Convergence of Iterative Hard Thresholding Variants with Application to Asynchronous Parallel Methods for Sparse Recovery

Jamie Haddock[1], Deanna Needell[1], Alireza Zaeemzadeh[2], Nazanin Rahnavard[2]

[1]Department of Mathematics, University of California, Los Angeles

Emails: {jhaddock and deanna}@math.ucla.edu

[2]School of Electrical Engineering and Computer Science, University of Central Florida

Emails: {zaeemzadeh and nazanin}@eecs.ucf.edu

Recently several asynchronous parallel algorithms for sparse recovery have been proposed. These methods share an estimation of the support of the signal between nodes, which then use this information in addition to their local estimation of the support to update via an iterative hard thresholding (IHT) method. We analyze a generalized version of the IHT method run on each of the nodes and show that this method performs at least as well as the standard IHT method. We perform numerical simulations that illustrate the potential advantage these methods enjoy over the standard IHT.

I. Introduction

Sparse recovery problems have received significant attention in the past decade, particularly in the compressed sensing (CS) literature [1, 2]. CS techniques have revolutionized sensing and sampling, with applications in image reconstruction [3, 4], hyper spectral imaging [5], wireless communications [6–9], and analog to digital conversion [10]. Meanwhile, complex data-gathering devices have been developed, leading to the rapid growth of *big data*. For instance, the size of problems in hyperspectral imaging [5] are so large that they cannot be stored or solved in conventional computers. This, as well as the proliferation of inexpensive multi-processor computing systems, has motivated the study of *parallel sparse recovery*.

In parallel sparse recovery, the goal is to solve a large-scale sparse recovery problem by partitioning it among multiple processing nodes, thus reducing both the storage and computation requirements [11]. However, many recent studies [11–16] focus on *synchronous* parallel recovery of the sparse signal, meaning that some subset of the processing nodes need to wait for another subset of the nodes to complete their tasks. Of course, this approach is sensitive to slow or nonfunctional nodes. Thus, it is natural to look for algorithms that divide the large-scale sparse recovery problems among several computing nodes and solve it *asynchronously*. In [17, 18], strategies to utilize the stochastic hard thresholding (StoIHT) [19, 20] in an asynchronous manner were proposed. Instead of sharing the current solution among the processors, which is the conventional approach [11–13], an estimate of the support of the signal is shared. Nodes use not only the local information regarding their estimation of the signal but also the shared estimation of the support of the signal to learn the support locations of the signal faster than conventional serial IHT.

This material is based upon work supported by the National Science Foundation under Grant No. ECCS-1810256, CCF-1718195, Grant No. DMS-1522158, Award No. CAREER-1348721, and Grant No. BIGDATA-1740325, and the University of California, Davis Dissertation Fellowship.

In this paper, we analyze the behavior of iterative hard thresholding algorithms which use additional information (besides the current estimation of the signal) to estimate the support. In future work, we will leverage this analysis to prove recovery guarantees for the asynchronous parallel approaches of [17, 18].

II. Notation

We begin with a review of the necessary notation describing the problem set-up and the class of algorithms to be analyzed. We consider the *sparse recovery* problem of reconstructing approximately sparse $\boldsymbol{x} \in \mathbb{R}^N$ from few nonadaptive, linear, and noisy measurements, $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e}$, where $\boldsymbol{A} \in \mathbb{R}^{m \times N}$ is the measurement matrix and $\boldsymbol{e} \in \mathbb{R}^m$ is noise. A common approach to recover the approximately sparse signal $\boldsymbol{x}$ is utilizing the *iterative hard thresholding (IHT)* algorithm which applies the iterative step

$$\boldsymbol{x}^{(n+1)} = H_k(\boldsymbol{x}^{(n)} + A^T(\boldsymbol{y} - A\boldsymbol{x}^{(n)})),$$

where $H_k(\boldsymbol{z})$ is the non-linear operator that sets all but the k largest magnitude entries of $\boldsymbol{z}$ to zero. This algorithm was shown to converge to the signal (up to a convergence horizon which depends on the norm of the error $\boldsymbol{e}$ and the misestimation of the sparsity) in [19]. We extend this analysis to the class of iterative hard thresholding algorithms we denote as $\text{IHT}_{k,\tilde{k}}$, which applies the iterative step

$$\boldsymbol{x}^{(n+1)} = H_{k,\tilde{k}}(\boldsymbol{x}^{(n)} + A^T(\boldsymbol{y} - A\boldsymbol{x}^{(n)})),$$

where $H_{k,\tilde{k}}(\boldsymbol{z})$ is the non-linear operator that sets all but the k largest magnitude entries and $\tilde{k}$ additional entries of $\boldsymbol{z}$ to zero, and $\boldsymbol{x}^{(0)} = \boldsymbol{0}$. This class of algorithms allows for the additional flexibility of selecting coordinates according to a non-greedy strategy (e.g., randomly, according to prior information, according to shared information between nodes, etc.).

