Iteratively Reweighted Algorithms for Robust Sparse Recovery with ℓ_p-Norm Minimization

Shweta Bharti[1], Samrat Mukhopadhyay[1], Govind Murmu[1]
[1]Department of Electronics Engineering, IIT (ISM) Dhanbad
Email: 21dr0178@ece.iitism.ac.in, samrat@iitism.ac.in, govind@iitism.ac.in

***Abstract*—Recovering sparse signals from limited and noisy measurements is a fundamental problem in compressed sensing, particularly when the measurements are corrupted by impulsive or heavy–tailed noise. Conventional recovery techniques based on quadratic loss functions often lack of robustness in such scenarios. In this work, sparse signal reconstruction is addressed using an ℓ_p loss formulation with $0 < p < 2$, which improves resilience to non-Gaussian noise while encouraging sparsity. We propose an iteratively reweighted framework that integrates hard thresholding with proximal gradient updates to solve the sparsity-constrained optimization problem. The resulting algorithms, termed Iteratively Reweighted Hard Thresholding (IRHT) and Iteratively Reweighted Hard Thresholding Pursuit (IRHTP), transform the nonconvex ℓ_p objective into a sequence of weighted least-squares subproblems that can be solved efficiently. This approach maintains computational simplicity comparable to greedy sparse recovery methods while improving robustness to outliers. Theoretical analysis establishes convergence of the proposed framework to a local minimizer of a smoothed ℓ_p-based objective. The proposed methods demonstrate effective sparse recovery performance in environments characterized by impulsive noise and heavy-tailed measurement distributions.**

***Index Terms*—Robust recovery, ℓ_p-norm, majorization-minimization, iterative hard thresholding**

I. Introduction

Compressed sensing has emerged as a powerful tool to recover a long unknown vector, from a few linear measurements [1]. We consider the problem of sparse signal recovery from linear measurements: recovering a vector $\boldsymbol{x} \in \mathbb{R}^n$, given $\boldsymbol{y} \in \mathbb{R}^m$ obtained from $\boldsymbol{x}$ (unknown) via linear measurements $\boldsymbol{y} = \boldsymbol{\Phi x} + \boldsymbol{e}$, where $\boldsymbol{\Phi} \in \mathbb{R}^{m\times n}$ is the measurement matrix and $\boldsymbol{e} \in \mathbb{R}^m$ is the measurement error vectors. Mathematically, the solution to the following problem is desired: $\min_{\boldsymbol{z}\in\mathbb{R}^n:\|\boldsymbol{z}\|_0\leq K} J(\boldsymbol{y}-\boldsymbol{\Phi z})$, where, $J(\cdot)$ is some scalar loss function,and $\|\boldsymbol{z}\|_0$ counts the number of nonzero coordinates of the vector $\boldsymbol{z}$.

In this paper we propose greedy algorithms to solve the following problem:

$$\min_{\boldsymbol{z}\in\mathbb{R}^n} \|\boldsymbol{y}-\boldsymbol{\Phi z}\|_p^p, \qquad \text{s.t. } \|\boldsymbol{z}\|_0 \leq K,$$

when $p \in (0,2)$. Here the function $\|\cdot\|_p : \mathbb{R}^n \to \mathbb{R}^+$ is defined as $\|\boldsymbol{z}\|_p^p = \sum_{i=1}^n |z_i|^p$, is the l_p norm of vector $\boldsymbol{z} \in \mathbb{R}^n$.

For $p \in (0,2)$, the l_p norm based estimators are found to be more suitable against impulsive noise, as their stability requires only finite $p^{\text{th}}(p<2)$ order moment of the measurement noise. A wide range of impulsive noise distributions, like the generalized Gaussian distribution (GGD) and the symmetric-α-stable ($S\alpha S$) distributions satisfy this condition and measurement noise with this kind of distribution is frequently encountered in applications like channel estimation in OTFS [2], time delay estimation [3] etc. Several robust variants of iterative hard thresholding have been proposed to address such non-Gaussian noise scenarios. In particular, Lorentzian Iterative Hard Thresholding (LIHT) [4] replaces the quadratic loss with a Lorentzian function to improve robustness against impulsive noise. Similarly, Huber-based Iterative Hard Thresholding (HIHT) [5] employs the Huber loss to balance sensitivity to small residuals and robustness to outliers. However, these approaches do not directly optimize a general l_p loss estimator which arise naturally as the maximum likelihood estimator (MLE) for generalized Gaussian distributed (GGD) noise . Moreover, their theoretical convergence properties are often limited or difficult to establish under general noise conditions.Motivated by these limitations, we propose a novel method for sparse recovery based on l_p loss minimization $(0 < p < 2)$, which integrates iteratively reweighted with hard thresholding and hard thresholding pursuit.

