$

Although (1) is NP-hard in general, it can be solved efficiently under regularity conditions on $\boldsymbol{A}$, notably the restricted isometry property (RIP) [1]. Among the many algorithmic approaches, including convex relaxations such as Basis Pursuit [2] and LASSO [3], and greedy methods such as OMP [4], CoSaMP [5], and SP [6], iterative hard thresholding (IHT) [7] stands out for its simplicity: each iteration requires only a matrix-vector product and a hard thresholding step, yet IHT achieves near-optimal recovery guarantees under RIP.

Since IHT is essentially projected gradient descent onto the set of K-sparse vectors, a natural question is whether it can benefit from momentum-based acceleration. In smooth convex optimization over convex sets, Polyak's heavy-ball method [8] and Nesterov's accelerated gradient [9] are known to achieve optimal first-order convergence rates, substantially outperforming vanilla gradient descent. This has motivated several accelerated IHT variants, such as, Normalized Iterative hard Thresholding (NIHT) [10], Fast Lipschitz Iterative Hard Tresholding (FLIHT) [11], Heavy Ball Hard Thresholding (HBHT) [12] and Momentum IHT (MIHT) [13]. Among these, MIHT has demonstrated particularly strong empirical performance by incorporating an exponentially weighted sum of all past gradient directions as the descent direction, rather than only the current gradient direction, which makes it conceptually analogous to the heavy-ball method.

Despite its empirical success, MIHT has no clear guideline on how to tune the momentum parameter β and the step size parameter μ. Although one might attempt to directly use the optimal parameter tuning rule of heavy-ball method for strongly convex quadratics, the task is complicated by the frequent switching of the effective Hessian as the *active support changes across iterations*. When the support set $S_t = \operatorname{supp}(\boldsymbol{x}^t)$ shifts, the restricted Hessian $\boldsymbol{A}_{S_t}^\top \boldsymbol{A}_{S_t}$ changes abruptly, invalidating any parameter choice tuned to the previous support.

To this end, we propose **`CosMIHT`** (Consecutive Support Matching Induced Momentum IHT), which integrates an adaptive parameter tuning rule into MIHT. The core mechanism is simple: the algorithm starts with conservative parameter values which are used until two consecutive iterates share the same support, distinct from the one before. Following this consecutive support matching, `CosMIHT` estimates the extreme eigenvalues of the restricted Gram matrix $\boldsymbol{A}_{S_t}^\top \boldsymbol{A}_{S_t}$ via a lightweight power method and switches to the optimal heavy-ball parameters using these eigenvalue estimates. These updated parameters are kept in use until again two consecutive supports match, albeit, distinct from the one before. This allows the algorithm to interpolate automatically between safe MIHT-like behavior during support discovery and near-optimal accelerated convergence after support identification.

Our specific contributions are as follows:

(i) **Algorithm design.** We propose `CosMIHT` (Algorithm 2), an IHT variant where the momentum and step-size parameters are updated adaptively based on consecutive

support matching. The only hyperparameters requiring manual selection are the conservative defaults (μ, β) and the number of power iterations r.

(ii) **Two-phase convergence theory.** We establish a rigorous two-phase convergence analysis under RIP. We prove a global linear contraction (upto noise floor) of CosMIHT in the *wandering phase* during support discovery (Proposition 1). Then we show that under a sufficient signal-to-noise ratio the algorithm locks-in to the correct support after a finite *lock-in* time, (Theorem 1), following it enters the *lock-in phase* where the dynamics reduces to that of the classical heavy-ball method (Section IV-A3).

(iii) **Power method accuracy analysis.** We analyze the accuracy of the eigenvalue estimates produced by the power method (Theorem 2), providing high-probability bounds on the number of power iterations r needed to ensure that the tuned parameters ensure near-optimal convergence rate in the lock-in phase.

(iv) **Numerical validation.** We corroborate the theoretical findings with simulations comparing CosMIHT against IHT, FLIHT, and MIHT in both noiseless and noisy settings, demonstrating improved recovery probability and faster convergence.

A. Paper Organization

Section II collects notation and preliminary results. Section III presents the CosMIHT algorithm. Section IV contains the convergence analysis, including wandering phase bounds, lock-in time guarantees, and lock-in phase convergence rates as well as a discussion on the computational complexity. Section V presents numerical experiments.

