RECOVERING JOINTLY SPARSE VECTORS VIA HARD THRESHOLDING PURSUIT

Simon Foucart[1]

[1]Department of Mathematics
Drexel University, Philadelphia
Email: foucart@math.drexel.edu

ABSTRACT

We introduce an iterative algorithm designed to find row-sparse matrices $X \in \mathbb{R}^{N \times K}$ solution of an underdetermined linear system $AZ = Y$, where $A \in \mathbb{R}^{m \times N}$ and $Y \in \mathbb{R}^{m \times K}$ are given. In the case $K = 1$, the algorithm is a simple combination of popular Compressive Sensing algorithms, which we had previously coined Hard Thresholding Pursuit. After recalling the main results concerning this algorithm, we generalize them to the case $K \geq 1$ for the new Simultaneous Hard Thresholding Pursuit algorithm. In particular, we prove that any s-row-sparse matrix can be exactly recovered using a finite number of iterations of the algorithm provided that the $3s$th Restricted Isometry Constant of the matrix A satisfies $\delta_{3s} < 1/\sqrt{3}$. We also discuss the cost of recovering matrices at once via Simultaneous Hard Thresholding Pursuit versus recovering their columns one by one via Hard Thresholding Pursuit.

Keywords— compressive sensing, sparse recovery, iterative algorithms, thresholding, joint sparsity

1. INTRODUCTION

A fundamental problem in Compressive Sensing is to find linear measurement schemes allowing for the representation of sparse vectors using an information level close to the sparsity level rather than the full dimension of the vectors. Such measurement schemes exist, as shown by probabilistic methods. Setting this problem aside, the other fundamental problem is to find efficient reconstruction schemes allowing for the recovery of sparse vectors from their measurement vectors. Mathematically speaking, given a measurement matrix $A \in \mathbb{R}^{m \times N}$ with $m \ll N$, and given a measurement vector $\mathbf{y} = A\mathbf{x} \in \mathbb{R}^m$ associated with an s-sparse vector $\mathbf{x} \in \mathbb{R}^N$ (a vector that has at most s nonzero entries), we wish to access this vector in a numerically tractable way. A very popular strategy — the ℓ_1-minimization — consists in solving the convex optimization program (which can be recast as a linear optimization program)

$$\underset{\mathbf{z} \in \mathbb{R}^N}{\text{minimize}} \ \|\mathbf{z}\|_1 \quad \text{subject to } A\mathbf{z} = \mathbf{y}. \tag{1}$$

There are other popular strategies, for instance the Subspace Pursuit (SP) [2], the Compressive Sampling Matching Pursuit (CoSaMP) [9], and the Iterative Hard Thresholding (IHT) [1] algorithms, to name a few. In [3], we have introduced a simple combination of the latter algorithms, which we called the Hard Thresholding Pursuit (HTP) algorithm. It may be viewed as the Iterative Hard Thresholding algorithm where each iteration is augmented with a debiasing step, shown to improve performance in other algorithms as well. It reads:

Start with an s-sparse $\mathbf{x}^0 \in \mathbb{C}^N$, say $\mathbf{x}^0 = 0$, and iterate the steps

$$S^{[n+1]} = \big\{\text{indices of } s \text{ largest } \big|\big(\mathbf{x}^{[n]} + A^\top(\mathbf{y} - A\mathbf{x}^{[n]})\big)_i\big|\big\}, \tag{HTP$_1$}$$

$$\mathbf{x}^{[n+1]} = \operatorname{argmin}\big\{\|\mathbf{y} - A\mathbf{z}\|_2, \operatorname{supp}(\mathbf{z}) \subseteq S^{[n+1]}\big\}, \tag{HTP$_2$}$$

until a stopping criterion is met. This criterion is chosen to be $S^{[n+1]} = S^{[n]}$, since no new outputs are created afterwards. We became aware that the algorithm has (unsurprisingly) been suggested in other places [5, 8, 6], but [3] presents in particular a simple analysis in terms of Restricted Isometry Constants that was not available before. We recall that the sth Restricted Isometry Constant $\delta_s = \delta_s(A)$ of the matrix A is defined as the smallest quantity $\delta \geq 0$ such that

$$(1-\delta)\|\mathbf{z}\|_2^2 \leq \|A\mathbf{z}\|_2^2 \leq (1+\delta)\|\mathbf{z}\|_2^2, \quad \text{all } s\text{-sparse } \mathbf{z} \in \mathbb{R}^N.$$

