Iterative Soft/Hard Thresholding with Homotopy Continuation for Sparse Recovery

Yuling Jiao, Bangti Jin, Xiliang Lu

***Abstract*—In this note, we analyze an iterative soft / hard thresholding algorithm with homotopy continuation for recovering a sparse signal $x^\dagger$ from noisy data of a noise level ϵ. Under suitable regularity and sparsity conditions, we design a path along which the algorithm can find a solution x^* which admits a sharp reconstruction error $\|x^* - x^\dagger\|_{\ell^\infty} = O(\epsilon)$ with an iteration complexity $O(\frac{\ln \epsilon}{\ln \gamma} np)$, where n and p are problem dimensionality and $\gamma \in (0,1)$ controls the length of the path. Numerical examples are given to illustrate its performance.**

***Index Terms*—iterative soft/hard thresholding, continuation, solution path, convergence**

I. Introduction

SPARSE recovery has attracted much attention in machine learning, signal processing, statistics and inverse problems over the last decade. Often the problem is formulated as

$$y = \Psi x^\dagger + \eta, \tag{1}$$

where $x^\dagger \in \mathbb{R}^p$ is the unknown sparse signal, $y \in \mathbb{R}^n$ is the data with the noise $\eta \in \mathbb{R}^n$ of level $\epsilon = \|\eta\|$, and the matrix $\Psi \in \mathbb{R}^{n\times p}$ with $p \gg n$ has normalized columns $\{\psi_i\}$, i.e., $\|\psi_i\| = 1$, $i = 1, \ldots, p$. The desired sparsity structure can be enforced by either the ℓ^0 or ℓ^1 penalty, i.e.,

$$\min_{x\in\mathbb{R}^p} \tfrac{1}{2}\|\Psi x - y\|^2 + \lambda \|x\|_t, \quad t \in \{0, 1\}, \tag{2}$$

where $\lambda > 0$ is the regularization parameter.

Among existing algorithms for minimizing (2), iterative soft / hard thresholding (IST/IHT) algorithm [1]–[4] and their accelerated extension [5], [6] are extremely popular. These algorithms are of the form

$$x^{k+1} = T_{\tau_k\lambda}(x^k + \tau_k \Psi^t(y - \Psi x^k)), \tag{3}$$

where τ_k is the stepsize, and T_λ is a soft- or hard-thresholding operator defined componentwise by

$$T_\lambda(t) = \begin{cases} \max(|t| - \lambda, 0)\mathrm{sgn}(t), & \text{IST}, \\ \chi_{\{|t|>\sqrt{2\lambda}\}}(t), & \text{IHT}, \end{cases} \tag{4}$$

where $\chi(t)$ is the characteristic function. Their convergence was analyzed in many works, mostly under the condition $\tau_k < 2/\|\Psi\|^2$. This condition ensures a (asymptotically) contractive thresholding and thus the desired convergence [1]–[4]. Meanwhile, it was observed that the continuation along λ can greatly speed up the algorithms [6]–[10]. Nonetheless, as pointed out by [11] "... the design of a robust, practical, and theoretically effective continuation algorithm remains an interesting open question ..." There were several works aiming at filling this gap. In the works [12], [13], a proximal gradient method with continuation for ℓ^1 problem was analyzed with linear search, under sparse restricted eigenvalue/restricted strong convexity condition. Recently, a Newton type method with continuation was studied for ℓ^1 and ℓ^0 problems [14], [15]. In this work, we present a unified approach to analyze IST/IHT with continuation and a fixed stepsize $\tau = 1$, denoted by ISTC/IHTC. The challenge in the analysis is the lack of monotonicity of function values due to the choice $\tau = 1$.

The overall procedure is given in Algorithm 1. Here λ_0 is an initial guess of λ, supposedly large, $\gamma \in (0,1)$ is the decreasing factor for λ, and K_{max} is the maximum number of inner iterations (for a fixed λ). The choice of the final λ^* is given in (5) below. Distinctly, the inner iteration does not need to be solved exactly (actually one inner iteration suffices the desired accuracy of the final solution x^*, cf. Theorem 2 below), and there is no need to perform stepsize selection.