II. Preliminaries

Notation: Bold lowercase/uppercase denote vectors/matrices. Transpose is $(\cdot)^\top$. For an index set $S \subseteq [N] := \{1, \dots, N\}$, $\boldsymbol{x}_S$ is the restriction, and $\mathrm{supp}(\boldsymbol{x})$ the support. $\mathcal{H}_K(\boldsymbol{z})$ keeps the K largest-magnitude entries of $\boldsymbol{z}$ and zeros the rest.

Assumption 1 (Restricted Isometry Property)**.** The matrix $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ satisfies the restricted isometry property (RIP) of order K, with a restricted isometry constant (RIC) δ_K if the following holds:

$$\delta_K := \max_{S \subset [N]: |S| \le K} \left\| \boldsymbol{I}_K - \boldsymbol{A}_S^\top \boldsymbol{A}_S \right\|_{2\to 2} < 1, \tag{2}$$

where $\|\cdot\|_{2\to 2}$ denotes the operator norm.

Lemma 1 (Theorem 1 [14])**.** *Let $\boldsymbol{x}^\star$ be K-sparse with support S. For any $\boldsymbol{z}$, $\|\mathcal{H}_K(\boldsymbol{z}) - \boldsymbol{x}^\star\|_2 \le \omega \|(\boldsymbol{z} - \boldsymbol{x}^\star)_{S\cup T}\|_2$ where $T = \mathrm{supp}(\mathcal{H}_K(\boldsymbol{z}))$, and $\omega = \frac{1+\sqrt{5}}{2}$ is the hard-thresholding expansion constant.*

III. Proposed Algorithm

We propose CosMIHT (Algorithm 2) by integrating a *consecutive support matching induced parameter tuning step* with the MIHT computations. Specifically, at step t, we first estimate the support S_t of the current iterate $\boldsymbol{x}^t$ and update the momentum parameter β_t and step size μ_t based on S_t, S_{t-1}, and S_{t-2}, as follows. If there is a support change at step t, i.e., $S_{t-1} \neq S_t$, then the supports have not stabilized yet and we retain the previous parameters $\beta_t = \beta_{t-1}, \mu_t = \mu_{t-1}$. If $S_t = S_{t-1}$, then there are two cases to consider. If $S_{t-1} \neq S_{t-2}$, i.e., the support just stabilized at t, then we update the parameters β_t and μ_t to the optimal parameters for Polyak's heavy-ball method (see lines 5–8 of Algorithm 2). If $S_{t-1} = S_{t-2}$, then the support stabilized before $t-1$, and CosMIHT reuses the parameters from step $t-1$.

Algorithm 1 POWER Method

Require: $\boldsymbol{B} \in \mathbb{R}^{K\times K}$, $\boldsymbol{b} \in \mathbb{R}^K$, $r \in \mathbb{Z}_+$
1: **Initialize** $\boldsymbol{b}_0 = \boldsymbol{b}$
2: **for** $k = 1, 2, \cdots, r$ **do**
3: $\quad \boldsymbol{b}_k = \boldsymbol{B}\boldsymbol{b}_{k-1}$
4: **end for**
5: **Return** $\frac{\boldsymbol{b}_r}{\|\boldsymbol{b}_r\|_2}$

Algorithm 2 CosMIHT: Consecutive Support Matching Induced Momentum Iterative Hard Thresholding