As in [19], we will analyze the $\text{IHT}_{k,\tilde{k}}$ algorithms for measurement matrices which satisfy the non-symmetric isometry property (which is equivalent to the standard restricted isometry property for a scaled matrix). We say that A satisfies the non-symmetric isometry property if

$$(1-\beta_k)\|\boldsymbol{z}\|_2^2 \leq \|A\boldsymbol{z}\|_2^2 \leq \|\boldsymbol{z}\|_2^2$$

for all k-sparse $\boldsymbol{z}$. Throughout what follows, we use $\|\cdot\|$ to denote the Euclidean norm and $\|\cdot\|_1$ the ℓ_1-norm.

We use the following standard notation. We let $\mathrm{supp}(\boldsymbol{a})$ denote the set of indices of the non-zero entries of $\boldsymbol{a}$ while $\mathrm{supp}_k(\boldsymbol{a})$ denotes the set of indices of the k-largest magnitude entries of $\boldsymbol{a}$. For an index set Λ, we let $\boldsymbol{a}_\Lambda$ denote the vector in which the entries indexed by Λ are equal to those of $\boldsymbol{a}$, and all others are zero. We let $\boldsymbol{x}^k = \boldsymbol{x}_{\mathrm{supp}_k(\boldsymbol{x})}$ be the best k-sparse approximation of the vector $\boldsymbol{x}$. Thus, for exactly s-sparse $\boldsymbol{x}$, $\boldsymbol{x}^s = \boldsymbol{x}$. We let the true support of signal $\boldsymbol{x}$ be denoted $\Gamma^* = \mathrm{supp}(\boldsymbol{x})$ and the support of the best k-sparse approximation to $\boldsymbol{x}$ be denoted $\Gamma^k = \mathrm{supp}_k(\boldsymbol{x})$. We let the support of the nth iterate be denoted $\Gamma^{(n)} = \mathrm{supp}(\boldsymbol{x}^{(n)})$, the support of the best k-sparse approximation to $\boldsymbol{x}^{(n)}$ be denoted $\Gamma_k^{(n)} = \mathrm{supp}_k(\boldsymbol{x}^{(n)})$, and the remaining support of the iterate be denoted $\Gamma_{\tilde{k}}^{(n)} = \Gamma^{(n)} \setminus \Gamma_k^{(n)}$. Then, for ease of analysis, we denote $\Omega^{(n)} = \Gamma^k \cup \Gamma^{(n)}$. Note that $|\Gamma^k| = k$, $|\Gamma^{(n)}| = k+\tilde{k}$, and $|\Omega^{(n)}| \leq 2k+\tilde{k}$. We additionally let the iterative difference between the $\Omega^{(n)}$ sets be denoted $\Omega_{(n)}^{(n-1)} = \Omega^{(n-1)} \setminus \Omega^{(n)}$. We define the residual vector $\boldsymbol{r}^{(n)} = \boldsymbol{x}^k - \boldsymbol{x}^{(n)}$, and the proxy vector $\boldsymbol{b}^{(n)} = \boldsymbol{x}^{(n-1)} + A^T(\boldsymbol{y} - A\boldsymbol{x}^{(n-1)})$; note then that $\boldsymbol{x}^{(n)} = H_{k,\tilde{k}}(\boldsymbol{b}^{(n)})$. Finally, we define $\tilde{\boldsymbol{e}} = A(\boldsymbol{x}-\boldsymbol{x}^k)+\boldsymbol{e}$ and $\tilde{\boldsymbol{x}}^{(n)} = \boldsymbol{x}_{\Gamma^k \cup \Gamma_{\tilde{k}}^{(n)}}$.

III. Main Result

We now show that the $\mathrm{IHT}_{k,\tilde{k}}$ algorithms converge at least as quickly as the IHT_k algorithm, and the convergence horizon depends on the approximability of the signal $\boldsymbol{x}$ by a k-sparse vector and the magnitude of the noise $\boldsymbol{e}$.