Algorithm 1 Iterative Soft/Hard-Thresholding with Continuation (ISTC/IHTC)

1: Input: $\Psi \in \mathbb{R}^{n\times p}$, y, λ_0, $\gamma \in (0,1)$, λ^*, $K_{max} \in \mathbb{N}$, $x(\lambda_0) = 0$.
2: **for** $\ell = 1, 2, \ldots$ **do**
3: Let $\lambda_\ell = \gamma\lambda_{\ell-1}$, $x^0 = x(\lambda_{\ell-1})$.
4: If $\lambda_\ell < \lambda^*$, stop and output $x^* = x^0$.
5: **for** $k = 0, 1, \ldots, K_{max} - 1$ **do**
6: $x^{k+1} = T_{\lambda_\ell}(x^k + \Psi^t(y - \Psi x^k))$.
7: **end for**
8: Set $x(\lambda_\ell) = x^{K_{max}}$
9: **end for**

In Theorem 2, we prove that under suitable mutual coherence condition on the matrix Ψ (cf. Assumption 2.1 and Remark 2.2), ISTC/IHTC always converges.

II. Convergence analysis

The starting point of our analysis is the next lemma.

Lemma 1: For any $x, y \in \mathbb{R}$, there holds

$$|T_\lambda(x + y) - x| \le \begin{cases} |y| + \lambda & \text{IST}, \\ |y| + \sqrt{2\lambda} & \text{IHT}. \end{cases}$$

School of Statistics and Mathematics and Big Data Institute of ZUEL, Zhongnan University of Economics and Law, Wuhan, 430063, P.R. China. (yulingjiaomath@whu.edu.cn)

Department of Computer Science, University College London, Gower Street, London WC1E 6BT, UK. (bangti.jin@gmail.com, b.jin@ucl.ac.uk)

Corresponding author. School of Mathematics and Statistics and Hubei Key Laboratory of Computational Science, Wuhan University, Wuhan 430072, P.R. China. (xllv.math@whu.edu.cn)

Proof: By the definition of the operator T_λ, cf. (4),

$$|T_\lambda(x+y)-x| \le |T_\lambda(x+y)-(x+y)|+|y| \le \begin{cases} |y|+\lambda & \text{IST}, \\ |y|+\sqrt{2\lambda} & \text{IHT}, \end{cases}$$

which completes the proof of the lemma. ∎

Let the true signal $x^\dagger$ be s-sparse with a support $\mathcal{A}^\dagger$, i.e., $s = |\mathcal{A}^\dagger|$, and $\mathcal{I}^\dagger$ the complement of $\mathcal{A}^\dagger$. Recall also that the mutual coherence (MC) μ of the matrix Ψ is defined by $\mu = \max_{i\neq j} |\langle \psi_i, \psi_j\rangle|$ [16].

Assumption 2.1: The MC μ of Ψ satisfies $\mu s < 1/2$.

The proper choice of the regularization parameter λ is essential for successful sparse recovery. It is well known that under Assumption 2.1, the choice $\lambda = O(\epsilon)$ for the ℓ_1 penalty and $\lambda = O(\epsilon^2)$ for the ℓ_0 penalty ensures $\|x - x^\dagger\|_{\ell^\infty} = O(\epsilon)$ [15], [17]. Thus we consider the following *a priori* choice

$$\lambda^* = \begin{cases} C_1\epsilon, \text{ with } C_1 > \frac{1}{1-2\mu s}, & \text{for ISTC}, \\ C_0\epsilon^2, \text{ with } C_0 > \frac{1}{2(1-2\mu s)^2}, & \text{for IHTC}. \end{cases} \tag{5}$$

In practice, one may consider *a posteriori* choice rules [18]. Now we can state the global convergence of Algorithm 1.

Theorem 2: Let Assumption 2.1 hold, and λ^* be chosen by (5). Suppose that λ_0 is large, $K_{max} \in \mathbb{N}$, and

$$\gamma \in \begin{cases} [2\mu s/(1-1/C_1), 1), & \text{for ISTC}, \\ [(\frac{2\mu s}{1-1/(2C_0)^{1/2}})^2, 1), & \text{for IHTC}. \end{cases}$$

Then Algorithm 1 is well-defined, and the solution x^* satisfies:

(i) $\operatorname{supp}(x^*) \subset \mathcal{A}^\dagger$,

(ii) there holds the error estimate

$$\|x^* - x^\dagger\|_{\ell^\infty} \le \begin{cases} (C_1-1)\epsilon/(\mu s), & \text{for ISTC}, \\ (\sqrt{2C_0}-1)\epsilon/(\mu s), & \text{for IHTC}. \end{cases}$$

Further, if $\min_{i\in\mathcal{A}^\dagger} |x_i^\dagger|$ is large enough, then $\operatorname{supp}(x^*) = \mathcal{A}^\dagger$.