Our main contributions in this regard are the following:

1) Following Daubechies' IRLS [6] approach, we have proposed an iteratively reweighted algorithm, called the iteratively reweighted hard thresholding(IRHT) and iteratively reweighted hard thresholding pursuit (IRHTP), to solve the $l_p - l_0$ problem. In each iteration of the algorithm we use a proximal gradient step to solve a weighted l_2 function which resolves the difficulty associated with the non-convexity or non-smoothness of the original l_p loss function. Our formulation of the minimization procedure is similar to the one considered by Sigl *et al* [7]. However, the main novelty lies in modifying their procedure by incorporating the sparsity constraint of the problem, which is in the same spirit of the formulation of IHT algorithm using the majorization-minimization (MM) framework as in [8]. Moreover, we generalize some of the analysis of Sigl [7] for the case $p \in [0,1)$. As such, our formulation demonstrates a systematic way of formulating the LIHT algorithm [4].
2) We analytically prove that the algorithm converges to one of the local minimizers of the following related problem:

$$\min_{\boldsymbol{z}\in\mathbb{R}^n} \sum_{j=1}^{m} \left[(y_j - \boldsymbol{\Phi z})_j)^2 + \epsilon^2\right]^{p/2}, \text{s.t. } \|\boldsymbol{z}\|_0 \leq K,$$

Algorithm 1 IRHT and IRHTP Algorithms

Input: Order $p \in [0,2)$, $\mu > 0$, $\epsilon > 0$, max. iteration k_{it}.
Initialize: $k = 0$, $\boldsymbol{x}^0 = \mathbf{0}$, $\boldsymbol{w}^0 = [1\ 1\ \cdots\ 1]^T \in \mathbb{R}^m$
for $k = 0, \ldots, k_{\text{it}}$ **do**

1) $\boldsymbol{x}^{k+1}$ **update:** using (3) for IRHT; using (4) and (5) for IRHTP
2) $\boldsymbol{w}^{k+1}$ **update:** using (7)

end for

where, ϵ is a parameter of the algorithm.

3) We analyze the convergence rate and derive provable performance guarantees by imposing suitable conditions on the algorithm parameters to ensure stability.
4) Finally, we provide corroborate our theoretical claims with extensive numerical simulations.

Notations: We use $H_K(\cdot)$ to denote the hard thresholding operator, which sets all elements to zero except for the K largest entries by magnitude. The operator $L_K(\cdot)$ returns the support after applying $H_K(\cdot)$. The $\|\cdot\|_{2\to 2}$ denotes the l_2- operator norm of a matrix. For any positive integer K, a matrix $\boldsymbol{\Phi} \in \mathbb{R}^{m\times n}$ is said to satisfy the RIP of order K if it satisfies $\delta_K \in (0,1)$, where, $\delta_K := \max_{S\subset[N]:|S|\le K} \left\|\boldsymbol{\Phi}_S^t\boldsymbol{\Phi}_S - \boldsymbol{I}_K\right\|_{2\to 2}$ where,$\boldsymbol{\Phi}_S$ denotes the submatrix of $\boldsymbol{\Phi}$ formed with the columns of $\boldsymbol{\Phi}$ restricted to the index set S, and $\boldsymbol{I}_K$, denote the identity matrix of size $K \times K$.

II. Algorithm for the K sparse problem

We intend to solve the following problem :

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \sum_{j=1}^{m} \left[(y_j - (\boldsymbol{\Phi}\boldsymbol{x})_j)^2 + \epsilon^2\right]^{p/2} \quad \text{s.t. } \|\boldsymbol{x}\|_0 \le K, \quad \text{(P1)}$$

The **Algorithm** is proposed to solve the problem (P1). To formulate the algorithm, we first define the functional in Eq (1), and then use a series of minimizations on the functional to obtain the updates of $\boldsymbol{x}$ and $\boldsymbol{w}$.

Lemma 2.1. *Define the following surrogate functional:*

$$\begin{aligned}
&\mathcal{J}(\boldsymbol{x},\boldsymbol{z},\boldsymbol{w},\epsilon)\\
&= \frac{p}{2}\sum_{j=1}^{m} w_j \left[\left(y_j - (\boldsymbol{\Phi}\boldsymbol{x})_j\right)^2 - \left((\boldsymbol{\Phi}\boldsymbol{x})_j - (\boldsymbol{\Phi}\boldsymbol{z})_j\right)^2\right]\\
&+ \frac{p\,\|\boldsymbol{w}\|_\infty}{2\mu}\|\boldsymbol{x}-\boldsymbol{z}\|_2^2 + \frac{p}{2}\sum_{j=1}^{m}\left[\epsilon^2 w_j + \frac{2-p}{p} w_j^{\frac{p}{p-2}}\right], \quad (1)
\end{aligned}$$

where $\boldsymbol{x} \in \mathbb{R}^n$, $\epsilon > 0$, $\boldsymbol{w} \in \mathbb{R}^m$, with $w_j > 0$, $\forall j = 1, 2, \cdots, m$, and $\boldsymbol{y} \in \mathbb{R}^m$.

Then, for a fixed μ, the minimization procedure

$$\boldsymbol{x}^{k+1} = \underset{\boldsymbol{x}\in\mathbb{R}^n:\|x\|_0\le K}{\arg\min}\ \mathcal{J}(\boldsymbol{x},\boldsymbol{x}^k,\boldsymbol{w}^k,\epsilon). \quad (2)$$

defines the iterative procedure for IRHT algorithm

$$\boldsymbol{x}^{k+1} = H_K\left(\boldsymbol{x}^k + \mu\boldsymbol{\Phi}^t\boldsymbol{D}_k\left(\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}^k\right)\right). \quad (3)$$

where, $\boldsymbol{D}_k = \mathtt{diag}(\boldsymbol{w}^k)/\left\|\boldsymbol{w}^k\right\|_\infty$,
and define iterative procedure for IRHTP algorithm.