Theorem 1. *If A has the non-symmetric restricted isometry property with $\beta_{3k+2\tilde{k}} < \frac{1}{8}$, then in iteration n, the $\mathrm{IHT}_{k,\tilde{k}}$ algorithms with input observations $\boldsymbol{y} = A\boldsymbol{x}+\boldsymbol{e}$ recover the approximation $\boldsymbol{x}^{(n)}$ with*

$$\|\boldsymbol{x}-\boldsymbol{x}^{(n)}\| \leq 2^{-n}\|\boldsymbol{x}^k\| + 5\|\boldsymbol{x}-\boldsymbol{x}^k\| + \frac{4}{\sqrt{k}}\|\boldsymbol{x}-\boldsymbol{x}^k\|_1 + 4\|\boldsymbol{e}\|.$$

Proof. First, note that

$$\begin{aligned}\|\boldsymbol{x}-\boldsymbol{x}^{(n)}\| &\leq \|\boldsymbol{x}^k-\boldsymbol{x}^{(n)}\| + \|\boldsymbol{x}-\boldsymbol{x}^k\| \\ &= \|\boldsymbol{r}^{(n)}\| + \|\boldsymbol{x}-\boldsymbol{x}^k\|.\end{aligned}$$

Now, since $\boldsymbol{x}^{(n)}$ is at least as good as the best k-sparse approximation to $\boldsymbol{b}_{\Omega^{(n)}}^{(n)}$, it is better than $\boldsymbol{x}^k$, so we have

$$\begin{aligned}\|\boldsymbol{r}^{(n)}\| &\leq \|\boldsymbol{x}^k - \boldsymbol{b}_{\Omega^{(n)}}^{(n)}\| + \|\boldsymbol{x}^{(n)} - \boldsymbol{b}_{\Omega^{(n)}}^{(n)}\| \\ &\leq 2\|\boldsymbol{x}^k - \boldsymbol{b}_{\Omega^{(n)}}^{(n)}\|.\end{aligned}$$

Now, expanding $\boldsymbol{b}_{\Omega^{(n)}}^{(n)}$ and noting $\Gamma^k \subset \Omega^{(n)}$, we have

$$\begin{aligned}\|\boldsymbol{r}^{(n)}\| &\leq 2\|\boldsymbol{x}_{\Omega^{(n)}}^k - \boldsymbol{x}_{\Omega^{(n)}}^{(n-1)} - A_{\Omega^{(n)}}^T A\boldsymbol{r}^{(n-1)} - A_{\Omega^{(n)}}^T\tilde{\boldsymbol{e}}\| \\ &\leq 2\|\boldsymbol{r}_{\Omega^{(n)}}^{(n-1)} - A_{\Omega^{(n)}}^T A\boldsymbol{r}^{(n-1)}\| + 2\|A_{\Omega^{(n)}}^T\tilde{\boldsymbol{e}}\| \\ &\leq 2\|(I - A_{\Omega^{(n)}}^T A_{\Omega^{(n)}})\boldsymbol{r}_{\Omega^{(n)}}^{(n-1)}\| \\ &\quad + 2\|A_{\Omega^{(n)}}^T A_{\Omega^{(n-1)}\setminus\Omega^{(n)}}\boldsymbol{r}_{\Omega^{(n-1)}\setminus\Omega^{(n)}}^{(n-1)}\| \\ &\quad + 2\|A_{\Omega(n)}^T\tilde{\boldsymbol{e}}\|.\end{aligned}$$

We have $2\|(I - A_{\Omega^{(n)}}^T A_{\Omega^{(n)}})\boldsymbol{r}_{\Omega^{(n)}}^{(n-1)}\| \leq 2\beta_{2k+\tilde{k}}\|\boldsymbol{r}_{\Omega^{(n)}}^{(n-1)}\|$ by Lemma 1 of [19]. Additionally, we have $2\|A_{\Omega^{(n)}}^T A_{\Omega^{(n-1)}\setminus\Omega^{(n)}}\boldsymbol{r}_{\Omega^{(n-1)}\setminus\Omega^{(n)}}^{(n-1)}\| \leq 2\beta_{3k+2\tilde{k}}\|\boldsymbol{r}_{\Omega^{(n-1)}\setminus\Omega^{(n)}}^{(n-1)}\|$ by Proposition 3.2 of [21]. By the non-symmetric isometry property, $\|A_{\Omega(n)}^T\tilde{\boldsymbol{e}}\| \leq \|\tilde{\boldsymbol{e}}\|$ and $\beta_{2k+\tilde{k}} \leq \beta_{3k+2\tilde{k}}$ by definition. Thus,

$$\begin{aligned}\|\boldsymbol{r}^{(n)}\| &\leq 2\beta_{3k+2\tilde{k}}\|\boldsymbol{r}_{\Omega^{(n)}}^{(n-1)}\| + 2\beta_{3k+2\tilde{k}}\|\boldsymbol{r}_{\Omega^{(n-1)}\setminus\Omega^{(n)}}^{(n-1)}\| + 2\|\tilde{\boldsymbol{e}}\| \\ &\leq 4\beta_{3k+2\tilde{k}}\|\boldsymbol{r}^{(n-1)}\| + 2\|\tilde{\boldsymbol{e}}\|.\end{aligned}$$

