An Armijo-Type Hard Thresholding Algorithm for Joint Sparse Recovery

LILI PAN AND XUNZHI ZHU
School of Mathematics and Statistics, Shandong University of Technology, Zibo 255000, China
Corresponding author: Xunzhi Zhu (zhxzh1997@126.com)

The work was supported by the National Natural Science Foundation of China (11801325).

ABSTRACT Joint sparse recovery (JSR) in compressed sensing simultaneously recovers sparse signals with a common sparsity structure from their multiple measurement vectors obtained through a common sensing matrix. In this paper, we present an Armijo-type hard thresholding (AHT) algorithm for joint sparse recovery. Under the restricted isometry property (RIP), we show that the AHT can converge to a local minimizer of the optimization problem for JSR. Furthermore, we compute the AHT convergence rate with the above conditions. Numerical experiments show the good performance of the new algorithm for JSR.

INDEX TERMS Joint sparse recovery, Armijo-type hard thresholding, convergence, numerical experiment.

I. INTRODUCTION
Sparse recovery in signal processing [1], [2] addresses the problem of recovering sparse signals from underdetermined linear measurements. That is, one aims to recover a sparse vector \( x \in \mathbb{R}^n \) from the measurement \( y = Ax \in \mathbb{R}^m \) with \( m < n \). A vector \( x \) is sparse if most of its entries are zeros. This kind of problem is referred to as the single measurement vector (SMV) problem, which is a significant problem in the era of big data. Mathematically, it can be solved by the following optimization problem:

\[
\min \frac{1}{2} \|Ax - y\|_2^2 \quad \text{s.t.} \quad \|x\|_0 \leq s, \tag{1}
\]

where \( \|x\|_0 \) counts the number of nonzero entries of \( x \), \( s < n \) and \( \| \cdot \| \) is the \( \ell_2 \) norm of a vector.

In some applications, such as magnetoencephalography (MEG), there are multiple measurement vectors, each corresponding to a different unknown vector, with the property that all unknown vectors share a common support. This kind of sparse recovery problem is called the multiple measurement vector (MMV) problem or joint sparse recovery (JSR) problem. It recovers the source matrix \( X \in \mathbb{R}^{n \times p} \) with joint sparsity from the matrix \( Y \in \mathbb{R}^{m \times p} \) with multiple measurement vectors given by \( Y = AX \). Here, the joint sparsity of matrix \( X \) is called row sparsity, defined as the number of nonzero rows in \( X \). Usually, the JSR problem can be written as

\[
\min \frac{1}{2} \|AX - Y\|_F^2 \quad \text{s.t.} \quad \|X\|_{r, 0} \leq s, \tag{2}
\]

where \( \| \cdot \|_F \) is the Frobenius norm of the matrix, \( \|X\|_{r, 0} \) counts the number of rows of \( X \) containing nonzero entries and \( s < n \). Denote \( X_{(i)} \) as the ith row of \( X \) and \( \Gamma_r(X) = \{ i = 1, \ldots, n : \|X_{(i)}\|_0 \neq 0 \} \) as the row support of \( X \). Then, \( \|X\|_{r, 0} = |\Gamma_r(X)| \) where \( |J| \) is the cardinality of index set \( J \).

In recent years, a substantial body of work has been performed for the JSR problem by relaxation methods [3]–[8]. In addition, many greedy methods for a single measurement vector were extended to the JSR problem. Tropp et al. [9], [10] introduced simultaneous recovery algorithms based on orthogonal matching pursuit (OMP), called simultaneous orthogonal matching pursuit (SOMP) [11]. In [12], Foucart extended the hard thresholding pursuit (HTP) [13] to the joint sparse recovery problem. The authors in [14] provided a comprehensive investigation of the extension to the joint sparse recovery problem of five known greedy algorithms designed for the single measurement vector problem: iterative hard thresholding (IHT) [15], normalized IHT (NIHT) [16], HTP [13], normalized HTP (NHTP) [13], and compressive sampling matching pursuit (CoSaMP) [17]. By considering the rank of the unknown signal matrix \( X \) or the subspace spanned by measurement vectors \( Y \), various joint sparse recovery algorithms have been proposed, such as rank-aware-order recursive matching pursuit (RA-ORMP) [22], subspace-augmented multiple signal classification (SA-MUSIC) [23], and signal space matching pursuit (SSMP) [24].