This turns out to be a fruitful concept, since it allows to establish s-sparse recovery for a wide range of algorithms using the optimal number $m \asymp s\ln(N/s)$ of linear measurements. Typically, a condition of the type $\delta_t < \delta_*$ for some specified δ_* and some t related to s guarantees that all s-sparse vectors $\mathbf{x} \in \mathbb{R}^N$ are recoverable from $\mathbf{y} = A\mathbf{x} \in \mathbb{R}^m$ via a given algorithm. For instance, we mention the conditions $\delta_{2s} < 0.4652$ for ℓ_1-minimization, $\delta_{3s} < 0.5$ for IHT, and $\delta_{4s} < 0.3843$ for CoSaMP, see [4], while keeping in mind that these are subject to improvements. Since the conditions $\delta_t < \delta_*$ are only known to be satisfied for random matrices for which $m \asymp (t/\delta_*^2)\ln(N/t)$, they can heuristically be assessed by the smallness of the ratio t/δ_*^2. As such, the upcoming sufficient condition (2) for HTP is preferable to the ones cited above. It has to be noted that this condition is also sufficient for the IHT algorithm, and in fact for other fast variants of the HTP algorithm where one performs only a few gradient descent steps instead of solving the linear system coming out of (HTP$_2$), see [3]. It also has to be noted that the HTP algorithm, like all algorithms mentioned here, is stable with respect to sparsity defect and robust with respect to measurement error. Indeed, the main theorem of [3] reads:

Theorem 1. *Suppose that the $3s$th Restricted Isometry Constant of the matrix $A \in \mathbb{R}^{m \times N}$ satisfies*

$$\delta_{3s} < \frac{1}{\sqrt{3}} \approx 0.57735. \tag{2}$$

Then, for every $\mathbf{x} \in \mathbb{R}^N$ and every $\mathbf{e} \in \mathbb{R}^m$, if S is an index set of s largest entries of $\mathbf{x}$ in absolute value, the sequence $(\mathbf{x}^{[n]})$ defined by (HTP) with $\mathbf{y} = A\mathbf{x} + \mathbf{e}$ satisfies, for all $n \geq 0$,

$$\|\mathbf{x}^{[n]} - \mathbf{x}_S\|_2 \leq \rho^n \|\mathbf{x}^{[0]} - \mathbf{x}_S\|_2 + \tau \, \frac{1-\rho^n}{1-\rho} \, \|A\mathbf{x}_{\overline{S}} + \mathbf{e}\|_2,$$

where $\tau \leq 5.15$ and

$$\rho := \sqrt{\frac{2\delta_{3s}^2}{1-\delta_{2s}^2}} < 1.$$

To retrieve the ideal case discussed in this paper, we just set $\mathbf{x}_{\overline{S}} = 0$ and $\mathbf{e} = 0$. In this case, we have shown in [3] that the original vector $\mathbf{x} \in \mathbb{R}^N$ is exactly recovered after at most

$$n_\rho(\mathbf{x}) := \left\lceil \frac{\ln\left(\sqrt{2/3}\,\|\mathbf{x}\|_2/\xi\right)}{\ln\left(1/\rho\right)} \right\rceil + 1 \tag{3}$$

iterations of the algorithm started with $\mathbf{x}^{[0]} = 0$. Here ξ denotes the smallest nonzero absolute value of entries of $\mathbf{x}$. Although the convergence was not established in the general case, we pointed out that the sequence $(\mathbf{x}^{[n]})$ is eventually periodic, and we could estimate the distance between $\mathbf{x}$ and any cluster point of the sequence $(\mathbf{x}^{[n]})$. The following result was derived in [3] using Theorem 1 to show that the HTP algorithm yields error estimates similar to the ones available for ℓ_1-minimization.