Now, if $\beta_{3k+2\tilde{k}} < \frac{1}{8}$, then $\|\boldsymbol{r}^{(n)}\| \leq \frac{1}{2}\|\boldsymbol{r}^{(n-1)}\| + 2\|\tilde{\boldsymbol{e}}\|$, so iterating this inequality yields

$$\begin{aligned}\|\boldsymbol{r}^{(n)}\| &\leq 2^{-n}\|\boldsymbol{r}^{(0)}\| + 4\|\tilde{\boldsymbol{e}}\| \\ &= 2^{-n}\|\boldsymbol{x}^k\| + 4\|\tilde{\boldsymbol{e}}\|.\end{aligned}$$

Finally, we have

$$\begin{aligned}\|\boldsymbol{x}-\boldsymbol{x}^{(n)}\| &\leq 2^{-n}\|\boldsymbol{x}^k\| + 4\|\tilde{\boldsymbol{e}}\| + \|\boldsymbol{x}-\boldsymbol{x}^k\| \\ &\leq 2^{-n}\|\boldsymbol{x}^k\| + 5\|\boldsymbol{x}-\boldsymbol{x}^k\| + \frac{4}{\sqrt{k}}\|\boldsymbol{x}-\boldsymbol{x}^k\|_1 + 4\|\boldsymbol{e}\|\end{aligned}$$

since $\|\tilde{\boldsymbol{e}}\| \leq \|\boldsymbol{x}-\boldsymbol{x}^k\| + \frac{1}{\sqrt{k}}\|\boldsymbol{x}-\boldsymbol{x}^k\|_1 + \|\boldsymbol{e}\|$ by Lemma 6.1 in [21]. ☐

Now, while we see now that the $\mathrm{IHT}_{k,\tilde{k}}$ algorithms perform at least as well as the IHT_k algorithm, we consider a special case in which these methods perform better than standard IHT_k algorithm. In this case, the convergence horizon defined by the difference between the signal and its sparse approximation is decreased by the selection of the additional $\tilde{k}$ indices in the approximation.

Theorem 2. *Suppose the signal $\boldsymbol{x}$ has constant values on its support, and the $\tilde{k}$ indices selected (non-greedily) by the $\mathrm{IHT}_{k,\tilde{k}}$ algorithm each lie uniformly in the support of $\boldsymbol{x}$ with probability p. If A has the non-symmetric restricted isometry property with $\beta_{3k+2\tilde{k}} < \frac{1}{8}$, then in iteration n, the $\mathrm{IHT}_{k,\tilde{k}}$ algorithms with input observations $\boldsymbol{y} = A\boldsymbol{x}+\boldsymbol{e}$ recover the approximation $\boldsymbol{x}^{(n)}$ with*

$$\begin{aligned}\mathbb{E}_{\tilde{k}}\|\boldsymbol{x}-\boldsymbol{x}^{(n)}\| &\leq 2^{-n}\|\boldsymbol{x}\| + 5\mathbb{E}_{\tilde{k}}\|\boldsymbol{x}-\tilde{\boldsymbol{x}}^{(n)}\| \\ &\quad + \frac{4}{\sqrt{k}}\mathbb{E}_{\tilde{k}}\|\boldsymbol{x}-\tilde{\boldsymbol{x}}^{(n)}\|_1 + 4\|\boldsymbol{e}\| \\ &\leq 2^{-n}\|\boldsymbol{x}\| + \left(5\alpha + \frac{4\alpha}{\sqrt{k}}\right)\|\boldsymbol{x}\|_1 + 4\|\boldsymbol{e}\|\end{aligned}$$

where $\alpha = \left(\frac{|supp(\boldsymbol{x})|-k}{|supp(\boldsymbol{x})|}\right)\left(\frac{|supp(\boldsymbol{x})|-p\tilde{k}}{|supp(\boldsymbol{x})|}\right)$.

Proof. We begin by applying the triangle inequality,

$$\|\boldsymbol{x} - \boldsymbol{x}^{(n)}\| \le \|\tilde{\boldsymbol{x}}^{(n)} - \boldsymbol{x}^{(n)}\| + \|\boldsymbol{x} - \tilde{\boldsymbol{x}}^{(n)}\|.$$

We have that $\boldsymbol{x}^{(n)}$ is at least as good as the best k-sparse approximation to $\boldsymbol{b}^{(n)}_{\Omega^{(n)}}$, it is better than $\tilde{\boldsymbol{x}}^{(n)}_{\Gamma^k}$ and $\boldsymbol{x}^{(n)}$ agrees with $\boldsymbol{b}^{(n)}_{\Omega^{(n)}}$ on $\Gamma^{(n)}_{\tilde{k}}$, so we have

$$\begin{aligned}\|\tilde{\boldsymbol{x}}^{(n)} - \boldsymbol{x}^{(n)}\| &\le \|\tilde{\boldsymbol{x}}^{(n)} - \boldsymbol{b}^{(n)}_{\Omega^{(n)}}\| + \|\boldsymbol{x}^{(n)} - \boldsymbol{b}^{(n)}_{\Omega^{(n)}}\| \\ &\le 2\|\tilde{\boldsymbol{x}}^{(n)} - \boldsymbol{b}^{(n)}_{\Omega^{(n)}}\|.\end{aligned}$$