Proof: We only prove the assertion for ISTC, since that for IHTC is similar. The choice of C_1 in (5) implies $C_1 > 1$ and $\frac{2\mu s}{1-1/C_1} < 1$, and thus the choice of γ makes sense.

First we consider the inner loop at lines 5 - 7 of Algorithm 1 and omit the index ℓ for notational simplicity. Let $E^k = \|x^k - x^\dagger\|_{\ell^\infty}$, and $\alpha = \frac{1-1/C_1}{\mu s}$. Consider one IST iteration from x^k to x^{k+1}. The key step to the convergence proof is the following implication: with $\mathcal{A}^k = \operatorname{supp}(x^k)$

$$\begin{aligned} &\mathcal{A}^k \subset \mathcal{A}^\dagger \text{ and } E^k \le \alpha\lambda \\ \Rightarrow &\mathcal{A}^{k+1} \subset \mathcal{A}^\dagger \text{ and } E^{k+1} \le \alpha\gamma\lambda \quad \forall \lambda \ge \lambda^*. \end{aligned} \tag{6}$$

Now we show this claim. It follows from (1) and $\|\Psi_i\| = 1$ the following componentwise expression for the update

$$\begin{aligned} x_i^{k+1} &= T_\lambda(x_i^k + \Psi_i^t(y - \Psi x^k)) \\ &= T_\lambda(x_i^\dagger + \Psi_i^t(\Psi_{\mathcal{A}^\dagger\cup\mathcal{A}^k\setminus\{i\}}(x^\dagger - x^k)_{\mathcal{A}^\dagger\cup\mathcal{A}^k\setminus\{i\}} + \eta)). \end{aligned}$$

By the hypothesis in (6), $\mathcal{A}^k \subset \mathcal{A}^\dagger$, $E^k \le \alpha\lambda$, $\lambda \ge \lambda^*$ and (5), we deduce that for any $i \in \mathcal{I}^\dagger$

$$\begin{aligned} &|x_i^\dagger + \Psi_i^t(\Psi_{\mathcal{A}^\dagger\cup\mathcal{A}^k\setminus\{i\}}(x^\dagger - x^k)_{\mathcal{A}^\dagger\cup\mathcal{A}^k\setminus\{i\}} + \eta)| \\ \le &|\Psi_i^t(\Psi_{\mathcal{A}^\dagger}(x^\dagger - x^k)_{\mathcal{A}^\dagger}| + |\Psi_i^t\eta| \\ \le &\mu s E^k + \epsilon \le (\tfrac{1}{C_1} + \mu s\alpha)\lambda = \lambda, \end{aligned}$$

by the definition of α, and the second inequality follows from [15, Lemma 2.1]. Hence, $|x_i^{k+1}| \le |T_\lambda(\mu s E^k + \epsilon)| = 0$, which implies directly $\mathcal{A}^{k+1} \subset \mathcal{A}^\dagger$. Meanwhile, under (6) and (5), for any $i \in \mathcal{A}^\dagger$, by Lemma 1, we deduce

$$\begin{aligned} |x_i^{k+1} - x_i^\dagger| &\le \lambda + |\Psi_i^t(\Psi_{\mathcal{A}^\dagger\setminus\{i\}}(x^\dagger - x^k)_{\mathcal{A}^\dagger\setminus\{i\}}| + |\Psi_i^t\eta| \\ &\le \lambda + \mu(s-1)E^k + \epsilon \le \lambda + \mu s\alpha\lambda + \tfrac{1}{C_1}\lambda \\ &= (1 + \tfrac{1}{C_1} + \alpha\mu s)\lambda = 2\lambda \le \alpha\gamma\lambda. \end{aligned}$$

Thus we have $E^{k+1} \le \alpha\gamma\lambda$, i.e., the claim (6) holds.