In this paper, we pay particular attention to the IHT algorithm [15] for problem (1), which was proposed by Blumensath and Davies in [15], since IHT has good recovery

Digital Object Identifier 10.1109/ACCESS.2021.3097216

Received June 1, 2021, accepted July 8, 2021, date of publication July 14, 2021, date of current version July 26, 2021.
properties, and the frame is simple and easy to implement. The convergence of IHT to stationary points was established in [18]. In [16], the authors showed that the numerical studies of IHT are not very promising and the algorithm often fails to converge when the conditions fail. Then, they gave NIHT with an adaptive step size and line search and proved that it converges to a local minimizer if the matrix \( A \) is row full-rank and \( s \)-regular, where \( A \) is \( s \)-regular, which means that any \( s \) columns of \( A \) are linearly independent [18]. Curtis and Thompson [19] showed that NIHT converges to a local minimizer if \( A \) is \( 2s \)-regular.

To guarantee the convergence of the algorithm, some line search rules were proposed. Among these rules, the Armijo-type line search rule [25] is the most popular and easier to implement on computers than other line searches [26]. Inspired by these works, we propose an Armijo-type hard thresholding (AHT) algorithm to solve problem (2).

Since problem (2) is, in general, nonconvex programming, we can only expect to find the local minimum rather than the global minimum. Then, we show that the accumulation point of the AHT algorithm is a stationary point. With the restricted isometry property (RIP) of matrix \( A \), we prove that the stationary point is a local minimum. Moreover, we give the full convergence of AHT and its linear rate of convergence under RIP. Finally, we simulate several experiments to demonstrate the effectiveness of the algorithm.

This paper is organized as follows. Section II presents the AHT algorithm for problem (2) and proves its convergence properties. Numerical results are given in Section III. The last section makes concluding remarks.

II. ALGORITHM AND CONVERGENCE

In this section, we present the Armijo-type hard thresholding (AHT) algorithm for problem (2) and then analyze its convergence properties. Denote \( f(X) := \frac{1}{2}||AX - Y||_F^2 \). Then, the gradient of \( f \) is denoted as \( \nabla f(X) := A^\top(AX - Y) \). Denote the row sparse set

\[
S := \{ X \in \mathbb{R}^{n \times p} : \|X\|_{r,0} \leq s \}.
\]

For matrix \( A \in \mathbb{R}^{m \times n} \), the restricted isometry constants (RIC) of order \( s \), denoted \( L_s \) and \( U_s \), are defined as follows: for all \( x \in \mathbb{R}^n \) with \( \|x\|_0 \leq s \), it holds that

\[
L_s \|x\|_2^2 \leq \|Ax\|_2^2 \leq U_s \|x\|_2^2.
\]

It is easy to verify that if \( A \) has RIP constants \( L_s \) and \( U_s \), then for any \( X \in \mathbb{S} \), we have

\[
L_s \|X\|_F^2 \leq \|AX\|_F^2 \leq U_s \|X\|_F^2.
\]

For \( X \in \mathbb{R}^{n \times p} \) and set \( \Omega \subseteq \mathbb{R}^{n \times p} \), projecting \( X \) onto \( \Omega \) is

\[
\Pi_\Omega(X) \triangleq \arg \inf_{Z \in \Omega} \|X - Z\|_F.
\]

Note that the projection onto sparse set \( S \), written \( \Pi_S(\cdot) \) sets all but the \( s \) largest \( \ell_2 \) norms of the rows in \( X \) to zeros.

The definition of \( \alpha \)-stationarity was proposed in [18] based on the notion of a fixed-point equation. Here, we extend it to problem (2).

\textbf{Definition 1:} The point \( X^* \in S \) is called an \( \alpha \)-stationary point of problem (2) if it holds that

\[
X^* \in \Pi_S \left( X^* - \alpha A^\top(AX^* - Y) \right).
\]

Note that \( X^* \in S \) is an \( \alpha \)-stationary point of problem (2) if and only if [18]

\[
\|A^\top(AX^* - Y)\|_{\Gamma_r} \begin{cases}
0, & i \in \Gamma_r^s, \\
\leq \frac{1}{\alpha} \mathcal{M}_s(X^*), & i \notin \Gamma_r^s,
\end{cases}
\]

where \( \mathcal{M}_i(X) \) denotes the \( i \)th largest \( \ell_2 \) norms of the rows in \( X \) and \( \Gamma_r^s := \Gamma_r(X^*) \).