Expanding $\boldsymbol{b}^{(n)}_{\Omega^{(n)}}$ and applying [19, Lemma 1] and [21, Proposition 3.2], we have

$$\begin{aligned}\|\tilde{\boldsymbol{x}}^{(n)} - \boldsymbol{x}^{(n)}\| &\le 2\|\tilde{\boldsymbol{x}}^{(n)} - \boldsymbol{x}^{(n-1)}_{\Omega^{(n)}} - A^T_{\Omega^{(n)}}(\boldsymbol{y} - A\boldsymbol{x}^{(n-1)})_{\Omega^{(n)}}\| \\ &\le 2\|\tilde{\boldsymbol{x}}^{(n)} - \boldsymbol{x}^{(n-1)}_{\Omega^{(n)}} - A^T_{\Omega^{(n)}}A(\tilde{\boldsymbol{x}}^{(n)} - \boldsymbol{x}^{(n-1)})\| \\ &\quad + 2\|A^T_{\Omega^{(n)}}A\boldsymbol{x} - A^T_{\Omega^{(n)}}A\tilde{\boldsymbol{x}}^{(n)} + A^T_{\Omega^{(n)}}\boldsymbol{e}\| \\ &\le 2\|(I - A^T_{\Omega^{(n)}}A_{\Omega^{(n)}})(\tilde{\boldsymbol{x}}^{(n)}_{\Omega^{(n)}} - \boldsymbol{x}^{(n-1)}_{\Omega^{(n)}})\| \\ &\quad + 2\|A^T_{\Omega^{(n)}}A_{\Omega^{(n-1)}_{(n)}}(\tilde{\boldsymbol{x}}^{(n)} - \boldsymbol{x}^{(n-1)})_{\Omega^{(n-1)}_{(n)}}\| \\ &\quad + 2\|A^T_{\Omega^{(n)}}A\boldsymbol{x} - A^T_{\Omega^{(n)}}A\tilde{\boldsymbol{x}}^{(n)} + A^T_{\Omega^{(n)}}\boldsymbol{e}\| \\ &\le 2\beta_{2k+\tilde{k}}\|\tilde{\boldsymbol{x}}^{(n)}_{\Omega^{(n)}} - \boldsymbol{x}^{(n-1)}_{\Omega^{(n)}}\| \\ &\quad + 2\beta_{3k+2\tilde{k}}\|\tilde{\boldsymbol{x}}^{(n)}_{\Omega^{(n-1)}_{(n)}} - \boldsymbol{x}^{(n-1)}_{\Omega^{(n-1)}_{(n)}}\| \\ &\quad + 2\|A^T_{\Omega^{(n)}}A\boldsymbol{x} - A^T_{\Omega^{(n)}}A\tilde{\boldsymbol{x}}^{(n)} + A^T_{\Omega^{(n)}}\boldsymbol{e}\|\end{aligned}$$

By the non-symmetric isometry property, $\|A^T_{\Omega(n)}\boldsymbol{w}\| \le \|\boldsymbol{w}\|$ and $\beta_{2k+\tilde{k}} \le \beta_{3k+2\tilde{k}}$ by definition. Thus,

$$\begin{aligned}\|\tilde{\boldsymbol{x}}^{(n)} - \boldsymbol{x}^{(n)}\| &\le 4\beta_{3k+2\tilde{k}}\|\tilde{\boldsymbol{x}}^{(n)} - \boldsymbol{x}^{(n-1)}\| \\ &\quad + 2\|A(\boldsymbol{x} - \tilde{\boldsymbol{x}}^{(n)}) + \boldsymbol{e}\|\end{aligned}$$

Now, if $\beta_{3k+2\tilde{k}} < \frac{1}{8}$, then we can iterate this inequality to yield

$$\begin{aligned}\|\tilde{\boldsymbol{x}}^{(n)} - \boldsymbol{x}^{(n)}\| &\le 2^{-n}\|\tilde{\boldsymbol{x}}^{(n)} - \boldsymbol{x}^{(0)}\| \\ &\quad + 4\|A(\boldsymbol{x} - \tilde{\boldsymbol{x}}^{(n)}) + \boldsymbol{e}\| \\ &= 2^{-n}\|\tilde{\boldsymbol{x}}^{(n)}\| \\ &\quad + 4\|A(\boldsymbol{x} - \tilde{\boldsymbol{x}}^{(n)}) + \boldsymbol{e}\|\end{aligned}$$