Require: $\boldsymbol{y}, \boldsymbol{A}$, sparsity K, stepsize $\mu > 0$, momentum $\beta \in [0, 1)$, no. of power iterations r, time horizon T
1: **Initialize:** $\boldsymbol{x}^0 = \boldsymbol{0}$, $\boldsymbol{v}^0 = \boldsymbol{0}, \boldsymbol{b}_0 = \frac{\boldsymbol{g}}{\|\boldsymbol{g}\|_2}$, where $\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, $S_{-1} = \emptyset$, $\beta_{-1} = \beta$, $\mu_{-1} = \mu$
2: **for** $t = 0, 1, 2, \dots, T-1$ **do**
3: $\quad S_t = \mathrm{supp}(\boldsymbol{x}^t)$
Consecutive support matching induced parameter tuning
4: $\quad$ **if** $t > 0$, $S_t = S_{t-1}$ and $S_{t-1} \neq S_{t-2}$ **then**
5: $\quad\quad \boldsymbol{b}_{t,r} = \mathrm{POWER}(\boldsymbol{A}_{S_t}^\top \boldsymbol{A}_{S_t}, \boldsymbol{b}_0, r)$
6: $\quad\quad \boldsymbol{c}_{t,r} = \mathrm{POWER}(\|\boldsymbol{A}_{S_t}\|_F^2 \boldsymbol{I}_K - \boldsymbol{A}_{S_t}^\top \boldsymbol{A}_{S_t}, \boldsymbol{b}_0, r)$
7: $\quad\quad l_t = \|\boldsymbol{A}_{S_t}\boldsymbol{c}_{t,r}\|_2^2$, $L_t = \|\boldsymbol{A}_{S_t}\boldsymbol{b}_{t,r}\|_2^2$
8: $\quad\quad \mu_t \leftarrow \frac{4}{(\sqrt{L_t}+\sqrt{l_t})^2}$, $\beta_t \leftarrow \left(\frac{\sqrt{L_t}-\sqrt{l_t}}{\sqrt{L_t}+\sqrt{l_t}}\right)^2$
9: $\quad$ **else**
10: $\quad\quad \beta_t \leftarrow \beta_{t-1}$, $\mu_t \leftarrow \mu_{t-1}$
11: $\quad$ **end if**
12: $\quad \boldsymbol{g}^t = -\boldsymbol{A}^\top(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^t)$
13: $\quad \boldsymbol{v}^{t+1} = \beta_t \boldsymbol{v}^t - \mu_t \boldsymbol{g}^t$
14: $\quad \boldsymbol{x}^{t+1} \leftarrow \mathcal{H}_K(\boldsymbol{x}^t + \boldsymbol{v}^{t+1})$
15: **end for**
16: **Output:** $\boldsymbol{x}^T$

IV. Theoretical Results

A. Convergence analysis of CosMIHT

We begin this section with a convergence analysis of CosMIHT. Recall that we aim to solve the optimization problem

$$\min \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 \text{ s.t. } \|\boldsymbol{x}\|_0 \le K, \quad (P_1) \tag{3}$$

where we adopt the measurement model $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}^\star + \boldsymbol{e}$ where $\boldsymbol{x}^\star$ is K-sparse with support $S^\star$. It can be shown that under the following signal-to-noise ratio condition,

$$\frac{\|\boldsymbol{x}^\star\|_2}{\|\boldsymbol{e}\|_2} > \frac{2}{1 - \delta_{2K}} \tag{4}$$

the optimization problem (P_1) has a unique K-sparse solution $\widehat{\boldsymbol{x}}$, supported on $S^\star$, such that

$$\widehat{\boldsymbol{x}}_{S^\star} = \boldsymbol{x}^\star_{S^\star} + \boldsymbol{A}^\dagger_{S^\star}\boldsymbol{e}. \tag{5}$$

See the technical report [15] for a proof. The rest of the analysis aims to establish convergence of the sequence $\{\boldsymbol{x}^t\}$ generated by CosMIHT to $\widehat{\boldsymbol{x}}$.

We divide the convergence analysis into two phases. The first phase is termed the **wandering phase**, where essentially the algorithm wanders through different supports, and the second phase is referred to as the **lock-in phase** where the algorithm locks in to a fixed support. To define these phases mathematically, we first define the *lock-in* time as below:

$$T_0 = \arg\min \{t \geq 1 : S_\tau = S^\star,\ \forall \tau \geq t\}. \tag{6}$$

Using this definition, the two phases are equivalent to the time intervals $I_i,\ i = 1, 2$, defined as below:

$$I_1 = \{1 \leq t \leq T_0 - 1\}, \qquad (\textbf{Wandering Phase}) \tag{7}$$

$$I_2 = \{T_0 \leq t \leq T\}. \qquad (\textbf{Lock-in Phase}) \tag{8}$$

We now analyze the CosMIHT algorithm for each of these phases.

1) Wandering phase convergence analysis of CosMIHT: In this phase, as the support is not stabilized, CosMIHT undergoes frequent support switching. By combining the analysis of MIHT [13] with the parameter selection rule of CosMIHT (Lines 5–10 in Algorithm 2), we state a result which analyze the evolution of $\boldsymbol{x}^t$ and $\boldsymbol{v}^t$ in this phase. To state the result, we use the following notations:

$$\gamma = \max\left\{\frac{1}{1-\delta_{3K}}, \mu\right\},\ \alpha = \max\left\{\delta_{3K}^2, \beta\right\}, \tag{9}$$

$$\rho = \max\left\{|\mu - 1| + \mu\delta_{3K}, \frac{2\delta_{3K}}{1-\delta_{3K}}\right\},\ \xi = \omega\gamma\sqrt{1+\delta_{3K}}. \tag{10}$$