Corollary 1. *Suppose that the matrix $A \in \mathbb{R}^{m \times N}$ satisfies $\delta_{6s} < 1/\sqrt{3}$. Then, for any $\mathbf{x} \in \mathbb{R}^N$ and any $\mathbf{e} \in \mathbb{R}^m$, each cluster point $\mathbf{x}^\star$ of the sequence $(\mathbf{x}^{[n]})$ defined by (HTP) with s replaced by $2s$ and with $\mathbf{y} = A\mathbf{x}+\mathbf{e}$ satisfies, for all $1 \leq p \leq 2$,*

$$\|\mathbf{x} - \mathbf{x}^\star\|_p \leq \frac{C}{s^{1-1/p}} \sigma_s(\mathbf{x})_1 + D s^{1/2-1/p} \|\mathbf{e}\|_2,$$

where the constants C and D depend only on δ_{6s}.

2. ROW-SPARSE RECOVERY

Let us suppose that several sparse vectors $\mathbf{x}_1, \ldots, \mathbf{x}_K \in \mathbb{R}^N$ are to be recovered from $\mathbf{y}_1 = A\mathbf{x}_1, \ldots, \mathbf{y}_K = A\mathbf{x}_K \in \mathbb{R}^m$, with the additional assumption that $\mathbf{x}_1, \ldots, \mathbf{x}_K$ are jointly sparse, in other words that $\mathbf{x}_1, \ldots, \mathbf{x}_K$ are all supported on a set of small cardinality. Note that complex vectors $\mathbf{x} \in \mathbb{C}^N$ measured with a matrix $A \in \mathbb{R}^{m \times N}$ constitute a typical example for $K = 2$. Intuitively, we anticipate a gain in computational complexity by recovering $\mathbf{x}_1, \ldots, \mathbf{x}_K$ all at the same time rather than one by one. This fact is discussed after Theorem 2. For now, we reformulate the problem by defining the $N \times K$ matrix

$$X = \left[\mathbf{x}_1 \,|\, \mathbf{x}_2 \,| \cdots |\, \mathbf{x}_K\right].$$

The joint sparsity assumption just says that X is s-row-sparse, i.e., that its row-support

$$\mathrm{supp}(X) := \left\{ i : X_{(i)} \neq 0 \right\}$$

has cardinality at most s. Here $X_{(i)}$ stands for the ith row fo X, and in general the notations $M_{(1)}, \ldots, M_{(k)}$ and $M_1, \ldots, M_\ell$ will be used to represent the rows and columns, respectively, of a matrix $M \in \mathbb{R}^{k \times \ell}$, so that

$$M = \begin{bmatrix} M_{(1)} \\ \hline \vdots \\ \hline M_{(k)} \end{bmatrix} = \left[M_1 \,|\cdots|\, M_\ell \right].$$

The Frobenius norm of M will also be used, that is

$$\|M\|_F^2 = \sum_{i,j=1}^{k,\ell} m_{i,j}^2 = \sum_{i=1}^{k} \|M_{(i)}\|_2^2 = \sum_{j=1}^{\ell} \|M_j\|_2^2.$$

Note that this norm is derived from the Frobenius inner product

$$\langle M, M' \rangle_F = \sum_{i,j=1}^{k,\ell} m_{i,j} m'_{i,j} = \sum_{i=1}^{k} \langle M_{(i)}, M'_{(i)} \rangle = \sum_{j=1}^{\ell} \langle M_j, M'_j \rangle.$$