Finally, we have

$$\begin{aligned}\|\boldsymbol{x} - \boldsymbol{x}^{(n)}\| &\le 2^{-n}\|\boldsymbol{x}\| + 4\|A(\boldsymbol{x} - \tilde{\boldsymbol{x}}^{(n)})\| \\ &\quad + 4\|\boldsymbol{e}\| + \|\boldsymbol{x} - \tilde{\boldsymbol{x}}^{(n)}\|\end{aligned}$$

and since $\tilde{\boldsymbol{x}}^{(n)}$ is a best $r := |\mathrm{supp}_k(\boldsymbol{x}) \cup (\mathrm{supp}(\boldsymbol{x}) \cap \Gamma^{(n)}_{\tilde{k}})|$-sparse approximation to $\boldsymbol{x}$, by [21, Lemma 6.1], we have $\|A(\boldsymbol{x} - \tilde{\boldsymbol{x}}^{(n)})\| \le \|\boldsymbol{x} - \tilde{\boldsymbol{x}}^{(n)}\| + \frac{1}{\sqrt{r}}\|\boldsymbol{x} - \tilde{\boldsymbol{x}}^{(n)}\|_1$. Thus, finally we have

$$\begin{aligned}\mathbb{E}_{\tilde{k}}\|\boldsymbol{x} - \boldsymbol{x}^{(n)}\| &\le 2^{-n}\|\boldsymbol{x}\| + 5\mathbb{E}_{\tilde{k}}\|\boldsymbol{x} - \tilde{\boldsymbol{x}}^{(n)}\| \\ &\quad + 4\|\boldsymbol{e}\| + 4\mathbb{E}_{\tilde{k}}\left[\frac{1}{\sqrt{r}}\|\boldsymbol{x} - \tilde{\boldsymbol{x}}^{(n)}\|_1\right] \\ &\le 2^{-n}\|\boldsymbol{x}\| + 5\mathbb{E}_{\tilde{k}}\|\boldsymbol{x} - \tilde{\boldsymbol{x}}^{(n)}\| \\ &\quad + 4\|\boldsymbol{e}\| + \frac{4}{\sqrt{k}}\mathbb{E}_{\tilde{k}}\|\boldsymbol{x} - \tilde{\boldsymbol{x}}^{(n)}\|_1.\end{aligned}$$

Now, we need only analyze $\mathbb{E}_{\tilde{k}}\|\boldsymbol{x} - \tilde{\boldsymbol{x}}^{(n)}\|_q$ where $q = 1, 2$. We do these analysis for both $q = 1$ and $q = 2$ next. We apply the triangle inequality and then simplify, yielding

$$\begin{aligned}\mathbb{E}_{\tilde{k}}\|\boldsymbol{x} - \tilde{\boldsymbol{x}}^{(n)}\|_q &\le \mathbb{E}_{\tilde{k}}\|\boldsymbol{x}_{\mathrm{supp}_k(\boldsymbol{x})} - \tilde{\boldsymbol{x}}^{(n)}_{\mathrm{supp}_k(\boldsymbol{x})}\|_q \\ &\quad + \mathbb{E}_{\tilde{k}}\|\boldsymbol{x}_{\Gamma_{\tilde{k}(n)}} - \tilde{\boldsymbol{x}}^{(n)}_{\Gamma_{\tilde{k}(n)}}\|_q \\ &\quad + \mathbb{E}_{\tilde{k}}\|\boldsymbol{x}_{\mathrm{supp}(\boldsymbol{x})\setminus\mathrm{supp}(\tilde{\boldsymbol{x}}^{(n)})}\|_q \\ &\le \mathbb{E}_{\tilde{k}}\|\boldsymbol{x}_{\mathrm{supp}(\boldsymbol{x})\setminus\mathrm{supp}(\tilde{\boldsymbol{x}}^{(n)})}\|_1 \\ &\le \frac{\|\boldsymbol{x}\|_1}{|\mathrm{supp}(\boldsymbol{x})|}\mathbb{E}_{\tilde{k}}|\mathrm{supp}(\boldsymbol{x})\setminus\mathrm{supp}(\tilde{\boldsymbol{x}}^{(n)})| \\ &\le \frac{\|\boldsymbol{x}\|_1}{|\mathrm{supp}(\boldsymbol{x})|}\Big[|\mathrm{supp}(\boldsymbol{x})| - k \\ &\qquad - \mathbb{E}_{\tilde{k}}\Big|\mathrm{supp}(\boldsymbol{x}) \cap (\Gamma^{(n)}_{\tilde{k}}\setminus\mathrm{supp}_k(\boldsymbol{x})\Big|\Big] \\ &= \frac{\|\boldsymbol{x}\|_1}{|\mathrm{supp}(\boldsymbol{x})|}[|\mathrm{supp}(\boldsymbol{x})| - k]\left[1 - \frac{p\tilde{k}}{|\mathrm{supp}(\boldsymbol{x})|}\right] \\ &= \alpha\|\boldsymbol{x}\|_1.\end{aligned}$$