Proposition 1. *Under the conditions*

$$\delta_{3K} < \frac{1}{2\omega+1},\ \frac{\omega\gamma\sqrt{1+\delta_{3K}}}{1-\omega\rho} < \frac{1-\alpha}{\alpha},\ \mu < \frac{1}{1-\delta_{3K}}, \tag{11}$$

CosMIHT satisfies

$$E_t \leq \frac{\|\widehat{\boldsymbol{x}}\|_2}{2}(\lambda_1^t(1-\sin\theta) + \lambda_2^t(1+\sin\theta)) + \frac{\xi\,\|\boldsymbol{e}\|_2\,((\omega-1)\alpha+1)}{(1-\lambda_1)(1-\lambda_2)}, \tag{12}$$

$$V_t \leq \frac{\|\widehat{\boldsymbol{x}}\|_2\, b\cos\theta}{2\omega\alpha}(\lambda_1^t - \lambda_2^t) + \frac{\xi\,\|\boldsymbol{e}\|_2\cos\theta}{(1-\lambda_1)(1-\lambda_2)}\left[\gamma\sqrt{1+\delta_{3K}} - \omega\rho + \frac{\alpha+1}{2}\right], \tag{13}$$

$$\lambda_2 < 0 < \lambda_1 < 1, \tag{14}$$

where $\tan\theta = \frac{\alpha-\omega\rho}{2b}$, $b = \sqrt{\omega\alpha\gamma\sqrt{1+\delta_{3K}}}$ *and*

$$\lambda_i = \omega\rho + b(\tan\theta + (-1)^{i-1}\sec\theta),\ i = 1, 2. \tag{15}$$

Proof. Please refer to the technical report [15] for a proof. □

2) Lock-in Time analysis of CosMIHT: We now show that if the noise is sufficiently small, i.e., if $\|\boldsymbol{e}\|_2$ is small enough, then CosMIHT always locks in to the correct support $S^\star$ after a finite number of iterations.

Theorem 1. *Under the condition* (11)*, the lock-in time is upper bounded as below:*

$$T_{\text{lock-in}} \leq \left\lceil \frac{\ln\left(\frac{\sqrt{2}\|\boldsymbol{x}^\star\|_2\left(\rho + \frac{b\cos\theta}{\omega}\right)}{x^\star_{\min} - \xi\|\boldsymbol{e}\|_2 \upsilon}\right)}{\ln(1/\lambda_1)} \right\rceil, \tag{16}$$

where

$$\upsilon = \frac{\rho}{(1-\lambda_1)(1-\lambda_2)}\Bigg[(\omega-1)\alpha + 1 + \alpha\cos\theta\left(\gamma\sqrt{1+\delta_{3K}} - \omega\rho + \frac{\alpha+1}{2}\right)\Bigg]. \tag{17}$$

Remark 1. Since $T_{\text{lock-in}}$ is essentially $\mathcal{O}\left(\ln\left(\frac{\|\boldsymbol{x}^\star\|_2}{x^\star_{\min}}\right)\right)$, and since $\boldsymbol{x}^\star$ with subGaussian non-zero entries satisfies $\frac{\|\boldsymbol{x}^\star\|_2}{x^\star_{\min}} \approx K^{3/2}$ with high probability, $T_{\text{lock-in}}$ is approximately $\mathcal{O}(\ln K)$ for sparse vectors with subGaussian non-zero entries.

Proof. The proof is postponed to Appendix A. □

3) Lock-in phase convergence analysis of CosMIHT: Theorem 1 guarantees that CosMIHT identifies the correct support in finite time. Once CosMIHT locks in to the correct support $S^\star$ (which is guaranteed under sufficient signal-to-noise ratio as given in Theorem 1), the parameters β_t, μ_t are fixed, and the trajectory of CosMIHT becomes identical to that of the heavy-ball method, with the Hessian $\boldsymbol{A}_{S^\star}^\top \boldsymbol{A}_{S^\star}$.