Next we prove the following assertion by mathematical induction: for all ℓ with $\lambda_\ell \ge \lambda^*$, there holds

$$\operatorname{supp} x(\lambda_\ell) \subset \mathcal{A}^\dagger, \quad \|x(\lambda_\ell) - x^\dagger\|_{\ell^\infty} \le \alpha\gamma\lambda_\ell. \tag{7}$$

Since λ_0 is large, it satisfies (7). Now assume (7) holds for $\lambda_{\ell-1}$, i.e., $\operatorname{supp}\ x(\lambda_{\ell-1}) \subset \mathcal{A}^\dagger$ and $\|x(\lambda_{\ell-1}) - x^\dagger\|_{\ell^\infty} \le \alpha\gamma\lambda_{\ell-1}$. When Algorithm 1 runs lines 3 - 7 for λ_ℓ, since $x^0 = x(\lambda_{\ell-1})$, then we have $\mathcal{A}^0 \subset \mathcal{A}^\dagger$ and $E^0 \le \alpha\lambda_\ell$. From (6), we obtain that for all $k \ge 1$, $\mathcal{A}^k \subset \mathcal{A}^\dagger$ and $E^k \le \alpha\gamma\lambda_\ell$. In particular, if we choose $k = K_{max}$, then (7) holds for λ_ℓ. When Algorithm 1 terminates for some $\lambda_\ell < \lambda^*$, then $\lambda_{\ell-1} \ge \lambda^*$ and $x^* = x(\lambda_{\ell-1})$. From (7) we have $\operatorname{supp} x^* \subset \mathcal{A}^\dagger$ and $\|x^* - x^\dagger\|_{\ell^\infty} \le \alpha\lambda^* = (C_1-1)\epsilon/(\mu s)$. Likewise, if $\min_{i\in\mathcal{A}^\dagger} |x_i| > (C_1-1)\epsilon/(\mu s)$, property (ii) implies $\operatorname{supp}(x^*) = \mathcal{A}^\dagger$.

Last, we briefly discuss IHTC. For the choice C_0 in (5), $\gamma \in [(\frac{2\mu s}{1-1/(2C_0)^{1/2}})^2, 1)$ makes sense. With $\alpha = \frac{1-1/(2C_0)^{1/2}}{\mu s}$, a similar argument yields

$$\begin{aligned} &\mathcal{A}^k \subset \mathcal{A}^\dagger \text{ and } E^k \le \alpha\sqrt{2\lambda} \\ \Rightarrow &\mathcal{A}^{k+1} \subset \mathcal{A}^\dagger \text{ and } E^{k+1} \le \alpha\sqrt{2\gamma\lambda}. \end{aligned}$$

The rest follows like before, and thus it is omitted. ∎

Remark 2.1: The proof works for any choice $K_{max} \ge 1$, including $K_{max} = 1$. In practice, we fix it at $K_{max} = 5$. This together with Theorem 2 allows estimating the complexity of Algorithm 1. At each iteration, one needs to compute matrix-vector product Ψx and $\Psi^t y$, and for each λ, the number of iterations is bounded by K_{max}. The overall cost depends on the decreasing factor γ by $O(\frac{\ln\lambda^*}{\ln\gamma} np) = O(\frac{\ln\epsilon}{\ln\gamma} np)$.

Remark 2.2: Conditions similar to Assumption 2.1 have been widely used in the literature, for analyzing OMP [17], [19], [20] (with $(2s-1)\mu \le 1$) and for bounding the estimation error of Lasso [21], [22] (with $7s\mu < 1$ and $4s\mu \le 1$). Thus Assumption 2.1 is fairly standard. Examples of matrices with small MC μ include that formed by equiangular tight frame and random subgaussian matrices [23]. Further, we note that other similar conditions, e.g., restricted eigenvalue condition and RIP conditions, were also used to derive error bounds of the type $\|x - x^\dagger\|_2 = O(\epsilon)$ for proximal gradient homotopy algorithms [12], [13] and Greedy methods, e.g., CoSaMP [24], NIHT [25] and CGIHT [26].

III. Numerical Results and Discussions

Now we present numerical examples to show the convergence and the performance of Algorithm 1. First, we give implementation details, e.g., data generation, parameter setting for the algorithm. Then our method is compared with several

state-of-the-art algorithms in terms of reconstruction error and recovery ability via phase transition.