Applying this bound into the previous bound on $\mathbb{E}_{\tilde{k}}\|\boldsymbol{x} - \boldsymbol{x}^{(n)}\|$ yields the required result. □

IV. Experiments

In the first experiment, we take the signal dimension $N = 1000$, the sparsity level $s := |\mathrm{supp}(\boldsymbol{x})| = 20$, and the number of measurements $m = 300$. Here, the nonzero entries of signal $\boldsymbol{x}$ are standard normal random variables (so they do not satisfy the identical assumption of Theorem 2). Even so, when we perform 100 iterations of $\mathrm{IHT}_{k,\tilde{k}}$ where $k = \tilde{k} = 5$ and the $\tilde{k}$ indices each lie in supp($\boldsymbol{x}$) with probability p (for various values of p), we see that the convergence horizon is decreased over standard IHT_k as in Theorem 2. Note that in this case, standard IHT_k coincides with $\mathrm{IHT}_{k,\tilde{k}}$ where $p = 0$. These results are averaged over 100 trials and plotted in Figure 1.

We additionally explore the rate at which the extra indices shared between nodes by the *Asynchronous StoIHT (AStoIHT)* [17] and the *Bayesian Asynchronous StoIHT (BAStoIHT)* [18] lie in the true support of the signal $\boldsymbol{x}$. In these experiments, we run AStoIHT and BAStoIHT on various numbers of nodes n_c and plot the rate that the indices shared between nodes (i.e., not selected greedily using local node information), $T^{(n)}$, lie in the true support, Γ^*. This is computed as $\frac{T^{(n)} \cap \Gamma^*}{s}$ and is plotted in Figure 2. Similar to the previous experiment and following [17, 18], we $N = 1000$, $s = 20$, $m = 300$ and the convergence criteria is $\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^{(n)}\| \le 10^{-7}$. Furthermore, in each trial, half of the cores are four times slower than the other half. These results are averaged over 50 trials.

V. Conclusions

We have analyzed both theoretically and empirically the behavior of the $\mathrm{IHT}_{k,\tilde{k}}$ algorithms. This provides heuristic evidence for why BAStoIHT outperforms its undistributed counterpart IHT. We prove that $\mathrm{IHT}_{k,\tilde{k}}$ performs at least as well as IHT_k. We additionally prove that $\mathrm{IHT}_{k,\tilde{k}}$ outperforms IHT_k in the special case where the signal $\boldsymbol{x}$ has constant

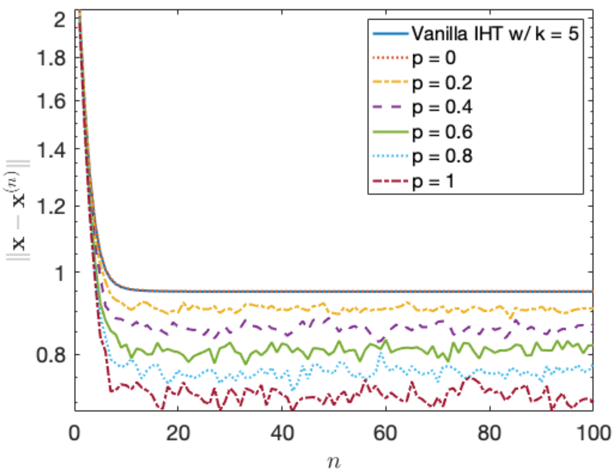
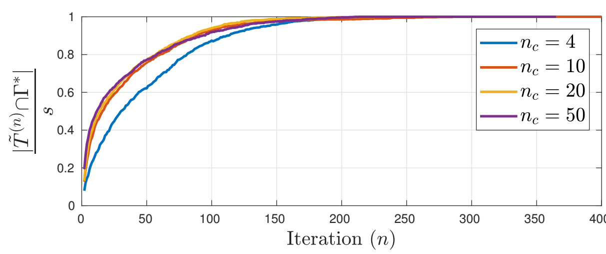
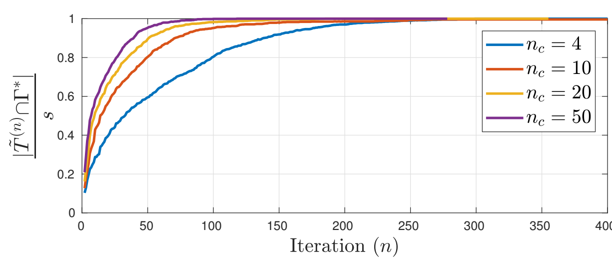


Figure 1: Plot of error $\|\boldsymbol{x}-\boldsymbol{x}^{(n)}\|$ vs. iteration for 100 iterations of $\mathrm{IHT}_{k,\tilde{k}}$ with various probabilities p that the $\tilde{k}$ indices lie in $\mathrm{supp}(\boldsymbol{x})$.