With these notations set up, we first recall that solving the convex optimization problem

$$\underset{Z \in \mathbb{R}^{N \times K}}{\text{minimize}} \sum_{i=1}^{N} \|Z_{(i)}\|_2 \quad \text{subject to } AZ = Y \tag{4}$$

allows to recover the s-row-sparse matrix $X \in \mathbb{R}^{N \times K}$ from $Y := AX \in \mathbb{R}^{m \times K}$ as soon as the measurement matrix A has small Restricted Isometry Constants. It is indeed known that successfully recovering all s-row-sparse $X \in \mathbb{R}^{N \times K}$ by solving (4) is exactly equivalent to successfully recovering all s-sparse $\mathbf{x} \in \mathbb{R}^N$ by solving (1), see [7]. As another strategy for the recovery of s-row-sparse matrices X from $Y := AX$, we are now suggesting a natural extension of the HTP algorithm, which we call Simultaneous Hard Thresholding Pursuit. The implementation of this algorithm can be found on the author web page at
`www.math.drexel.edu/~foucart/software.htm`
It reads:

Start with an s-row-sparse $X^{[0]} \in \mathbb{R}^{N \times n}$, say $X^{[0]} = 0$, and iterate the steps

$$S^{[n+1]} = \left\{ \text{indices of } s \text{ largest } \left\| \left(X^{[n]} + A^\top (Y - AX^{[n]}) \right)_{(i)} \right\|_2 \right\}, \tag{SHTP$_1$}$$

$$X^{[n+1]} = \mathrm{argmin}\left\{ \|Y - AZ\|_F, \mathrm{supp}(Z) \subseteq S^{[n+1]} \right\}, \tag{SHTP$_2$}$$

until the stopping criterion $S^{[n+1]} = S^{[n]}$ is met. The convergence of the SHTP algorithm is guaranteed by the following result. A stable and robust version, in the spirit of Theorem 1, also holds in this context.

Theorem 2. *Suppose that the $3s$th Restricted Isometry Constant of the matrix $A \in \mathbb{R}^{m \times N}$ satisfies*

$$\delta_{3s} < \frac{1}{\sqrt{3}} \approx 0.57735.$$

Then, for every s-row-sparse $X \in \mathbb{R}^N$, the sequence $(X^{[n]})$ defined by (SHTP) with $Y = AX$ satisfies, for all $n \geq 0$,

$$\|X^{[n]} - X\|_F \leq \rho^n \, \|X^{[0]} - X\|_F. \tag{5}$$

The proof of this result appears in the next section. As in [3], it is possible to show that $X \in \mathbb{R}^{N \times K}$ is exactly recovered after at most

$$n'_\rho(X) := \left\lceil \frac{\ln\left(\sqrt{2/3}\,\|X\|_F/\Xi\right)}{\ln\left(1/\rho\right)} \right\rceil + 1 \tag{6}$$

iterations of the algorithm started with $X^{[0]} = 0$. Here Ξ denotes the smallest nonzero ℓ_2-norm of rows of X. For instance, if the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_K$ have nonzero entries that are almost all similar, say equal to 1, except for some very small entries at non-overlapping positions, say one $0 < \epsilon \ll 1$ per vector, then the number of iterations (6) for the SHTP run is much smaller than the number of iterations (3) for any of the HTP runs, since

$$\frac{\|X\|_F^2}{\Xi^2} = \frac{K(s-1+\epsilon^2)}{K-1+\epsilon^2} < \frac{\|\mathbf{x}_k\|_2^2}{\xi_k^2} = \frac{s-1+\epsilon^2}{\epsilon^2}.$$

In fact, we can observe that $n'_\rho(X)$ never exceeds the largest $n_\rho(\mathbf{x}_k)$. Indeed, we would otherwise have, for all $1 \leq k \leq K$,

$$\frac{\|X\|_F^2}{\Xi^2} > \frac{\|\mathbf{x}_k\|_2^2}{\xi_k^2}, \quad \text{i.e.,} \quad \|X\|_F^2 \, \xi_k^2 > \Xi^2 \, \|\mathbf{x}_k\|_2^2.$$

Summing over k and simplifying by $\|X\|_F^2 = \sum_{k=1}^K \|\mathbf{x}_k\|_2^2$ would give the contradiction

$$\sum_{k=1}^K \xi_k^2 > \Xi^2.$$

Let us now examine the cost per iteration of the SHTP algorithm in terms of number of multiplications and divisions. This cost can be split into two contributions:

- $NK(s+1) \approx NKs$ to form the row-norms in (SHTP$_1$) — note that $A^\top Y$ and $A^\top A$ are calculated once and for all at the beginning of the algorithm;
- $\sum_{\ell=1}^{s-1} \ell^2 + K \sum_{\ell=1}^{s-1} \ell \approx s^2(2s+3K)/6$ to solve the K simultaneous $s \times s$ systems of normal equations arising in (SHTP$_2$) using Gaussian elimination.