$$\begin{aligned}
\boldsymbol{u}^{k+1} &= \boldsymbol{x}^k + \mu\boldsymbol{\Phi}^t\boldsymbol{D}_k\left(\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}^k\right),\\
\Lambda^{k+1} &= L_K(\boldsymbol{u}^{k+1}), \quad (4)\\
\boldsymbol{x}^{k+1} &= \underset{supp(\boldsymbol{x})\subseteq\Lambda^{k+1}}{\arg\min}\left\|\tilde{\boldsymbol{y}}_k - \tilde{\boldsymbol{\Phi}}_k\boldsymbol{x}\right\|_2^2. \quad (5)
\end{aligned}$$

Moreover, the minimization procedure

$$\boldsymbol{w}^{k+1} = \underset{\boldsymbol{w}\in\mathbb{R}^n_+}{\arg\min}\ \mathcal{J}(\boldsymbol{x}^{k+1},\boldsymbol{x}^{k+1},\boldsymbol{w},\epsilon) \quad (6)$$

defines the sequence of weight updates

$$w_j^{k+1} = \left[\left(y_j - \left(\boldsymbol{\Phi}\boldsymbol{x}^{k+1}\right)_j\right)^2 + \epsilon^2\right]^{\frac{p-2}{2}},\ 1 \le j \le m. \quad (7)$$

Proof. For detailed proof, please refer to Appendix (A) ∎

III. Analysis of the IRHT and IRHTP algorithm

Lemma 3.1. *At any step $k \ge 0$, under the assumption $\mu \in \left(0, \frac{1}{1+\delta_{2K}}\right]$, for any K-sparse vector $\boldsymbol{z} \in \mathbb{R}^n$, the functional $\mathcal{J}$, defined in Eq* (1) *satisfies*

$$\left\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}^k\right\|_p^p \le \mathcal{J}(\boldsymbol{x}^k,\boldsymbol{x}^k,\boldsymbol{w}^k,\epsilon) \le \mathcal{J}(\boldsymbol{x}^k,\boldsymbol{z},\boldsymbol{w}^k,\epsilon). \quad (8)$$

Proof. It is straightforward to obtain from the definition in Eq (1), using the update equation for $\boldsymbol{w}^k$ from Eq (7), that

$$\mathcal{J}(\boldsymbol{x}^k,\boldsymbol{x}^k,\boldsymbol{w}^k,\epsilon) = \sum_{j=1}^{m}\left[\left(y_j - (\boldsymbol{\Phi}\boldsymbol{x}^k)_j\right)^2 + \epsilon^2\right]^{p/2}. \quad (9)$$

This verifies the first inequality.

For the second inequality, observe that, under the assumptions on μ, for any two K-sparse vectors $\boldsymbol{x},\boldsymbol{z} \in \mathbb{R}^n$, and $\boldsymbol{w} \in \mathbb{R}^m$, one gets

$$\begin{aligned}
&\Rightarrow \sum_{j=1}^{m} w_j\left((\boldsymbol{\Phi}\boldsymbol{x})_j - (\boldsymbol{\Phi}\boldsymbol{z})_j\right)^2 = \left\|\boldsymbol{D}^{1/2}\boldsymbol{\Phi}\left(\boldsymbol{x}-\boldsymbol{z}\right)\right\|_2^2\\
&\le \left\|\boldsymbol{D}^{1/2}\right\|_{2\to 2}^2 \|\boldsymbol{\Phi}\left(\boldsymbol{x}-\boldsymbol{z}\right)\|_2^2 \quad (\text{where } \boldsymbol{D} = \mathtt{diag}(\boldsymbol{w}))\\
&\le (\max_j w_j)(1+\delta_{2K})\|\boldsymbol{x}-\boldsymbol{z}\|_2^2\\
&\le \frac{\|\boldsymbol{w}\|_\infty}{\mu}\|\boldsymbol{x}-\boldsymbol{z}\|_2^2. \quad (10)
\end{aligned}$$

Thereafter the inequality follows by simply using the above inequality in the expression for $\mathcal{J}(\boldsymbol{x}^k,\boldsymbol{z},\boldsymbol{w}^k,\epsilon)$. ∎

Lemma 3.2. *If $\{\boldsymbol{x}^k\}_{k\ge 0}$ is the sequence generated by IRHT (resp. IRHTP), the sequence $\{\mathcal{J}(\boldsymbol{x}^k,\boldsymbol{x}^k,\boldsymbol{w}^k,\epsilon)\}_{k\ge 0}$ is a positive valued monotone non-increasing sequence.*

Proof. That $\mathcal{J}(\boldsymbol{x}^k,\boldsymbol{x}^k,\boldsymbol{w}^k,\epsilon)$ is positive for all $k \ge 0$, follows from the definition of $\mathcal{J}$ in Eq (1), noting that $\boldsymbol{w}^k > 0, \epsilon > 0$, $\forall k \ge 0$.

To prove the non-increasing property, we obtain

$$\mathcal{J}(\boldsymbol{x}^{k+1},\boldsymbol{x}^{k+1},\boldsymbol{w}^{k+1},\epsilon) \overset{\xi_1}{\le} \mathcal{J}(\boldsymbol{x}^{k+1},\boldsymbol{x}^{k+1},\boldsymbol{w}^k,\epsilon),$$

which follows from the update procedure of $\boldsymbol{w}^k$, as established in Eq 6 in Lemma 2.1. To further proceed we need to use the update mechanism $\boldsymbol{x}^k$ to $\boldsymbol{x}^{k+1}$, for which we need separate analysis for IRHT and IRHTP.