To conduct a convergence analysis of CosMIHT in the lock-in phase, we first look at the update steps constrained to $S^\star$. We first denote $\widehat{\mu} = \mu_{T_{\text{lock-in}}}$ and $\widehat{\beta} = \beta_{T_{\text{lock-in}}}$. Now note that for $t \geq T_{\text{lock-in}}$, the support of $\boldsymbol{x}^t$ is $S^\star$, so that

$$\boldsymbol{x}^{t+1}_{S^\star} = \boldsymbol{x}^t_{S^\star} + \boldsymbol{v}^{t+1}_{S^\star}, t \geq T_{\text{lock-in}}. \tag{18}$$

Then, using the fact that $\boldsymbol{P}_{S^\star}^\perp \boldsymbol{A}_{S^\star} = \boldsymbol{0}_{S^\star}$, for $t \geq T_{\text{lock-in}} + 1$,

$$\begin{aligned}
&\boldsymbol{x}^{t+1}_{S^\star} = \boldsymbol{x}^t_{S^\star} + \boldsymbol{v}^{t+1}_{S^\star} \\
&= \boldsymbol{x}^t_{S^\star} + \widehat{\beta}\boldsymbol{v}^t_{S^\star} + \widehat{\mu}\boldsymbol{A}_{S^\star}^\top(\boldsymbol{y} - \boldsymbol{A}_{S^\star}\boldsymbol{x}^t_{S^\star}) \\
&\implies \boldsymbol{x}^{t+1}_{S^\star} - \widehat{\boldsymbol{x}}_{S^\star} = (\boldsymbol{I}_K - \widehat{\mu}\boldsymbol{A}_{S^\star}^\top\boldsymbol{A}_{S^\star})(\boldsymbol{x}^t_{S^\star} - \widehat{\boldsymbol{x}}_{S^\star}) \\
&+ \widehat{\beta}(\boldsymbol{x}^t_{S^\star} - \boldsymbol{x}^{t-1}_{S^\star}) \\
&\implies \boldsymbol{h}^{t+1} = (\boldsymbol{I}_K - \widehat{\mu}\boldsymbol{\Phi})\boldsymbol{h}^t + \widehat{\beta}(\boldsymbol{h}^t - \boldsymbol{h}^{t-1}),
\end{aligned} \tag{19}$$

where $\boldsymbol{h}^t = \boldsymbol{x}^t_{S^\star} - \widehat{\boldsymbol{x}}_{S^\star}$ and $\boldsymbol{\Phi} = \boldsymbol{A}_{S^\star}^\top\boldsymbol{A}_{S^\star}$. This is the update equation of the classical Polyak's heavy-ball method. Using standard analysis [16] one can obtain the following (see the technical report [15] for details):

$$\left\|\boldsymbol{h}^t\right\|_2 = \mathcal{O}\left(\sqrt{K}\widetilde{\beta}^{\frac{t-T_{\text{lock-in}}-1}{2}}\right), \tag{20}$$

where $g(x) = x + \sqrt{x^2 - 1}$ and

$$\sqrt{\widetilde{\beta}} = \sqrt{\widehat{\beta}} g\left(\max\left\{\max_{1\le i\le K}\left|\frac{1+\widehat{\beta}-\widehat{\mu}\lambda_i^\star}{2\sqrt{\widehat{\beta}}}\right|, 1\right\}\right), \quad (21)$$

where $\lambda_1^\star \ge \cdots \ge \lambda_K^\star$ are the eigenvalues of $\boldsymbol{A}_{S^\star}^\top \boldsymbol{A}_{S^\star}$. Once lock-in is achieved, we have (line 8 of Algorithm 2)

$$\widehat{\beta} = \left(\frac{\sqrt{L}-\sqrt{l}}{\sqrt{L}+\sqrt{l}}\right)^2, \widehat{\mu} = \frac{4}{(\sqrt{L}+\sqrt{l})^2}, \quad (22)$$

where

$$L = \|\boldsymbol{A}_{S^\star}\boldsymbol{b}\|_2^2, l = \|\boldsymbol{A}_{S^\star}\boldsymbol{c}\|_2^2, \quad (23)$$

$$\boldsymbol{b} = \text{POWER}(\boldsymbol{A}_{S^\star}^\top \boldsymbol{A}_{S^\star}, \boldsymbol{b}_0, r),$$
$$\boldsymbol{c} = \text{POWER}(\|\boldsymbol{A}_{S^\star}\|_F^2 \boldsymbol{I}_K - \boldsymbol{A}_{S^\star}^\top \boldsymbol{A}_{S^\star}, \boldsymbol{b}_0, r). \quad (24)$$