The cost per iteration of the basic HTP algorithm is obtained by substituting $K = 1$ above. We then observe that, thanks to the second step of the algorithm, a run of SHTP cost less per iteration than the K runs of HTP , since

$$\frac{s^2(2s+3K)}{6} \leq K \frac{s^2(2s+3)}{6}.$$

The simple numerical experiment of Fig. 1, carried out for Gaussian measurement matrices A and Gaussian sparse vectors $\mathbf{x}_1, \ldots, \mathbf{x}_K$, confirms that one run of SHTP is faster than K runs of HTP. It is also more reliable, according to the intuition that more information about the support is available.

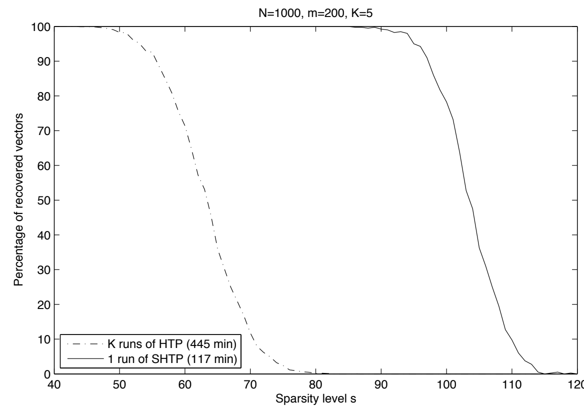

Fig. 1. Comparison of HTP and SHTP on 400 trials

3. PROOF OF THEOREM 2

This section is dedicated to the proof of the main theorem of this paper. The argument closely follows the one given in [3] for the HTP algorithm. It makes use of the following simple characterization of Restricted Isometry Constants, known for $K = 1$, although not used very often. Below, the notations $M_{(S)}$, respectively M_S, stand for the submatrices of a matrix M formed by the rows, respectively by the columns, indexed by a set S.

Lemma 1. *Given a matrix $A \in \mathbb{R}^{m \times N}$ and integers $s, K \geq 1$,*

$$\delta_s(A) = \max_{\mathrm{card}(S)=s} \; \max_{Z \in \mathbb{R}^{s \times K}, \|Z\|_F = 1} \|(I - A_S^\top A_S)Z\|_F.$$

Proof. This is just a consequence of the known characterization

$$\delta_s(A) = \max_{\mathrm{card}(S)=s} \|I - A_S^\top A_S\|_{2 \to 2},$$

and the observation that, for any matrix $M \in \mathbb{R}^{k \times \ell}$,

$$\|M\|_{2 \to 2} = \max_{Z \in \mathbb{R}^{\ell \times K}, \|Z\|_F = 1} \|MZ\|_F.$$

This observation follows, on the one hand, from

$$\begin{aligned} \|M\|_{2 \to 2} &= \max_{\|\mathbf{z}\|_2 = 1} \|M\mathbf{z}\|_2 = \max_{\|\mathbf{z}\|_2 = 1} \|M \left[\mathbf{z} \,|\, 0 \,|\cdots|\, 0\right] \|_F \\ &\leq \max_{\|Z\|_F = 1} \|MZ\|_F, \end{aligned}$$

and on the other hand from

$$\begin{aligned} \|MZ\|_F^2 &= \| \left[MZ_1 \,|\cdots|\, MZ_\ell\right] \|_F^2 = \sum_{j=1}^{\ell} \|MZ_j\|_2^2 \\ &\leq \sum_{j=1}^{\ell} \|M\|_2^2 \|Z_j\|_2^2 = \|M\|_2^2 \, \|Z\|_F^2, \end{aligned}$$

after taking the maximum over Z. □

With this lemma at hand, we now proceed with the proof of Theorem 2. The first step of the argument is a consequence of