IRHT analysis: For IRHT, using the majorization property of the functional $\mathcal{J}$ as established in Lemma 3.1 and the update equaton for IRHT in (2), we obtain that

$$\begin{aligned}\mathcal{J}(\boldsymbol{x}^{k+1},\boldsymbol{x}^{k+1},\boldsymbol{w}^k,\epsilon) &\overset{\xi_2}{\leq} \mathcal{J}(\boldsymbol{x}^{k+1},\boldsymbol{x}^k,\boldsymbol{w}^k,\epsilon)\\ &\leq \mathcal{J}(\boldsymbol{x}^k,\boldsymbol{x}^k,\boldsymbol{w}^k,\epsilon),\end{aligned} \tag{11}$$

which is the desired result IRHT.

IRHTP analysis: For IRHTP, first denote $\boldsymbol{v}^{k+1} = H_K(\boldsymbol{u}^{k+1})$ and then note that

$$\boldsymbol{v}^{k+1} = \arg\min_{\boldsymbol{x}\in\mathbb{R}^n:\|x\|_0\leq K} \mathcal{J}(\boldsymbol{x},\boldsymbol{x}^k,\boldsymbol{w}^k,\epsilon) \tag{12}$$

Note that, with $\boldsymbol{W}^k = \texttt{diag}\left(\boldsymbol{w}^k\right)$, we get,

$$\begin{aligned}&\mathcal{J}(\boldsymbol{x}^{k+1},\boldsymbol{x}^{k+1},\boldsymbol{w}^k,\epsilon) - \mathcal{J}(\boldsymbol{x}^k,\boldsymbol{x}^k,\boldsymbol{w}^k,\epsilon) =\\ &\frac{p}{2}\left[\left\|\boldsymbol{W}_k^{1/2}(\boldsymbol{y}_k-\boldsymbol{\Phi}\boldsymbol{x}^{k+1})\right\|^2 - \left\|\boldsymbol{W}_k^{1/2}(\boldsymbol{y}_k-\boldsymbol{\Phi}\boldsymbol{x}^{k})\right\|^2\right]\\ &\overset{(5)}{\leq} \frac{p}{2}\left[\left\|\boldsymbol{W}_k^{1/2}(\boldsymbol{y}_k-\boldsymbol{\Phi}\boldsymbol{v}^{k+1})\right\|^2 - \left\|\boldsymbol{W}_k^{1/2}(\boldsymbol{y}_k-\boldsymbol{\Phi}\boldsymbol{x}^{k})\right\|^2\right]\\ &= \mathcal{J}(\boldsymbol{v}^{k+1},\boldsymbol{v}^{k+1},\boldsymbol{w}^k,\epsilon) - \mathcal{J}(\boldsymbol{x}^k,\boldsymbol{x}^k,\boldsymbol{w}^k,\epsilon)\\ &\overset{(8)}{\leq} \mathcal{J}(\boldsymbol{v}^{k+1},\boldsymbol{x}^k,\boldsymbol{w}^k,\epsilon) - \mathcal{J}(\boldsymbol{x}^k,\boldsymbol{x}^k,\boldsymbol{w}^k,\epsilon)\\ &\overset{(4)}{\leq} \mathcal{J}(\boldsymbol{x}^k,\boldsymbol{x}^k,\boldsymbol{w}^k,\epsilon) - \mathcal{J}(\boldsymbol{x}^k,\boldsymbol{x}^k,\boldsymbol{w}^k,\epsilon) = 0.\end{aligned} \tag{13}$$

■

Theorem 3.1. *Let the true measuremernt model be given by $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{e}$, where $\boldsymbol{x}$ is a K-sparse vector, and $\boldsymbol{e}$ is the measurement noise vector. Then, under the condition $\mu < 1/(1+\delta_{2K})$,and $\gamma \geq \min_{k\geq 0}\gamma_k, \forall k \geq 0$, the iterates $\boldsymbol{x}^k$ generated by IRHT (resp.IRHTP) algorithm satisfy,*

$$\left\|\boldsymbol{x}^{k+1}-\boldsymbol{x}\right\|_2 \leq \rho_{3K}\left\|\boldsymbol{x}^k-\boldsymbol{x}\right\|_2 + \tau_{2K}\left\|\boldsymbol{e}\right\|_2, \tag{14}$$

where,

$$\rho_{3K} = \begin{cases} 2(1-\mu(1-\delta_{3K})\gamma_k), & \textit{IRHT},\\ \sqrt{\frac{\omega(1-\gamma_k(1-\delta_{3K}))^2}{1-(1-\gamma_k(1-\delta_{2K}))^2}}, & \textit{IRHTP}\end{cases}. \tag{15}$$

and

$$\tau_{2K} = \begin{cases} \omega\mu\sqrt{1+\delta_{2K}}, & \textit{IRHT},\\ \frac{\sqrt{2\gamma_k(1-\delta_{2K})+\sqrt{1+\delta_K}}}{\gamma_k(1-\delta_{2K})}, & \textit{IRHTP}\end{cases}. \tag{16}$$

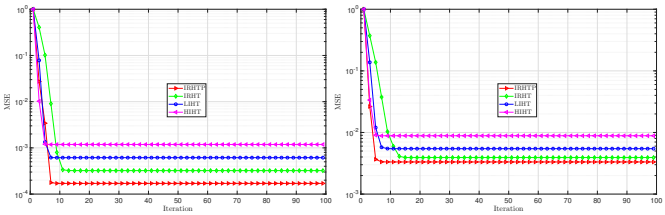
Moreover, if $\mu > 1/(2(1-\delta_{2K}))$, and ϵ are chosen such that

$$\epsilon > \left(\frac{\left\|\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x}^0\right\|_p}{\left[\left((2\mu(1-\delta_{3K}))^{\frac{p}{2-p}}\right)-1\right]^{1/p}}\right). \tag{17}$$

then $\rho_{3K} < 1$,and consequently,

$$\left\|\boldsymbol{x}^\star-\boldsymbol{x}\right\|_2 < \frac{\tau_{2K}}{1-\rho_{3K}}\left\|\boldsymbol{e}\right\|_2, \tag{18}$$

where $\boldsymbol{x}^\star$ is the limit of the sequence $\{\boldsymbol{x}^k\}_{k\geq 0}$.