Let $\boldsymbol{B} = \boldsymbol{A}_{S^\star}^\top \boldsymbol{A}_{S^\star}$, and denote the eigenvalues of $\boldsymbol{B}$ by $\lambda_1^\star \ge \cdots \ge \lambda_K^\star$, which satisfy $1-\delta_K \le \lambda_K^\star \le \lambda_1^\star \le 1+\delta_K$ by the RIP of $\boldsymbol{A}$. When $L = \lambda_1$ and $l = \lambda_K$, the parameters $\widehat{\beta}$ and $\widehat{\mu}$ reduce to the optimal heavy-ball parameters for $\boldsymbol{B}$. This guarantee is achievable in the limit of indefinitely many power iterations; however, since only finitely many iterations are performed in practice, there is an approximation error between L and λ_1, and between l and λ_K. This error renders $\widehat{\beta}$ and $\widehat{\mu}$ suboptimal and consequently degrades the convergence rate $\sqrt{\widetilde{\beta}}$. We now show that, with an appropriate initialization of the power method, a certain number of iterations suffices to control this approximation error within a prescribed tolerance, thereby bringing the resulting convergence rate within a desired accuracy of the optimal rate $\frac{\sqrt{\lambda_1^\star}-\sqrt{\lambda_K^\star}}{\sqrt{\lambda_1^\star}+\sqrt{\lambda_K^\star}}$.

Theorem 2. *Let there be a number $c > 1$ such that the matrix $\boldsymbol{A}_{S^\star}^\top \boldsymbol{A}_{S^\star}$ satisfies $\max\{\mu_2, \eta_{K-1}\} \le \frac{1}{c}$, where $\mu_2 = \frac{\lambda_2^\star}{\lambda_1^\star}$ and $\eta_{K-1} = \frac{\sum_{i\ne K-1}\lambda_i^\star}{\sum_{i\ne K}\lambda_i^\star}$. Then, with the initialization of Algorithm* `POWER` *used in* `CosMIHT`*, and with $K \ge 3$, $\epsilon > 0$, the following is ensured with probability greater than $1 - 1/K^p$, for any $p \ge 1$,*

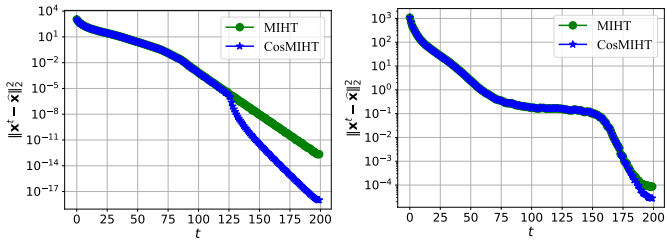
$$r \ge \left\lceil \frac{\ln\left(\frac{4K^{p+2}\lambda_1^\star}{\pi\epsilon}\right)}{2\ln c}\right\rceil \implies \begin{matrix}\lambda_1^\star - \epsilon \le L \le \lambda_1^\star, \\ \lambda_K^\star \le l \le \lambda_K^\star + \epsilon.\end{matrix} \quad (25)$$

Furthermore, for $t \ge T_{\text{lock-in}} + 1$,

$$\left\|\boldsymbol{x}^t - \widehat{\boldsymbol{x}}\right\|_2 = \mathcal{O}\left(\sqrt{K}\left(\frac{\delta_{3K}\left(1+\frac{2\epsilon}{\lambda_1^\star-\lambda_K^\star}\right)}{1+\sqrt{1-\delta_{3K}^2}}\right)^t\right). \quad (26)$$

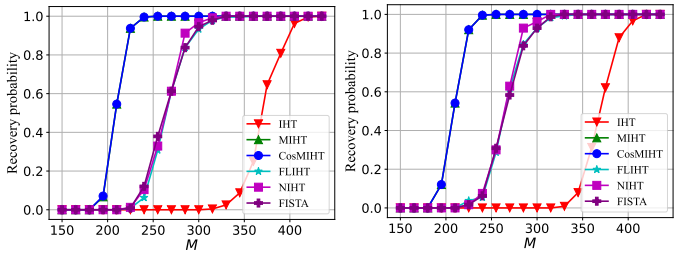
Remark 2. Since $\max\{\mu_2, \eta_{K-1}\} < 1$ and the logarithmic dependence on K and ϵ^{-1} is mild, the number of power iterations required is typically $r = \mathcal{O}(\log K)$ for moderate accuracy, making the eigenvalue estimation lightweight.