A. ALGORITHM

In NIHT for CS [16], to guarantee the objective function has a sufficient decrease per iteration, the authors added a step size that maximally decrease the objective function directly with the RIP of matrix \( A \). Here \( (\cdot)_r \) is the subvector (submatrix) obtained by discarding all but the elements (columns) in \( \Gamma_r \). The AHT framework is described as follows.

\textbf{Algorithm 1 Armijo-Type Hard Thresholding (AHT)}

\begin{itemize}
    \item [Step 1.] Initialize \( X^0, \sigma > 0, 0 < \beta < 1 \).
    \item [Step 2.] Let \( \Gamma_r^k = \Gamma_r(X^k) \) and compute
    \[
    \alpha_k^0 = \frac{\|A^\top(AX^k - Y)\|_{\Gamma_r^k}}{\|A^\top(AX^k - Y)\|_{\Gamma_r^k}}.
    \]
    \item [Step 3.] Compute
    \[
    X^{k+1} = \Pi_S(X^k - \alpha_k A^\top(AX^k - Y)),
    \]
    where \( \alpha_k = \alpha_k^0 \beta^{m_k} \) and \( m_k \) is the smallest positive integer \( m \) such that
    \[
    \|AX^k(\alpha_k^0 \beta^m - Y\|_F^2 \\
    \leq \|AX^k - Y\|_F^2 - \sigma \|X^k(\alpha_k^0 \beta^m - X^k\|_F^2,
    \]
    and \( X^k(\sigma) = \Pi_S(X^k - \alpha A^\top(AX^k - Y)) \).
    \item [Step 4.] If the stopping criterion is met, then stop.
\end{itemize}

Remark 1: We now briefly illustrate the algorithm.

- In Step 2, we utilize the optimal stepsize to compute an initial stepsize of \( X^{k+1} \) or a step size that maximally decreases the error in each iteration, which can be obtained by computing the following problem:

\[
\arg\min_{\alpha > 0} \{ \|A^\top(AX^k(\alpha A^\top(AX^k - Y))\|_{\Gamma_r^k} \}
\]
This step size is in accordance with the optimal step size in NIHT for CS in [16]. It should be noted that the maximal reduction in the objective function can only be guaranteed when the row supports of \( X^k \) and \( X^{k+1} \) coincide. When the row support of \( X^{k+1} \) differs from that of \( X^k \), we use an Armijo-type stepsize to sufficiently decrease the objective function in Step 3. The strategy of Armijo-type stepsize is as follows: we first calculate \( X^{k+1} \) with a larger stepsize and determine whether the objective function has a sufficient decrease, that is, whether (7) holds. If (7) holds, we keep this stepsize. Otherwise, we decrease the stepsize. Lemma 1 shows the stepsize exists. Combining (4) and (8), we can see that the optimal stepsize \( \alpha_k^0 \leq \frac{1}{U_{2s} + \sigma} \). From the following Lemma 1, the total number of decreasing step sizes to make \( 0 < \alpha = \alpha_k^0 (1 - \sigma) \leq \frac{1}{U_{2s} + \sigma} \) is at most \( \lceil \ln \frac{L_1}{\sigma U_{2s} + \sigma} / \ln \beta \rceil + 1 \).

- We adopt the same stopping criterion as that in [27], that is

\[
(A^T (AX^k - Y))_{(i)} = 0, \quad i \in \Gamma^+_r
\]

or the number of iterations exceeds the max iterations.

The following result shows that the Armijo stepsize is well defined under RIP.

**Lemma 1:** Assume matrix \( A \) satisfies RIP with order \( 2s \) and \( X^k \) is the iterative points generated by AHT. Then, for \( 0 < \alpha \leq \frac{1}{U_{2s} + \sigma} \), it holds

\[
\|AX^k(\alpha) - Y\|_F^2 \leq \|AX^k - Y\|_F^2 - \sigma \|X^k(\alpha) - X^k\|_F^2.
\]

Therefore, \( \alpha_k \) is well defined.