(a) $\gamma = 0.8$ (b) $\gamma = 5$

Fig. 1: MSE vs. Iteration

Remark III.1. If we consider $\mu < 1/(1+\delta_{2K})$, from Lemma 3.1 and require $\mu > 1/(\omega(1-\delta_{2K}))$, for $\rho_{3K} < 1$, such that,

$$\begin{aligned}\omega(1-\delta_{2K}) &> (1+\delta_{2K})\\ \Rightarrow \delta_{2K} &< 1/(\omega+1) \approx 0.382.\end{aligned} \tag{19}$$

Hence, under the RIP with $\delta_{2K} < 1/3$, there exists a feasible choice of μ satisfying the above conditions.

Proof. Please refer to Appendix (B) ■

IV. Numerical Experiments

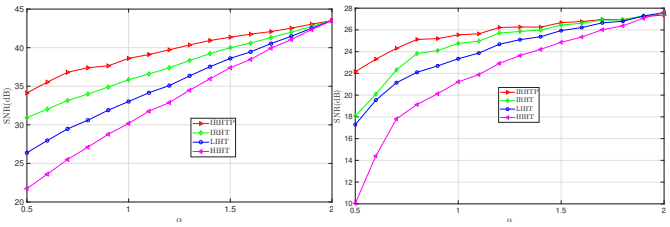
We evaluate the proposed IRHT and IRHTP algorithms under impulsive noise. The sparsity level is fixed at $K = 8$, with sensing matrix dimensions $m = 256$ and $n = 512$. The nonzero entries of the signal have amplitude 10 with equiprobable signs, and the support is selected uniformly at random from $\{1,\ldots,n\}$. The normalized sensing matrix $\boldsymbol{\Phi}$ has entries drawn i.i.d. from $\mathcal{N}(0,1/m)$. Each experiment is repeated 100 times and the results are averaged. The measurements are corrupted by SαS noise with characteristic exponent α and scale parameter γ. Both algorithms are run for 100 iterations. The threshold parameter is set to $\epsilon \in \{11,15\}$, with step sizes $\mu = 1/m$ for IRHT and $\mu = 1.5/m$ for IRHTP. Reconstruction quality is measured using

$$\text{SNR} = 20\log_{10}\left(\frac{\|x[i]\|}{\|\hat{x}[i]-x[i]\|}\right), \tag{20}$$

where $x[i]$ and $\hat{x}[i]$ denote the true and estimated signals in the i^{th} trial.

Fig. 1 shows convergence behavior under $\gamma = 0.8$ and $\gamma = 5$. All algorithms exhibit rapid initial MSE reduction; however, IRHT and IRHTP converge to significantly lower steady-state errors compared to LIHT [4] and HIHT [5]. IRHTP consistently achieves the lowest MSE, followed by IRHT.

Fig. 2 shows reconstruction SNR as α varies from 0.5 to 2. The SNR increases monotonically with α, reflecting diminishing noise impulsiveness. Across the entire range, IRHTP achieves the highest SNR, followed by IRHT, while LIHT [4] and HIHT [5] exhibit comparatively lower performance.



(a) $\gamma = 0.8$ (b) $\gamma = 5$

Fig. 2: SNR vs. α

Appendix A
Proof of Lemma 2.1

Proof. **Update for $\boldsymbol{x}^k$:** Define $\boldsymbol{W}_k = \texttt{diag}\left(\boldsymbol{w}^k\right)$. The functional $\mathcal{J}(\boldsymbol{x}, \boldsymbol{x}^k, \boldsymbol{w}^k, \epsilon)$ expands as

$$\begin{aligned}
&\mathcal{J}(\boldsymbol{x}, \boldsymbol{x}^k, \boldsymbol{w}^k, \epsilon) \\
&= \frac{p}{2}\left(\left\|\boldsymbol{W}_k^{1/2}(\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x})\right\|_2^2 - \left\|\boldsymbol{W}_k^{1/2}(\boldsymbol{\Phi}\boldsymbol{x}-\boldsymbol{\Phi}\boldsymbol{x}^k)\right\|_2^2\right) \\
&+ \frac{p}{2}\frac{\left\|\boldsymbol{w}^k\right\|_\infty}{\mu}\left\|\boldsymbol{x}-\boldsymbol{x}^k\right\|_2^2 + \text{terms independent of } \boldsymbol{x} \\
&= \frac{p}{2}\sum_{i=1}^{n}\left[\frac{\left\|\boldsymbol{w}^k\right\|_\infty}{\mu}x_i^2\right. \\
&\left.-2x_i\left(\frac{\left\|\boldsymbol{w}^k\right\|_\infty}{\mu}x_i^k + \boldsymbol{\phi}_i^T\boldsymbol{W}_k\left(\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x}^k\right)\right)\right] \\
&+ \text{terms independent of } \boldsymbol{x}.
\end{aligned}$$