Proof. The proof can be found in the technical report [15]. □



(a) $\sigma = 0$. (b) $\sigma = 10^{-3}$

Fig. 1: MSE vs iterations for noiseless and noisy measurements.



(a) $\sigma = 0$. (b) $\sigma = 10^{-3}$

Fig. 2: Recovery Probability vs M for noiseless and noisy measurements.

B. Computational Complexity Analysis

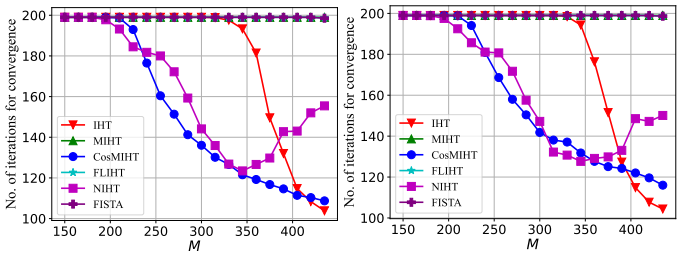
We carry out a computational complexity analysis of the proposed `CosMIHT` in terms of the number of floating point operations (flops) required to execute their steps. The description of `CosMIHT` (Algorithm 2) implies that any step t involves two types of computations: the compulsory computations (lines 11–13 of Algorithm 2) and the conditional computations (lines 5–8 of Algorithm 2). The conditional computations consume the following costs:

1) Power iterations for L_t computation (line 5): $\mathcal{O}(rMK)$,
2) Power iterations for l_t computation (line 6): $\mathcal{O}(rMK) + \mathcal{O}(rK^2)$,
3) L_t, l_t computations (line 7): $\mathcal{O}(KM) + \mathcal{O}(K)$.

The compulsory computations consume

1) Gradient computation (line 11): $\mathcal{O}(MN)$,
2) Momentum update (line 12): $\mathcal{O}(N)$,
3) Estimate update (line 13): $\mathcal{O}(N) + \mathcal{O}(K\ln N)$.

The extra computational burden of `CosMIHT` over MIHT stems solely from the conditional computations, which are triggered at most $T_{\text{lock-in}}/2 + 1$ times in total — at most $T_{\text{lock-in}}/2$ times during the wandering phase and once upon entering the lock-in phase. This amounts to $\mathcal{O}(rMKT_{\text{lock-in}}) = \mathcal{O}(rKM\ln K)$ additional flops (by Remark 1), which is substantially smaller than the $\mathcal{O}(TMN)$ compulsory cost. Hence `CosMIHT` is essentially as lightweight as MIHT, while enjoying the improved convergence established above.



(a) $\sigma = 0$. (b) $\sigma = 10^{-3}$

Fig. 3: No. of iterations for convergence vs M for noiseless and noisy measurements.

V. Numerical Experiments

We compare `CosMIHT` against MIHT [13], IHT [7], NIHT [17], FLIHT [11] and FISTA [18] via Monte Carlo simulations over 200 independent instances. We generate $\boldsymbol{A} \in \mathbb{R}^{M\times N}$ with i.i.d. $\mathcal{N}(0,1)$ entries and ℓ_2-normalized columns, a K-sparse $\boldsymbol{x}^\star$ with i.i.d. $\mathcal{N}(0,1)$ nonzero entries, and noise $\boldsymbol{e} \sim \mathcal{N}(\boldsymbol{0}, \sigma \boldsymbol{I}_M)$. We fix $N = 1000$, $K = 80$, and vary $M \in \{150, 165, \ldots, 450\}$. Parameters are set to $\mu = 0.6$, $\beta = 0.8$ for MIHT [13], $\mu = 0.6$ for IHT, $c = 0.1,\ \kappa = 2$ for NIHT [10], and $\mu = \frac{5}{(1+\sqrt{6K/M}+\sqrt{2})^2}$ for FLIHT [11].

Figure 1 plots the MSE $\|\boldsymbol{x}^t - \widehat{\boldsymbol{x}}\|_2^2$ vs. iterations. The two-phase convergence of `CosMIHT` is clearly visible in the noiseless case: the MSE curve exhibits a distinct acceleration fork at around iteration 125, marking the transition from the wandering phase to the lock-in phase. In the noisy case the MSE floor is dominated by measurement noise and the convergence performance of MIHT and `CosMIHT` is similar.