**Proof:** According to the computation in Step 3, we have

\[
X^k(\alpha) \in \arg\min \left\{ \|X - X^k + \alpha A^T (AX^k - Y)\|_F^2, \quad X \in S \right\}.
\]

It implies that

\[
\|X^k(\alpha) - X^k + \alpha A^T (AX^k - Y)\|_F^2 \leq \|\alpha A^T (AX^k - Y)\|_F^2,
\]

that is

\[
\|X^k(\alpha) - X^k\|_F^2 \leq -2\alpha \|A^T (AX^k - Y)\|_F^2 - \sigma \|X^k(\alpha) - X^k\|_F^2.
\]

Then we obtain

\[
\begin{align*}
\|AX^k(\alpha) - Y\|_F^2 & = \|AX^k - Y + \alpha (X^k(\alpha) - X^k)\|_F^2 \\
& \leq \|AX^k - Y\|_F^2 + 2\|AX^k - Y, A(X(\alpha) - X^k)\|_F^2 + U_{2s} \|X^k(\alpha) - X^k\|_F^2 \\
& \leq \|AX^k - Y\|_F^2 + 2(\|AX^k - Y\|_F^2) + U_{2s} \|X^k(\alpha) - X^k\|_F^2 \\
& \leq \|AX^k - Y\|_F^2 + 2\|AX^k(\alpha) - X^k\|_F^2 + U_{2s} \|X^k(\alpha) - X^k\|_F^2 \\
& \leq \|AX^k - Y\|_F^2 - \frac{1}{\alpha} \|X^k(\alpha) - X^k\|_F^2 \\
& + U_{2s} \|X^k(\alpha) - X^k\|_F^2 \\
& = \|AX^k - Y\|_F^2 - \frac{1}{\alpha} \|AX^k(\alpha) - X^k\|_F^2,
\end{align*}
\]

where the second inequality comes from (9). By letting \( \frac{1}{\sigma} - U_{2s} \geq \sigma \), we can obtain the desired result by the definition of \( \alpha \). □

**B. ANALYSIS OF CONVERGENCE**

In this subsection, we establish the convergence of AHT using RIP of \( A \) and Lemma 1.

**Theorem 1:** Let the RIP of matrix \( A \) hold and \( \{X^k\} \) be generated by AHT. Then,

\[
(i) \lim_{k \to \infty} \frac{|X^{k+1} - X^k|_F}{\|A^T (AX^k - Y)\|_F} = 0;
\]

\[
(ii) \text{any accumulation point of } \{X^k\} \text{ is an } \alpha \text{-stationary point of } (2).
\]

**Proof:** (i) It follows from Lemma 1 that

\[
\|AX^k - Y\|_F^2 - \|AX^{k+1} - Y\|_F^2 \geq \sigma \|X^{k+1} - X^k\|_F^2.
\]

Then

\[
\begin{align*}
\sum_{k=0}^{\infty} \|X^{k+1} - X^k\|_F^2 & \leq \frac{1}{\sigma} \sum_{k=0}^{\infty} (\|AX^k - Y\|_F^2 - \|AX^{k+1} - Y\|_F^2) \\
& < \frac{1}{\sigma} \|AX^0 - Y\|_F^2,
\end{align*}
\]

which shows \( \lim_{k \to \infty} \|X^{k+1} - X^k\|_F = 0 \).

It follows from (4) that

\[
\frac{\|A^T (AX^k - Y)\|_F^2}{\|A^T (AX^k - Y)\|_F^2} \geq \frac{1}{U_{2s}},
\]

then \( \alpha_k^0 \geq \frac{1}{U_{2s}} \). By Lemma 1 and the definition of \( \alpha_k \) in the algorithm, we have

\[
\alpha_k \geq \min \{ \alpha_k^0, \frac{\beta}{U_{2s} + \sigma} \}.
\]

Therefore, \( \alpha_k \) is bounded from below by a positive constant. We conclude that

\[
\lim_{k \to \infty} \frac{|X^{k+1} - X^k|_F}{\alpha_k} = 0.
\]

(ii) Assume that \( X^* \) is an accumulation point of sequence \( \{X^k\} \); then, there exists a subsequence \( \{X^{k_j}\} \) converging to \( X^* \) and \( \lim X^{k_j+1} = X^* \) by (i). For the simplification of notation, denote \( \alpha_k \) as \( \alpha \) by its boundedness and \( \Gamma^*_r = \Gamma_r(X^*) \). Based on

\[
X^{k_j+1} = \Pi_S (X^{k_j} - \alpha A^T (AX^{k_j} - Y))
\]

in Step 2, we consider two cases.