Let $T = \text{supp}(\boldsymbol{x})$. Since $w_j^k > 0$ and $\mu > 0$, each coordinate term is strictly convex and quadratic in x_i, so the minimizer over a fixed support T satisfies

$$x_i' = \begin{cases} x_i^k + \mu\boldsymbol{\phi}_i^T\boldsymbol{D}_k\left(\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x}^k\right), & i \in T, \\ 0, & i \notin T, \end{cases}$$

where $\boldsymbol{D}_k = \boldsymbol{W}_k / \left\|\boldsymbol{w}^k\right\|_\infty$. The value of the functional at this minimizer (up to terms independent of T) equals

$$-\frac{p\left\|\boldsymbol{w}^k\right\|_\infty}{2\mu}\sum_{i\in T}\left(x_i^k + \mu\boldsymbol{\phi}_i^T\boldsymbol{D}_k\left(\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x}^k\right)\right)^2,$$

which is minimized when T contains the indices of the K largest-magnitude entries of $\boldsymbol{x}^k + \mu\boldsymbol{\Phi}^T\boldsymbol{D}_k(\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x}^k)$. This yields the IRHT update

$$\boldsymbol{x}^{k+1} = H_K\left(\boldsymbol{x}^k + \mu\boldsymbol{\Phi}^T\boldsymbol{D}_k\left(\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x}^k\right)\right),$$

and the IRHTP update, which incorporates an additional orthogonal projection (pursuit step):

$$\Lambda^{k+1} = L_K(\boldsymbol{u}^{k+1}),$$

where $\boldsymbol{u}^{k+1} = \boldsymbol{x}^k + \mu\tilde{\boldsymbol{\Phi}}_k^T(\tilde{\boldsymbol{y}}_k - \tilde{\boldsymbol{\Phi}}_k\boldsymbol{x}^k)$, $\tilde{\boldsymbol{y}}_k = \boldsymbol{D}_k^{1/2}\boldsymbol{y}$, $\tilde{\boldsymbol{\Phi}}_k = \boldsymbol{D}_k^{1/2}\boldsymbol{\Phi}$, followed by

$$\boldsymbol{x}^{k+1} = \underset{\text{supp}(\boldsymbol{x})\subseteq\Lambda^{k+1}}{\arg\min}\left\|\tilde{\boldsymbol{y}}_k - \tilde{\boldsymbol{\Phi}}_k\boldsymbol{x}\right\|_2^2,$$

which is equivalent to $\boldsymbol{x}^{k+1} = \tilde{\boldsymbol{\Phi}}_{k,\Lambda^{k+1}}^{\dagger}\tilde{\boldsymbol{y}}_k$.

Update for $\boldsymbol{w}^k$: The expression for $\boldsymbol{w}^{k+1}$ is obtained by differentiating $\mathcal{J}(\boldsymbol{x}^{k+1}, \boldsymbol{x}^{k+1}, \boldsymbol{w}, \epsilon)$ with respect to each w_j and setting the result to zero. Since each term is a strictly convex quadratic in $w_j > 0$, the unique minimizer is given coordinate-wise by (7). ■

Appendix B
Proof of Theorem 3.1

Proof. **IRHT convergence analysis:** Applying Theorem 1 of [9],

$$\begin{aligned}
&\left\|\boldsymbol{x}^{k+1} - \boldsymbol{x}\right\|_2 \\
&\le \omega\left\|(\boldsymbol{x}^k - \boldsymbol{x} + \mu\boldsymbol{\Phi}^T\boldsymbol{D}_k(\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x}^k))_{\Lambda^{k+1}\cup\Lambda}\right\|_2 \\
&\le \omega\left\|\boldsymbol{I}_{S^{k+1}} - \mu\boldsymbol{\Phi}_{S^{k+1}}^T\boldsymbol{D}_k\boldsymbol{\Phi}_{S^{k+1}}\right\|_{2\to2}\left\|\boldsymbol{x}^k - \boldsymbol{x}\right\|_2 \\
&+ \omega\mu\sqrt{1+\delta_{2K}}\left\|\boldsymbol{e}\right\|_2, \qquad (21)
\end{aligned}$$

where $S^{k+1} = \Lambda^{k+1}\cup\Lambda^k\cup\Lambda$, $\omega = \frac{1+\sqrt{5}}{2}$, and $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{e}$. We now bound $\left\|\boldsymbol{I}_{S^{k+1}} - \mu\boldsymbol{\Phi}_{S^{k+1}}^T\boldsymbol{D}_k\boldsymbol{\Phi}_{S^{k+1}}\right\|_{2\to2}$. Since $\mu(1+\delta_{3K}) < 1$,

$$\begin{aligned}
&\left\|\boldsymbol{I} - \mu\boldsymbol{\Phi}_{S^{k+1}}^T\boldsymbol{D}_k\boldsymbol{\Phi}_{S^{k+1}}\right\|_{2\to2} \\
&= \max\left\{1 - \mu\lambda_{\min}\left(\boldsymbol{\Phi}_{S^{k+1}}^T\boldsymbol{D}_k\boldsymbol{\Phi}_{S^{k+1}}\right),\right. \\
&\qquad \left.\mu\lambda_{\max}\left(\boldsymbol{\Phi}_{S^{k+1}}^T\boldsymbol{D}_k\boldsymbol{\Phi}_{S^{k+1}}\right) - 1\right\} \\
&= 1 - \mu\lambda_{\min}\left(\boldsymbol{\Phi}_{S^{k+1}}^T\boldsymbol{D}_k\boldsymbol{\Phi}_{S^{k+1}}\right) \\
&\le 1 - \mu(1-\delta_{3K})\min_i D_{k,ii}. \qquad (22)
\end{aligned}$$