A. Implementation details

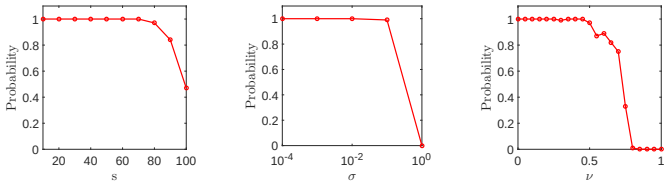
Following [6], the signals $x^\dagger$ are chosen as s-sparse with a dynamic range $DR := \max\{|x_i^\dagger| : x_i^\dagger \neq 0\}/\min\{|x_i^\dagger| : x_i^\dagger \neq 0\}$. The matrix $\Psi \in \mathbb{R}^{n\times p}$ is chosen to be either random Gaussian matrix, or random Bernoulli matrix, or the product of a partial FFT matrix and inverse Haar wavelet transform. Under proper conditions, such matrices satisfy Assumption 2.1. The noise η has entries following i.i.d. $N(0,\sigma^2)$.

We fix the algorithm parameters as follows: $\lambda_0 = \|\Psi^t y\|_\infty$ and $\lambda_0 = \|\Psi^t y\|_\infty^2/2$ for ISTC and IHTC, respectively [14], [15], decreasing factor $\gamma = 0.8$. Since the optimal λ^* depends on the noise level ϵ, which is often unknown in practice, we predefine a path $\Lambda = \{\lambda_\ell\}_{\ell=0}^N$ with $\lambda_\ell = \lambda_0\gamma^\ell$ and $N = 100$. Then we run Algorithm 1 on the path Λ and select the optimal λ^* by Bayesian information criterion [14]. All the computations were performed on an eight-core desktop with 3.40 GHz and 12 GB RAM using `MATLAB` 2014a. The `MATLAB` package `ISHTC` for reproducing all the numerical results can be found at http://www0.cs.ucl.ac.uk/staff/b.jin/companioncode.html.

First we illustrate Theorem 2 by examining the influence of sparsity level s, coherence μ and noise level σ on IHTC recovery on three settings ($n = 500$, $p = 1000$, $DR = 100$):

(a) random Gaussian Ψ, σ =1e-2, $s = 10 : 10 : 100$.
(b) random Gaussian Ψ, $s = 50$, σ =1e-4,1e-3,1e-2,1e-1,1.
(c) Ψ is random Gaussian with $\nu = 0 : 0.05 : 1$ (a larger ν gives a larger μ, cf. [27, Sect. 5.1]), $s = 10$, σ = 1e-3.

The results in Fig. 1 are computed from 100 independent realizations. It is observed that when the sparsity level s and noise level σ and incoherence ν are small, IHTC recovers the exact support with high probability as implied by Theorem 2.



(a) s (b) σ (c) ν

Fig. 1: The exact support recovery probability v.s. s, σ and ν

B. Comparison of ISTC with ℓ^1 solvers

Now we compare ISTC with four state-of-the-art ℓ^1 solvers: GPSR [8] (http://www.lx.it.pt/mtf/GPSR/), SpaRSA [9] (http://www.lx.it.pt/mtf/SpaRSA/), proximal-gradient homotopy method (PGH) [12] (https://www.microsoft.com/en-us/download/details.aspx?id=52421), and FISTA [5] (implemented as https://web.iem.technion.ac.il/images/user-files/becka/papers/wavelet_FISTA.zip)[1].

The numerical results (CPU time, number of matrix-vector multiplications (nMV), relative ℓ_2 error (Reℓ_2), and absolute ℓ_∞ error (Abℓ_∞)) are computed from 10 independent realizations of for random Bernoulli sensing matrices with different parameter tuples (n, p, s, DR, σ) are shown in Tables I. It is observed that ISTC yields reconstructions that are comparable with that by other methods but at least two to three times faster. Further, it scales well with the problem size p.

TABLE I: Numerical results (CPU time and errors), with random Bernoulli Ψ, of size p = 10000, 14000, 18000, $n = \lfloor p/4 \rfloor$, $s = \lfloor n/40 \rfloor$, with $DR = 100$ and σ = 5e-2.