**Case 1:** \( i \in \Gamma^*_r \). The convergence of \( \{X^{k_j}\} \) and \( \{X^{k_j+1}\} \) guarantees that for some \( n_1 > 0 \), \( \|X^{k_j}\|_F > 0 \), \( \|X^{k_j+1}\|_F > 0 \) for all \( j > n_1 \). The definition of the projection on \( S \) shows that

\[
X^{k_j+1} = X^{k_j} - \alpha A^T (AX^{k_j} - Y).
\]

Taking \( j \to \infty \), we have \( A^T (AX^* - Y) = 0 \).
Case 2: $i \notin \Gamma^r_s$. If there exists an $n_2 > 0$ such that for all $j > n_2$, $\|X_{k_j+1}^j\|_0 = 0$, the projection implies that

$$\|X_{k_j}^j - \alpha(A^\top(AX^* - Y))_0\| \leq \mathcal{M}_s(AX^*)$$

Letting $j \to \infty$ and using the continuity of the function $\mathcal{M}_s$, we obtain that

$$\|\alpha(A^\top(AX^* - Y))_0\| \leq \mathcal{M}_s(AX^*)$$

If there exists an infinite number of indices of $k_j$ for $\|X_{k_j+1}^j\|_0 > 0$, as the same proof in Case 1, it follows that

$$(A^\top(AX^* - Y))_0 = 0.$$ 

Since $\alpha_k$ is bounded from below by a positive constant, we have

$$\|A^\top(AX^* - Y)_0\| \begin{cases} = 0, & \text{if } i \in \Gamma^r_s, \\ \leq \frac{1}{\alpha} \mathcal{M}_s(AX^*), & \text{if } i \notin \Gamma^r_s, \end{cases}$$

which means $X^*$ is an $\alpha$-stationary point of (2). \qed

Theorem 2: Let the RIP of matrix $A$ hold and $\{X^k\}$ be generated by AHT. Then, the accumulation point of $\{X^k\}$ is a local minimizer of (2).

Proof: From Theorem 1, we know that the accumulation point of $\{X^k\}$ is an $\alpha$-stationary point of (2). Next, we only need to show that the $\alpha$-stationary point $X^*$ of (2) is a local minimizer of (2). We first prove the fact that $A^\top(AX^* - Y), X - X^* = 0$ by separating it into two cases.

Case 1: $\|X^\top\|_0 < s$. Then, $\mathcal{M}_s(AX^*) = 0$, which indicates that $A^\top(AX^* - Y) = 0$. It holds that $A^\top(AX^* - Y), X - X^* = 0$.

Case 2: $\|X^\top\|_0 \geq s$. For any $X$ sufficiently close to $X^*$, we have $\Gamma_r(X) \supseteq \Gamma^r_s$. Then, $\|X\|_\infty \geq \|X^\top\|_\infty = s$. Together with $\|X\|_\infty \leq s$, we obtain $\Gamma_r(X) = \Gamma^r_s$ and then $\|X\|_\infty = s$. Therefore,

$$\langle A^\top(AX^* - Y), X - X^* \rangle = \langle (A^\top(AX^* - Y))_{\Gamma^r_s}, (X - X^*)_{\Gamma^r_s} \rangle + \langle (A^\top(AX^* - Y))_{\Gamma^r_s^c}, (X - X^*)_{\Gamma^r_s^c} \rangle = 0,$$

where the last equality comes from (6), $\Gamma_r(X^*) = \Gamma^r_s$ and $\Gamma^r_s^c$ is the complementary set of $\Gamma^r_s$.

Since the objective function of problem (2) is convex, we have

$$\frac{1}{2} \|AX - Y\|^2 \geq \frac{1}{2} \|AX^* - Y\|^2 + \langle A^\top(AX^* - Y), X - X^* \rangle$$

$$= \frac{1}{2} \|AX^* - Y\|^2 + 0$$

for any $X \in S$ and is sufficiently close to $X^*$. This completes the proof. \qed

C. Rate of Convergence

To show the rate of convergence, we need the next lemma.