(a)



(b)

Figure 2: The rate at which the shared indices between nodes lie in the true support of signal $\boldsymbol{x}$ for iterations of (a) AStoIHT and (b) BAStoIHT.

entries on its support and the extra $\tilde{k}$ indices selected lie in the signal support with fixed probability. Finally, we show that this holds experimentally and provide evidence that the extra indices shared between nodes in BAStoIHT lies in the signal support at a high rate.

References

[1] D. L. D. Donoho, "Compressed sensing," vol. 52, pp. 1289–1306, 4 2006.

[2] E. E. J. Candès, "Compressive sampling," 2006.

[3] B. Shahrasbi and N. Rahnavard, "Model-Based Nonuniform Compressive Sampling and Recovery of Natural Images Utilizing a Wavelet-Domain Universal Hidden Markov Model," *IEEE Transactions on Signal Processing*, 2017.

[4] D. Needell and R. Ward, "Stable Image Reconstruction Using Total Variation Minimization," *SIAM Journal on Imaging Sciences*, vol. 6, pp. 1035–1058, 1 2013.

[5] J. Tan, Y. Ma, H. Rueda, D. Baron, and G. R. Arce, "Compressive Hyperspectral Imaging via Approximate Message Passing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, pp. 389–401, 3 2016.

[6] C. Jeon, R. Ghods, A. Maleki, and C. Studer, "Optimality of large MIMO detection via approximate message passing," in *2015 IEEE International Symposium on Information Theory (ISIT)*, pp. 1227–1231, IEEE, 6 2015.

[7] M. Joneidi, A. Zaeemzadeh, B. Shahrasbi, G.-J. Qi, and N. Rahnavard, "E-optimal Sensor Selection for Compressive Sensing-based Purposes," *IEEE Transactions on Big Data*, p. To Appear in, 2018.

[8] A. Talari and N. Rahnavard, "CStorage: Distributed Data Storage in Wireless Sensor Networks Employing Compressive Sensing," in *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, pp. 1–5, 2011.

[9] M. T. Nguyen, K. A. Teague, and N. Rahnavard, "CCS: Energy-efficient data collection in clustered wireless sensor networks utilizing block-wise compressive sensing," *Computer Networks*, vol. 106, pp. 171–185, 2016.

[10] S. Salehi, M. B. Mashhadi, A. Zaeemzadeh, N. Rahnavard, and R. F. De Mara, "Energy-Aware Adaptive Rate and Resolution Sampling of Spectrally Sparse Signals Leveraging VCMA-MTJ Devices," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, pp. 1–1, 2018.

[11] J. Zhu, R. Pilgrim, and D. Baron, "An overview of multi-processor approximate message passing," in *2017 51st Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–6, IEEE, 3 2017.

[12] S. Patterson, Y. C. Eldar, and I. Keidar, "Distributed Compressed Sensing for Static and Time-Varying Networks," *IEEE Transactions on Signal Processing*, vol. 62, pp. 4931–4946, 10 2014.

[13] C. Ravazzi, S. M. Fosson, and E. Magli, "Distributed iterative thresholding for ℓ_0/ℓ_1-regularized linear inverse problems," *IEEE Transactions on Information Theory*, vol. 61, pp. 2081–2100, April 2015.

[14] Y. Ma, Y. M. Lu, and D. Baron, "Multiprocessor approximate message passing with column-wise partitioning," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2017.

[15] P. Han, R. Niu, M. Ren, and Y. C. Eldar, "Distributed approximate message passing for sparse signal recovery," in *2014 IEEE Global Conference on Signal and Information Processing, GlobalSIP 2014*, 2014.

[16] P. Han, J. Zhu, R. Niu, and D. Baron, "Multi-processor approximate message passing using lossy compression," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2016.

[17] D. Needell and T. Woolf, "An asynchronous parallel approach to sparse recovery," in *2017 Information Theory and Applications Workshop (ITA)*, pp. 1–5, IEEE, 2 2017.

[18] A. Zaeemzadeh, J. Haddock, N. Rahnavard, and D. Needell, "A Bayesian approach for asynchronous parallel sparse recovery," in *Proc. 52nd Asilomar Conference on Signals, Systems and Computers*, 2018.

[19] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, pp. 265–274, 11 2009.

[20] N. Nguyen, D. Needell, and T. Woolf, "Linear Convergence of Stochastic Iterative Greedy Algorithms With Sparse Constraints," *IEEE Transactions on Information Theory*, vol. 63, pp. 6869–6895, 11 2017.

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