Figure 2 plots the empirical recovery probability (success criterion: $\|\boldsymbol{x}^T - \widehat{\boldsymbol{x}}\|_2 < 10^{-3}\|\widehat{\boldsymbol{x}}\|_2 + 5\,\|\boldsymbol{e}\|_2$) vs. M. In both the noiseless and noisy cases, `CosMIHT` matches MIHT and outperforms IHT, NIHT, FLIHT and FISTA, with `CosMIHT` having the smallest critical M (the minimum M beyond which recovery occurs with high probability) among all.

Figure 3 plots the mean iterations to convergence ($\|\boldsymbol{x}^{\tau+1} - \boldsymbol{x}^{\tau}\|_2 < 10^{-10}$): `CosMIHT` requires the fewest iterations across all comparisons, with NIHT and IHT being the only near-competitors but at a significant cost in recovery probability (see Figs. 2(a)–(b)).

References

[1] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[2] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[3] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society: Series B*, vol. 58, no. 1, pp. 267–288, 1996.

[4] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[5] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[6] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.

[7] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, 2009.

[8] B. T. Polyak, "Some methods of speeding up the convergence of iteration methods," *USSR Computational Mathematics and Mathematical Physics*, vol. 4, no. 5, pp. 1–17, 1964.

[9] Y. Nesterov, "A method of solving a convex programming problem with convergence rate $O(1/k^2)$," *Soviet Mathematics Doklady*, vol. 27, no. 2, pp. 372–376, 1983.

[10] T. Blumensath and M. E. Davies, "Normalized Iterative Hard Thresholding: Guaranteed Stability and Performance," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 298–309, 2010.

[11] V. Cevher, "An ALPS view of sparse recovery," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 5808–5811.

[12] Z. Sun, J. Zhou, Y.-B. Zhao, and N. Meng, "Heavy-ball-based hard thresholding algorithms for sparse signal recovery," *Journal of Computational and Applied Mathematics*, vol. 430, p. 115264, 2023.

[13] W. Jin and L.-J. Xie, "Momentum-based iterative hard thresholding algorithm for sparse signal recovery," *IEEE Signal Processing Letters*, vol. 32, pp. 1346–1350, 2025.

[14] J. Shen and P. Li, "A tight bound of hard thresholding," *Journal of Machine Learning Research*, vol. 18, no. 208, pp. 1–42, 2018.

[15] S. Mukhopadhyay and D. Mukherjee, "Consecutive Support Matching Induced Parameter Tuning Accelerates Momentum Iterative Hard Thresholding," 2026. [Online]. Available: https://arxiv.org/abs/2606.09382

[16] S. J. Wright and B. Recht, *Optimization for data analysis*. Cambridge University Press, 2022.

[17] T. Blumensath and M. E. Davies, "Normalized iterative hard thresholding: guaranteed stability and performance," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 298–309, 2010.

[18] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

Appendix A

We need to find a t such that $S_{t+1} = S^\star$ for all $\tau \geq t+1$. For this to happen, we must ensure that, at time t, for any $i \in S^\star$ and $j \notin S^\star$, the condition $\left|(\boldsymbol{x}^t + \boldsymbol{v}^{t+1})_i\right| \geq \left|(\boldsymbol{x}^t + \boldsymbol{v}^{t+1})_j\right|$ holds. Expanding $\boldsymbol{v}^{t+1}$ and using triangle inequality, the left hand side (LHS) and the right hand side (RHS) of the preceding inequality can be lower and upper bounded, respectively, and following some straightforward algebra (see the technical report for details [15]) it can be shown that the following condition is sufficient for the above condition to hold:

$$\sqrt{2}\left[\left|((\boldsymbol{I} - \mu_t \boldsymbol{A}^\top \boldsymbol{A})(\boldsymbol{x}^t - \widehat{\boldsymbol{x}}))_{i,j}\right| + \beta_t \left|(\boldsymbol{v}^t)_{i,j}\right|\right] < \left|\widehat{x}_i\right|. \tag{27}$$

Recalling the condition (5), as well as using the bounds of ρ_t and β_t (see the technical report [15]), a sufficient condition ensuring the above is $\sqrt{2}(\rho E_t + \alpha V_t) \leq x^\star_{\min}$, where $x^\star_{\min} = \min_{i\in S^\star} |x_i|$. By the evolution inequalities (12) and (13), and recalling that eigenvalue of $\boldsymbol{Q}$ satisfy $|\lambda_2| < \lambda_1$ from (14), it can be shown that the above is satisfied if the condition (16) holds. The details can be found in the technical report [15].