p	method	time (s)	nMV	Reℓ^2	Abℓ^∞
10000	ISTC	1.0	58	4.21e-3	2.66e-1
	PGH	1.7	419	4.14e-3	2.66e-1
	SpaRSA	3.4	302	4.13e-3	2.63e-1
	GPSR	3.0	256	4.25e-3	2.71e-1
	FISTA	5.3	505	4.30e-3	2.65e-1
14000	ISTC	2.0	58	4.30e-3	2.71e-1
	PGH	3.4	431	4.21e-3	2.68e-1
	SpaRSA	6.8	306	4.21e-3	2.67e-1
	GPSR	5.7	258	4.32e-3	2.75e-1
	FISTA	10.1	493	4.60e-3	2.76e-1
18000	ISTC	3.3	58	4.34e-3	2.88e-1
	PGH	5.6	443	4.25e-3	2.85e-1
	SpaRSA	11.4	309	4.25e-3	2.84e-1
	GPSR	9.5	258	4.36e-3	2.91e-1
	FISTA	17.2	506	4.40e-3	2.74e-1

Next, we compare the empirical performance of ISTC with other methods by their phase transition curves in the ρ-δ plane, with $\rho = s/n$ and $\delta = n/p$. When computing the curves, we fix the dimension $p = 1000$, and partition the range $(\delta, \rho) \times [0.1, 1]^2$ into a 30×30 equally spaced grid, and run 100 independent simulations at each grid point. The s-sparse signal $x^\dagger \in \mathbb{R}^p$, matrix $\Psi \in \mathbb{R}^{n\times p}$, and data $y \in \mathbb{R}^n$ are generated as [28, Fig. 13]. Fig. 2 plots the logistic regression curves identifying the 90% success rate for the algorithms. IHTC exhibits similar phase transition behavior as other methods.

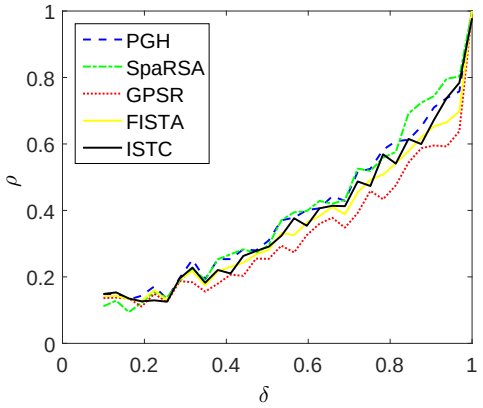


Fig. 2: The empirical phase transition curves for ISTC, PGH, SpaRSA and GPSR, with $\rho = s/n$ and $\delta = n/p$.

C. Comparison of IHTC with greedy solvers

Now we compare IHTC with four state-of-the-art greedy methods for the ℓ^0 problem, to recover 1D signal and benchmark MRI image. These methods include OMP [19] (https://sparselab.stanford.edu/SparseLab_

[1]All the codes were last accessed on February 23, 2017.

Left: 1D signal with $n = 665$, $p = 1024$, $s = 247$, and σ=1e-4. Right: 2D image with $n = 34489$, $p = 262144$, $s = 7926$, and σ=3e-2.

TABLE II: 1D signal

method	CPU time	PSNR
IHTC	0.41	51
OMP	1.20	49
NIHT	0.96	46
CoSaMP	0.49	26
CGIHT	0.98	49

TABLE III: 2D image

method	CPU time	PSNR
IHTC	6.1	28
OMP	932	28
NIHT	9.4	27
CoSaMP	14.3	26
CGIHT	7.9	27

files/Download_files/SparseLab21-Core.zip), normalized IHT (NIHT) [25] (http://www.gaga4cs.org/), CoSaMP [24] (http://mdav.ece.gatech.edu/software/SSCoSaMP-1.0.zip), and conjugate gradient IHT (CGIHT) [26] (http://www.gaga4cs.org/).

The underlying 1D signal and 2D MRI image are compressible under a wavelet basis. Thus, the data can be chosen as the wavelet coefficients sampled by the product of a partial FFT matrix and inverse Haar wavelet transform. For the 1D signal, the matrix Ψ is of size 665×1024, and consists of applying a partial FFT and an inverse two level Harr wavelet transform. The signal under wavelet transform has 247 nonzeros, and σ = 1e-4. The results are shown in Fig. 3 and Table II. The reconstruction by IHTC is visually more appealing than that of the others, cf. Fig. 3. The results by AIHT and CoSaMP suffer from pronounced oscillations. This is further confirmed by the PSNR value defined by $\mathrm{PSNR} = 10 \cdot \log \frac{V^2}{\mathrm{MSE}}$, where V is the maximum absolute value of the true signal, and MSE is the mean squared error of the reconstruction. Table II also presents the CPU time of the 1D example, which shows clearly that IHTC is the fastest one.