Lemma 2: Let the RIP of order $s$ hold. We have for some $\alpha > 0$ and $|\Gamma| \leq s$, $\Gamma_r(X) \subseteq \Gamma$

$$\|(I - \alpha A^\top A)X\|^2_2 \leq \Delta_\alpha(\alpha)\|X\|^2_F,$$

where

$$\Delta_\alpha(\alpha) := \max\{|\alpha U_s - 1|, |1 - \alpha L_s|\}.$$ 

Proof: Notice that the RIP (4) is equivalent to

$$(\alpha L_s - 1)\|X\|^2_2 \leq \alpha\|AX\|^2_F - \|X\|^2_F \leq (\alpha U_s - 1)\|X\|^2_2$$

for $\Gamma_r(X) \subseteq \Gamma$. Then, we have

$$\|\alpha A^\top AX - X\|^2_F \leq \max\{|\alpha U_s - 1|, |1 - \alpha L_s|\}\|X\|^2_2,$$

for all $|\Gamma| \leq s$ and $X$ with $\Gamma_r(X) \subseteq \Gamma$. Then, it follows that

$$\alpha\|A^\top AX - X\|^2_F \leq \max\{|\alpha U_s - 1|, |1 - \alpha L_s|\}\|X\|^2_2.$$

This means that

$$\max_{0 \neq X, \Gamma_r(X) \subseteq \Gamma} \frac{\langle (\alpha A^\top A - I)X, X \rangle}{\|X\|^2_2} = \max_{0 \neq X, \Gamma_r(X) \subseteq \Gamma} \frac{\alpha\|A^\top AX - X\|^2_F - \|X\|^2_F}{\|X\|^2_2} \leq \max\{|\alpha U_s - 1|, |1 - \alpha L_s|\}.$$ 

Then, we have

$$\max_{|\Gamma| \leq s} \|\alpha A^\top A - I\|_2 \leq \max\{|\alpha U_s - 1|, |1 - \alpha L_s|\}.$$ 

This completes the proof. \qed

The next theorem shows the linear convergence rate of AHT. Generally, for the first-order optimization method, we can only expect to obtain the linear convergence rate. The sufficient condition is in accordance with those for the methods based on HHT. For completeness, we give the proof.

Theorem 3: Assume that the RIP (4) of matrix $A$ with order $3s$ and $\Delta_3(\alpha) < 1/2$ (defined in (12)). If $X^*$ is an $s$-row sparse matrix, then the sequence $\{X^k\}$ generated by AHT converges to the matrix $X^*$ with $Y = AX^*$, i.e.,

$$\|X^{k+1} - X^*\|_F \leq \rho^{k+1}\|X^0 - X^*\|_F,$$

where $0 < \rho < 1$.

Proof: We use the fact that the $s$-row sparse matrix $X^{k+1}$ is a better $s$-term approximation to $X^k - \alpha A^\top (AX^k - Y)$ than $X^k$. This yields that

$$\|X^{k+1} - X^k + \alpha A^\top (AX^k - Y)\|^2_2 \leq \|X^k - X^k + \alpha A^\top (AX^k - Y)\|^2_2.$$ 

Inserting the term of $X^*$, it follows that

$$\|X^{k+1} - X^* + X^* - X^k + \alpha A^\top (AX^k - Y)\|^2_2 \leq \|X^* - X^k + \alpha A^\top (AX^k - Y)\|^2_2.$$
Expanding the left-hand side and eliminating $\|X^* - X^k + \alpha A^T (AX^k - Y)\|^2_F$ leads to

$$\|X^{k+1} - X^*\|^2_F \leq -2\langle X^* - X^k + \alpha A^T (AX^k - Y), X^{k+1} - X^* \rangle$$

$$= -2\langle X^* - X^k + \alpha A^T (AX^k - AX^*), X^{k+1} - X^* \rangle$$

$$= 2\langle X^* - X^k + \alpha A^T (AX^k - AX^*), X^{k+1} - X^* \rangle$$

$$= 2\langle (I - \alpha A^T)(X^k - X^*), X^{k+1} - X^* \rangle$$

$$\leq 2\Delta_3(\alpha)\|X^k - X^*\|_F\|X^{k+1} - X^*\|_F.$$

Simplifying by $\|X^{k+1} - X\|_F$, we derive

$$\|X^{k+1} - X\|_F \leq 2\Delta_3(\alpha)\|X^k - X\|_F.$$

Then

$$\|X^{k+1} - X\|_F \leq \rho^k\|X^1 - X\|_F.$$

Therefore, the sequence $\{X^k\}$ converges to $X$ whenever $\rho < 1$, i.e., $\Delta_3(\alpha) < \frac{1}{2}$. □

### III. NUMERICAL EXPERIMENTS

In this section, we conduct numerous experiments to illustrate the performance of AHT. The codes were written in MATLAB, and the experiments were performed in MATLAB R2019a on a desktop computer with 3.40 GHz and 16 GB RAM. For all the simulations, the data are generated as follows. More specifically, two types of sensing matrices of $A$ are generated, namely, random Gaussian matrix and random partial discrete cosine transform (pDCT) matrix:

- Gaussian: $A_{ij} \sim \mathcal{N}(0, I/m)$,
- $j = 1, 2, \ldots, n$,

- pDCT: $A_{ij} = m^{-1/2} \cos(2\pi(j - 1)i)$,
- $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$,

where $A_{ij}$ denotes the $i$th column of $A$ and $I$ is an $m$-dimensional identity matrix. After generating the matrices, we orthogonalize them to satisfy $AA^T = I$. We generate the ‘true’ original signal $X_{\text{true}}$ as follows: first produce an index set $T$ with $s$ indices randomly selected from $[1, \ldots, n]$; then, for each row of $X_{\text{true}}$ with index in $T$, uniformly choose from $[0, 10]$. The data are generated as follows (in MATLAB format):

$$X_{\text{true}} = \text{zeros}(n, p); T = \text{randperm}(n);$$

$$X_{\text{true}}(T(1:s), :) = 10 * \text{rand}(s, p);$$

$$Y = A * X_{\text{true}} + \sigma_0 * \text{randn}(m, p). \quad (13)$$

The case $\sigma_0 = 0$ is the exact recovery, and $\sigma_0 = 0.01$ is the stable recovery.

### A. COMPARISON OF DIFFERENT METHODS

We compare our algorithm with six state-of-the-art greedy methods: normalized iterative hard thresholding (NIHT) [16], compressive sampling matching pursuit (CoSaMP) [17], subspace pursuit (SP) [20], signal space matching pursuit (SSMP) [24], order recursive matching pursuit (ORMP) [21], and compressive multiple signal classification (MUSIC) [28].

To see the accuracy of the solutions and the speed of these seven methods, we now run 100 trials for each kind of matrix with $n = 512$, $m = 128$ and $p = 4$. Specific results produced by these seven methods are recorded, where the exact reconstruction ratio (ERR) defined as [20]

$$\text{ERR} := \frac{\text{number of exact reconstructions}}{\text{number of total trials}},$$

and the relative error (RE) is computed by

$$\text{RE} := \frac{\|X - X_{\text{true}}\|_F}{\|X\|_F}.$$

The CPU time (in seconds) reported here does not include the time for data initialization.

For Gaussian matrix with $\sigma_0 = 0$, when sparsity is no more than 25 (or $s/n \leq 4.9\%$), all the seven methods can make exact joint sparse recovery. With the increase of the sparsity, the quality of recovery of NIHT, ORMP, MUSIC and SP
decreases. And our AHT, CoSaMP and SSMP can recover signal accurately and their relative errors are less than those of other three methods. For $\sigma_0 = 0.01$, our AHT and SSMP can make exact recovery and also behave better than other five methods in regards of relative error. While for the CPU time, our AHT runs faster than other six methods.
For pDCT matrix with $\sigma_0 = 0$, when sparsity is no more than 35 (or $s/n \leq 6.8\%$), all the seven methods can make exact joint sparse recovery. For $\sigma_0 = 0.01$, our AHT, SSMP and ORMP can make exact joint sparse recovery and their errors are relative small. Our AHT also has advantage on the aspect of the CPU time.

In conclusion, for exact recovery ($\sigma = 0$) with the two kinds of matrices, almost all the methods can obtain exact reconstruction. For stable recovery ($\sigma = 0.01$), our AHT and SSMP behave better than other five methods. Significantly, our AHT possesses an apparent advantage in terms of CPU time.

IV. CONCLUSION

In this paper, we proposed an Armijo-type hard thresholding (AHT) algorithm to solve the joint sparsity recovery problem. We showed the sequence generated by AHT converges to a local minimizer of the optimization for JSR and we also gave the AHT convergence rate. Numerous experiments illustrated the efficiency of the AHT algorithm on JSR.