(SHTP$_2$), which says that $AX^{[n+1]}$ is the best approximation to Y from the space $\{AZ, \mathrm{supp}(Z) \subseteq S^{[n+1]}\}$. Therefore, it is characterized by the orthogonality condition

$$\langle AX^{[n+1]} - Y, AZ\rangle_F = 0 \qquad \text{whenever } \mathrm{supp}(Z) \subseteq S^{[n+1]}.$$

Since $Y = AX$, this may be rewritten as

$$\langle X^{[n+1]} - X, A^\top AZ\rangle = 0 \qquad \text{whenever } \mathrm{supp}(Z) \subseteq S^{[n+1]}.$$

We derive in particular

$$\begin{aligned}&\|(X^{[n+1]} - X)_{(S^{[n+1]})}\|_F^2 \\ &= \big\langle X^{[n+1]} - X, (X^{[n+1]} - X)_{(S^{[n+1]})}\big\rangle_F \\ &= \big\langle X^{[n+1]} - X, (I - A^\top A)\big((X^{[n+1]} - X)_{(S^{[n+1]})}\big)\big\rangle_F.\end{aligned}$$

With $S := \mathrm{supp}(X)$ and $T^{[n+1]} := S \cup S^{[n+1]}$, we obtain

$$\begin{aligned}&\|(X^{[n+1]} - X)_{(S^{[n+1]})}\|_F^2 \\ &= \big\langle X^{[n+1]}-X, (I-A_{T^{[n+1]}}^\top A_{T^{[n+1]}})\big((X^{[n+1]}-X)_{(S^{[n+1]})}\big)\big\rangle_F \\ &\le \|X^{[n+1]} - X\|_F \ \delta_{2s}\|(X^{[n+1]} - X)_{(S^{[n+1]})}\|_F.\end{aligned}$$

After simplification, we have

$$\|(X^{[n+1]} - X)_{(S^{[n+1]})}\|_F \le \delta_{2s}\|X^{[n+1]} - X\|_F.$$

It follows that

$$\begin{aligned}&\|X^{[n+1]} - X\|_F^2 \\ &= \|(X^{[n+1]} - X)_{(\overline{S^{[n+1]}})}\|_F^2 + \|(X^{[n+1]} - X)_{(S^{[n+1]})}\|_F^2 \\ &\le \|(X^{[n+1]} - X)_{(\overline{S^{[n+1]}})}\|_F^2 + \delta_{2s}^2\|X^{[n+1]} - X\|_F^2.\end{aligned}$$

After a rearrangement, we obtain

$$\|X^{[n+1]} - X\|_F^2 \le \frac{1}{1-\delta_{2s}^2}\|(X^{[n+1]} - X)_{(\overline{S^{[n+1]}})}\|_F^2. \tag{7}$$

The second step of the argument is as a consequence of (HTP$_1$). It starts by noticing that

$$\begin{aligned}&\|(X^{[n]} + A^\top(Y - AX^{[n]}))_{(S)}\|_F^2 \\ &\qquad \le \|(X^{[n]} + A^\top(Y - AX^{[n]}))_{(S^{[n+1]})}\|_F^2.\end{aligned}$$

Eliminating the contribution on $S \cap S^{[n+1]}$, we derive

$$\begin{aligned}&\|(X^{[n]} + A^\top(Y - AX^{[n]}))_{(S\setminus S^{[n+1]})}\|_F \\ &\qquad \le \|(X^{[n]} + A^\top(Y - AX^{[n]}))_{(S^{[n+1]}\setminus S)}\|_F.\end{aligned}$$