For the 2D MRI image, the matrix Ψ amounts to a partial FFT and an inverse wavelet transform, and it has a size 34489×262144. The image under eight level Haar wavelet transformation has 7926 nonzero entries and σ = 3e-2. The numerical results are shown in Fig. 4 and Table III. All ℓ^0 methods produce comparable results, but the IHTC is fastest.

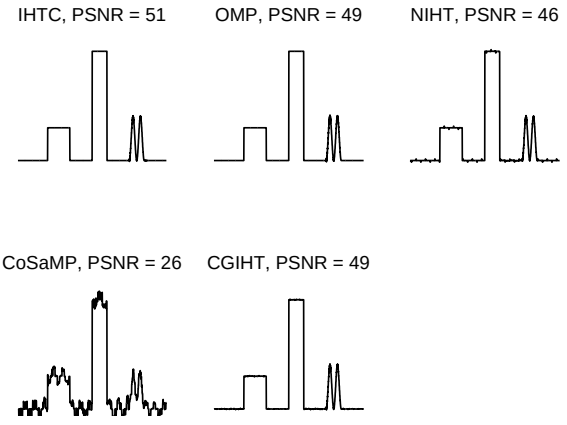


Fig. 3: Reconstructed signals and their PSNR values

Next, we compare the empirical sparse recovery performance of IHTC with these greedy methods by means of phase transition curves in the ρ-δ plane, with $\rho = s/n$ and $\delta = n/p$. When computing the curves, we fix the dimension $p = 1000$, partition the range $(\delta, \rho) \in [0.1, 1]^2$ into a 90×90 uniform grid, and run 100 independent simulations at each grid point.

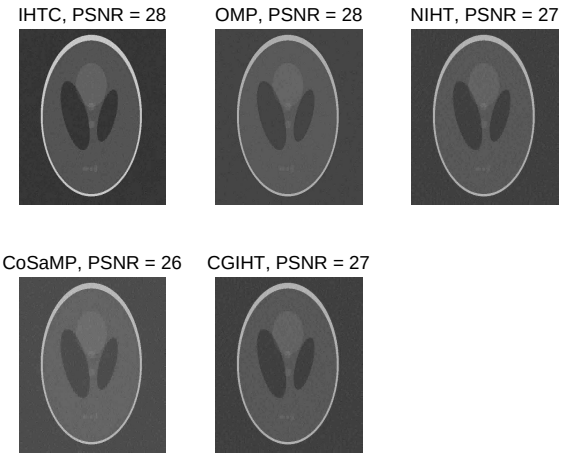


Fig. 4: Reconstructed MRI images and their PSNR values

Like before, the s-sparse signal $x^\dagger \in \mathbb{R}^p$, matrix $\Psi \in \mathbb{R}^{n \times p}$ and data $y \in \mathbb{R}^n$ are generated as [28, Fig. 13]. Fig. 5 plots the logistic regression curves identifying the 90% success rate for the algorithms. IHTC exhibits comparable phase transition phenomenon with other greedy methods, whereas CoSaMP performs slightly worse than others.

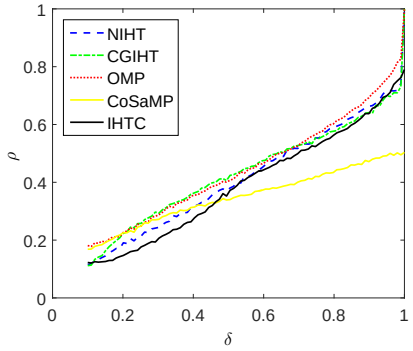


Fig. 5: The empirical phase transition curves of IHTC, OMP, CoSaMP, NIHT and CGIHT, with $\rho = s/n$ and $\delta = n/p$.

IV. Conclusion

In this paper, we analyze an iterative soft / hard thresholding algorithm with homotopy continuation for sparse recovery from noisy data. Under standard regularity condition and sparsity assumptions, sharp reconstruction errors can be obtained with an iteration complexity $O(\frac{\ln \epsilon}{\ln \gamma} np)$. Numerical results indicated its competitiveness with state-of-the-art sparse recovery algorithms. The results can be extended to other penalties, e.g., MCP [29] or SCAD [30].