Denoting $\gamma_k = \min_i D_{k,ii}$ and substituting above gives the IRHT bound

$$\begin{aligned}
&\left\|\boldsymbol{x}^{k+1} - \boldsymbol{x}\right\|_2 \\
&\le \omega(1-\mu(1-\delta_{3K})\gamma_k)\left\|\boldsymbol{x}^k - \boldsymbol{x}\right\|_2 + \omega\mu\sqrt{1+\delta_{2K}}\left\|\boldsymbol{e}\right\|_2. \qquad (23)
\end{aligned}$$

IRHTP convergence analysis: Since $\boldsymbol{x}^{k+1}$ is the least-squares solution over Λ^{k+1}, it satisfies the orthogonality condition $\tilde{\boldsymbol{\Phi}}_{\Lambda^{k+1}}^T(\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{\Phi}}\boldsymbol{x}^{k+1}) = \boldsymbol{0}$. Therefore,

$$\begin{aligned}
&\left\|(\boldsymbol{x}^{k+1} - \boldsymbol{x}^*)_{\Lambda^{k+1}}\right\|_2^2 \\
&= \left\langle(\boldsymbol{x}^{k+1} - \boldsymbol{x}^*)_{S^{k+1}}, (\boldsymbol{I} - \tilde{\boldsymbol{\Phi}}^T\tilde{\boldsymbol{\Phi}})(\boldsymbol{x}^{k+1} - \boldsymbol{x}^*)\right\rangle \\
&+ \left\langle(\boldsymbol{x}^{k+1} - \boldsymbol{x}^*)_{\Lambda^{k+1}}, \tilde{\boldsymbol{\Phi}}^T\tilde{\boldsymbol{\Phi}}(\boldsymbol{x}^{k+1} - \boldsymbol{x}^*)\right\rangle \\
&\le (1 - \gamma_k(1-\delta_{2K}))\left\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\right\|_2\left\|(\boldsymbol{x}^{k+1} - \boldsymbol{x}^*)_{\Lambda^{k+1}}\right\|_2 \\
&+ \sqrt{1+\delta_K}\left\|(\boldsymbol{x}^{k+1} - \boldsymbol{x}^*)_{\Lambda^{k+1}}\right\|_2\left\|\boldsymbol{e}\right\|_2.
\end{aligned}$$

Setting $\Gamma = (1 - \gamma_k(1-\delta_{2K}))$ and dividing through,

$$\left\|(\boldsymbol{x}^{k+1} - \boldsymbol{x}^*)_{\Lambda^{k+1}}\right\|_2 \le \Gamma\left\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\right\|_2 + \sqrt{1+\delta_K}\left\|\boldsymbol{e}\right\|_2. \qquad (24)$$

Since $\left\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\right\|_2^2 = \left\|(\boldsymbol{x}^{k+1} - \boldsymbol{x}^*)_{\Lambda^{k+1}}\right\|_2^2 +$

$\left\|(\boldsymbol{x}^{k+1}-\boldsymbol{x}^*)_{\overline{\Lambda^{k+1}}}\right\|_2^2$, substituting (24) gives

$$\left\|(\boldsymbol{x}^{k+1}-\boldsymbol{x}^*)\right\|_2^2 \leq \left(\Gamma\left\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^*\right\|_2+\sqrt{1+\delta_K}\left\|\boldsymbol{e}\right\|_2\right)^2 + \left\|(\boldsymbol{x}^{k+1}-\boldsymbol{x}^*)_{\overline{\Lambda^{k+1}}}\right\|_2^2.$$

Generalizing the analysis of [10, Thm. 3.8], $\left\|(\boldsymbol{x}^{k+1}-\boldsymbol{x}^*)\right\|_2$ is bounded by the largest root of the quadratic

$$f(t) := (1-\Gamma^2)t^2-(2\Gamma\sqrt{1+\delta_K}\left\|\boldsymbol{e}\right\|_2)t - \left(\left\|(\boldsymbol{x}^{k+1}-\boldsymbol{x}^*)_{\overline{\Lambda^{k+1}}}\right\|_2^2+(1+\delta_K)\left\|\boldsymbol{e}\right\|_2^2\right),$$

giving

$$\left\|(\boldsymbol{x}^{k+1}-\boldsymbol{x}^*)\right\|_2 \leq \frac{\Gamma\sqrt{1+\delta_K}\left\|\boldsymbol{e}\right\|_2}{1-\Gamma^2} + \frac{\sqrt{(1-\Gamma^2)\left\|(\boldsymbol{x}^{k+1}-\boldsymbol{x}^*)_{\overline{\Lambda^{k+1}}}\right\|_2^2+(1+\delta_K)\left\|\boldsymbol{e}\right\|_2^2}}{1-\Gamma^2}$$