For the right-hand side, we have

$$\begin{aligned}&\|(X^{[n]} + A^\top(Y - AX^{[n]}))_{(S^{[n+1]}\setminus S)}\|_F \\ &\qquad = \|\big((I - A^\top A)(X^{[n]} - X)\big)_{(S^{[n+1]}\setminus S)}\|_F.\end{aligned}$$

As for the left-hand side, we have

$$\begin{aligned}&\|(X^{[n]} + A^\top(Y - AX^{[n]}))_{(S\setminus S^{[n+1]})}\|_F \\ &= \|(X-X^{[n+1]})_{(\overline{S^{[n+1]}})} + \big((I-A^\top A)(X^{[n]}-X)\big)_{(S\setminus S^{[n+1]})}\|_F \\ &\ge \|(X-X^{[n+1]})_{(\overline{S^{[n+1]}})}\|_F - \|\big((I-A^\top A)(X^{[n]}-X)\big)_{(S\setminus S^{[n+1]})}\|_F.\end{aligned}$$

It follows that

$$\begin{aligned}&\|(X-X^{[n+1]})_{(\overline{S^{[n+1]}})}\|_F \\ &\quad \le \|\big((I - A^\top A)(X^{[n]}-X)\big)_{(S\setminus S^{[n+1]})}\|_F \\ &\qquad + \|\big((I - A^\top A)(X^{[n]}-X)\big)_{(S^{[n+1]}\setminus S)}\|_F \\ &\quad \le \sqrt{2}\ \|\big((I - A^\top A)(X^{[n]}-X)\big)_{(S\Delta S^{[n+1]})}\|_F.\end{aligned}$$

Thus, with $U^{[n]} := S \cup S^{[n]} \cup S^{[n+1]}$, we obtain

$$\begin{aligned}\|(X-X^{[n+1]})_{(\overline{S^{[n+1]}})}\|_F &\le \sqrt{2}\ \|(I-A_{U^{[n]}}^\top A_{U^{[n]}})(X^{[n]}-X)\|_F \\ &\le \sqrt{2}\,\delta_{3s}\|X^{[n]} - X\|_F.\end{aligned} \tag{8}$$

As a final step, we put (7) and (8) together to obtain

$$\|X^{[n+1]} - X\|_F \le \rho\,\|X^{[n]} - X\|_F, \qquad \rho := \sqrt{\frac{2\delta_{3s}^2}{1-\delta_{2s}^2}}.$$

The estimate (5) immediately follows. The coefficient ρ is less than one as soon as $2\delta_{3s}^2 < 1 - \delta_{2s}^2$. Since $\delta_{2s} \le \delta_{3s}$, this occurs as soon as $\delta_{3s} < 1/\sqrt{3}$.

4. REFERENCES

[1] T. Blumensath and M.E. Davies. Iterative hard thresholding for compressed sensing. *Appl. Comput. Harmon. Anal*, 27:265–274, 2009.

[2] W. Dai and O. Milenkovic. Subspace pursuit for compressive sensing signal reconstruction. *IEEE Trans. Inform. Theory*, 55:2230–2249, 2009.

[3] S. Foucart. Hard thresholding pursuit: an algorithm for compressive sensing. Preprint.

[4] S. Foucart. Sparse recovery algorithms: Sufficient conditions in terms of restricted isometry constants. *Proceedings of the 13th International Conference on Approximation Theory*, To appear.

[5] S. Huang and J. Zhu. Hybrid iterative thresholding for sparse signal recovery from incomplete measurements. Preprint.

[6] P. Jain, R. Meka, and I. Dhillon. Guaranteed rank minimization via singular value projection: Supplementary material. *CoRR*, abs/0909.5457, 2009.

[7] M.-J. Lai and Y. Liu. The null space property for sparse recovery from multiple measurement vectors. *Appl. Comput. Harmon. Anal.*, To appear.

[8] A. Maleki. Coherence analysis of iterative thresholding algorithms. *Proc. of Allerton Conference on Communication, Control, and Computing*, 2009.

[9] D. Needell and J.A. Tropp. Cosamp: Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmon. Anal.*, 26:301–321, 2009.