REFERENCES

[1] E. J. Candes and T. Tao, “Decoding by linear programming,” IEEE Trans. Inf. Theory, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
[2] D. L. Donoho, “Compressed sensing,” IEEE Trans. Inf. Theory, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[3] A. Argyriou, T. Evgeniou, and M. Pontil, “Multi-task feature learning,” in Proc. Adv. Neural Inf. Process. Syst., 2007, pp. 41–48.
[4] N. H. Nguyen, N. M. Nasrabadi, and T. D. Tran, “Robust multi-sensor classification via joint sparse representation,” in Proc. 14th Int. Conf. Inf. Fusion (FUSION), Jul. 2011, pp. 1–8.
[5] G. Obozinski, B. Taskar, and M. I. Jordan, “Joint covariate selection and joint subspace selection for multiple classification problems,” Statist. Comput., vol. 20, no. 2, pp. 231–252, Apr. 2010.
[6] G. Obozinski, M. J. Wainwright, and M. I. Jordan, “Support union recovery in high-dimensional multivariate regression,” Ann. Statist., vol. 39, no. 1, pp. 1–4, Feb. 2011.
[7] X.-T. Yuan, X. Liu, and S. Yan, “Visual classification with multitask joint sparse representation,” IEEE Trans. Image Process., vol. 21, no. 10, pp. 4349–4360, Oct. 2012.
[8] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, “Robust visual tracking via structured multi-task sparse learning,” Int. J. Comput. Vis., vol. 101, no. 2, pp. 367–383, Jan. 2013.
[9] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, “Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit,” Signal Process., vol. 86, no. 3, pp. 572–588, Mar. 2006.
[10] J. A. Tropp, “Algorithms for simultaneous sparse approximation. Part II: Convex relaxation,” Signal Process., vol. 86, no. 3, pp. 589–602, Mar. 2006.
[11] G. Davis, “Adaptive greedy approximations,” Constructive Approximation, vol. 13, no. 1, pp. 57–98, Mar. 1997.
[12] S. Foucart, “Recovering jointly sparse vectors via hard thresholding pursuit,” in Proc. Sampling Theory Appl., May 2011, pp. 1–4.
[13] S. Foucart, “Hard thresholding pursuit: An algorithm for compressive sensing,” SIAM J. Numer. Anal., vol. 49, no. 6, pp. 2543–2563, Jan. 2011.
[14] J. D. Blanchard, M. Cermak, D. Hanle, and Y. Jing, “Greedy algorithms for joint sparse recovery,” IEEE Trans. Signal Process., vol. 62, no. 7, pp. 1694–1704, Apr. 2014.
[15] T. Blumensath and M. E. Davies, “Iterative thresholding for sparse approximations,” J. Fourier Anal. Appl., vol. 14, nos. 5–6, pp. 629–654, Dec. 2008.
[16] T. Blumensath and M. E. Davies, “Normalized iterative hard thresholding: Guaranteed stability and performance,” IEEE J. Sel. Topics Signal Process., vol. 4, no. 2, pp. 298–309, Apr. 2010.
[17] 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, May 2009.
[18] A. Beck and Y. C. Eldar, “Sparsity constrained nonlinear optimization: Optimality conditions and algorithms,” SIAM J. Optim., vol. 23, no. 3, pp. 1480–1509, Jan. 2013.
[19] C. Curtis and A. Thompson, “A new and improved quantitative recovery analysis for iterative hard thresholding algorithms in compressed sensing,” IEEE Trans. Inf. Theory, vol. 61, no. 4, pp. 2019–2042, Apr. 2015.
[20] 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.
[21] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, “Sparse solutions to linear inverse problems with multiple measurement vectors,” IEEE Trans. Signal Process., vol. 53, no. 7, pp. 2477–2488, Jul. 2005.
[22] M. E. Davies and Y. C. Eldar, “Rank awareness in joint sparse recovery,” IEEE Trans. Inf. Theory, vol. 58, no. 2, pp. 1135–1146, Feb. 2012.
[23] K. Lee, Y. Bresler, and M. Junge, “Subspace methods for joint sparse recovery,” IEEE Trans. Inf. Theory, vol. 58, no. 6, pp. 3613–3641, Jun. 2012.
[24] J. Kim, J. Wang, L. T. Nguyen, and B. Shim, “Joint sparse recovery using signal space matching pursuit,” IEEE Trans. Inf. Theory, vol. 66, no. 8, pp. 5072–5096, Aug. 2020.

LILI PAN received the Ph.D. degree in operational research and cybernetics from Beijing Jiaotong University, China, in 2017. She is currently an Associate Professor in the School of Mathematics and Statistics, Shandong University of Technology, China. Her research interests include sparse optimization theory and algorithm, and sparse image recovery.

XUNZHI ZHU received the M.S. degree in operational research and cybernetics from the Dalian University of Technology, China, in 2005. He currently works with the School of Mathematics and Statistics, Shandong University of Technology, China. His research interest includes optimization algorithm.