Using $\sqrt{a^2+b^2}\leq a+b$ for $a,b\geq 0$,

$$\left\|(\boldsymbol{x}^{k+1}-\boldsymbol{x}^*)\right\|_2 \leq \frac{\left\|(\boldsymbol{x}^{k+1}-\boldsymbol{x}^*)_{\overline{\Lambda^{k+1}}}\right\|_2}{\sqrt{1-(1-\gamma_k(1-\delta_{2K}))^2}}+\frac{\sqrt{1+\delta_K}\left\|\boldsymbol{e}\right\|_2}{\gamma_k(1-\delta_{2K})}. \quad (25)$$

With

$$\left\|(\boldsymbol{x}^{k+1}-\boldsymbol{x}^*)_{\overline{\Lambda^{k+1}}}\right\|_2 \leq \sqrt{2}(1-\gamma_k(1-\delta_{3K}))\left\|(\boldsymbol{x}^{k}-\boldsymbol{x}^*)\right\|_2 + \sqrt{2(1+\delta_{2K})}\left\|\boldsymbol{e}\right\|_2, \quad (26)$$

combining (25) and (26) yields the IRHTP bound

$$\left\|(\boldsymbol{x}^{k+1}-\boldsymbol{x}^*)\right\|_2 \leq \sqrt{\frac{2(1-\gamma_k(1-\delta_{3K}))^2}{1-(1-\gamma_k(1-\delta_{2K}))^2}}\left\|(\boldsymbol{x}^{k}-\boldsymbol{x}^*)\right\|_2 + \frac{\sqrt{2\gamma_k(1-\delta_{2K})}+\sqrt{1+\delta_K}}{\gamma_k(1-\delta_{2K})}\left\|\boldsymbol{e}\right\|_2. \quad (27)$$

Lower bound on γ_k: Observing that

$$\gamma_k = \min_i D_{k,ii} = \left(\frac{\min_j (y_j-(\boldsymbol{\Phi}\boldsymbol{x}^k)_j)^2+\epsilon^2}{\max_j (y_j-(\boldsymbol{\Phi}\boldsymbol{x}^k)_j)^2+\epsilon^2}\right)^{1-p/2} \geq \left(\frac{\epsilon^2}{\max_j (y_j-(\boldsymbol{\Phi}\boldsymbol{x}^k)_j)^2+\epsilon^2}\right)^{1-p/2}. \quad (28)$$

Let $\pi_k = \arg\max_j \left((y_j-(\boldsymbol{\Phi}\boldsymbol{x}^k)_j)^2+\epsilon^2\right)$. Then,

$$\begin{aligned}
&\left[\max_j (y_j-(\boldsymbol{\Phi}\boldsymbol{x}^k)_j)^2+\epsilon^2\right]^{p/2} \\
&\leq \sum_{j=1}^{m}\left[(y_j-(\boldsymbol{\Phi}\boldsymbol{x}^k)_j)^2+\epsilon^2\right]^{p/2}-(m-1)\epsilon^p \\
&= \mathcal{J}(\boldsymbol{x}^k,\boldsymbol{x}^k,\boldsymbol{w}^k,\epsilon)-(m-1)\epsilon^p \\
&\overset{\text{Lemma 3.2}}{\leq} \mathcal{J}(\boldsymbol{x}^0,\boldsymbol{x}^0,\boldsymbol{w}^0,\epsilon)-(m-1)\epsilon^p \\
&\leq \left\|\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x}^0\right\|_p^p+\epsilon^p. \quad (29)
\end{aligned}$$

Substituting (29) into (28) yields a k-independent lower bound.

Defining $\gamma = \min_{k\geq 0}\gamma_k$,

$$\gamma \geq \frac{\epsilon^{2-p}}{\left(\left\|\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{x}^0\right\|_p^p+\epsilon^p\right)^{2/p-1}}. \quad (30)$$

Substituting into the IRHT and IRHTP bounds gives (14). When $\boldsymbol{e}=\boldsymbol{0}$, $\rho_{3K}<1$ implies $\boldsymbol{x}^k\to\boldsymbol{x}$ as $k\to\infty$. When $\boldsymbol{e}\neq\boldsymbol{0}$, the iterates converge with linear rate to a local minimum $\boldsymbol{x}^\star$ of (P1), and (17) follows from (30) and the requirement $\rho_{3K}<1$. ■

REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] A. Kumar, S. Jain, and S. Majhi, "Channel estimator and symbol detector for otfs systems impaired by impulsive noise," *IEEE Communications Letters*, vol. 27, no. 10, pp. 2742–2746, 2023.

[3] W.-J. Zeng, H.-C. So, and A. M. Zoubir, "An l_p-norm minimization approach to time delay estimation in impulsive noise," *Digital Signal Process.*, vol. 23, no. 4, pp. 1247–1254, 2013.

[4] R. E. Carrillo and K. E. Barner, "Lorentzian iterative hard thresholding: Robust compressed sensing with prior information," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4822–4833, 2013.

[5] E. Ollila, H.-J. Kim, and V. Koivunen, "Robust iterative hard thresholding for compressed sensing," in *Communications, Control and Signal Processing (ISCCSP), 2014 6th International Symposium on*. IEEE, 2014, pp. 226–229.

[6] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Commun. Pure Appl. Math.*, vol. 63, no. 1, pp. 1–38, 2010.

[7] J. Sigl, "Nonlinear residual minimization by iteratively reweighted least squares," *Comput. Optim. Appl.*, pp. 1–38, 2015.

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

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

[10] S. Foucart, "Hard thresholding pursuit: an algorithm for compressive sensing," *SIAM Journal on numerical analysis*, vol. 49, no. 6, pp. 2543–2563, 2011.