Acknowledgements

The authors thank anonymous referees for their helpful comments. The research of Y. Jiao is partially supported by National Science Foundation of China (NSFC) No. 11501579 and National Science Foundation of Hubei Province No. 2016CFB486, B. Jin by EPSRC grant EP/M025160/1, and X. Lu by NSFC Nos. 11471253 and 91630313.

REFERENCES

[1] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.

[2] P. Combettes and V. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Model. Simul.*, vol. 4, no. 4, pp. 1168–1200, 2005.

[3] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *J. Fourier Anal. Appl.*, vol. 14, no. 5-6, pp. 629–654, 2008.

[4] H. Attouch, J. Bolte, and B. F. Svaiter, "Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods," *Math. Program.*, vol. 137, no. 1-2, pp. 91–129, 2013.

[5] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[6] S. Becker, J. Bobin, and E. Candés, "NESTA: a fast and accurate first-order method for sparse recovery," *SIAM J. Imag. Sci.*, vol. 4, no. 1, pp. 1–39, 2011.

[7] E. Hale, W. Yin, and Y. Zhang, "Fixed-point continuation for ℓ_1-minimization: Methodology and convergence," *SIAM J. Optim.*, vol. 19, no. 3, pp. 1107–1130, 2008.

[8] M. Figueiredo, R. Nowak, and S. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Topics Signal Proc.*, vol. 1, no. 4, pp. 586–597, 2007.

[9] S. Wright, R. Nowak, and M. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Proc.*, vol. 57, no. 7, pp. 2479–2493, 2009.

[10] D. A. Lorenz, "Constructing test instances for basis pursuit denoising," *IEEE Trans. Signal Proc.*, vol. 5, no. 61, pp. 1210–1214, 2013.

[11] J. Tropp and S. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, 2010.

[12] L. Xiao and T. Zhang, "A proximal-gradient homotopy method for the sparse least-squares problem," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1062–1091, 2013.

[13] A. Agawal, S. Negahban, and M. J. Wainwright, "Fast global convergence of gradient methods for high-dimensional statistical recovery," *Ann. Stat.*, vol. 40, no. 5, pp. 2452–2482, 2012.

[14] Q. Fan, Y. Jiao, and X. Lu, "A primal dual active set algorithm with continuation for compressed sensing," *IEEE Trans. Signal Proc.*, vol. 62, no. 23, pp. 6276–6285, 2014.

[15] Y. Jiao, B. Jin, and X. Lu, "A primal dual active set with continuation algorithm for the ℓ^0-regularized optimization problem," *Appl. Comput. Harmon. Anal.*, vol. 39, no. 3, pp. 400–426, 2015.

[16] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.

[17] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, 2006.

[18] K. Ito and B. Jin, *Inverse Problems: Tikhonov Theory and Algorithms*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2015.

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

[20] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4680–4688, 2011.

[21] K. Lounici, "Sup-norm convergence rate and sign concentration property of Lasso and Dantzig estimators," *Electron. J. Stat.*, vol. 2, pp. 90–102, 2008.

[22] T. Zhang, "Some sharp performance bounds for least squares regression with l1 regularization," *Ann. Stat.*, vol. 37, no. 5A, pp. 2109–2144, 2009.

[23] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Birkhäuser, Basel, 2013.

[24] 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.

[25] T. Blumensath and M. E. Davies, "Normalized iterative hard thresholding: Guaranteed stability and performance," *IEEE J. Sel. Topics Signal Proc.*, vol. 4, no. 2, pp. 298–309, 2010.

[26] J. D. Blanchard, J. Tanner, and K. Wei, "CGIHT: conjugate gradient iterative hard thresholding for compressed sensing and matrix completion," *Inf. Inference*, vol. 4, no. 4, pp. 289–327, 2015.

[27] Y. Jiao, B. Jin, and X. Lu, "A primal dual active set algorithm for a class of nonconvex sparsity optimization," preprint, arXiv:1310.1147, 2013.

[28] D. L. Donoho and Y. Tsaig, "Fast solution of ℓ_1-norm minimization problems when the solution may be sparse," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 4789–4812, 2008.

[29] C.-H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *Ann. Stat.*, vol. 38, no. 2, pp. 894–942, 2010.

[30] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *J. Amer. Stat. Assoc.*, vol. 96, no. 456, pp. 1348–1360, 2001.