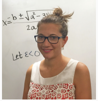
Convergence of Iterative Hard Thresholding Variants with Application to Asynchronous Parallel Methods for Sparse Recovery

Jamie Haddock

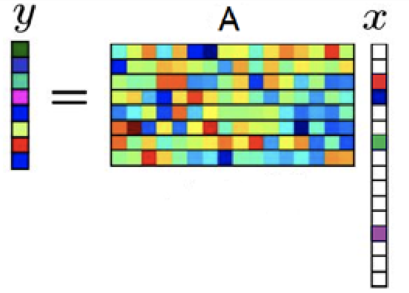
Asilomar Conference on Signals, Systems, and Computers,
November 4, 2019

Computational and Applied Mathematics
UCLA



joint with Deanna Needell, Nazanin Rahnavard, and Alireza Zaeezadeh

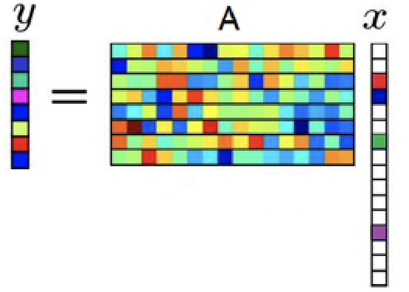
Sparse Recovery Problem



Sparse Recovery: reconstruct approximately sparse $\boldsymbol{x} \in \mathbb{R}^N$ from few nonadaptive, linear, and noisy measurements, $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e}$

$\boldsymbol{A} \in \mathbb{R}^{m \times N}$: measurement matrix
$\boldsymbol{e} \in \mathbb{R}^m$: noise

Sparse Recovery Problem



Sparse Recovery: reconstruct approximately sparse $\boldsymbol{x} \in \mathbb{R}^N$ from few nonadaptive, linear, and noisy measurements, $\boldsymbol{y} = \boldsymbol{Ax} + \boldsymbol{e}$

$\boldsymbol{A} \in \mathbb{R}^{m \times N}$: measurement matrix
$\boldsymbol{e} \in \mathbb{R}^m$: noise

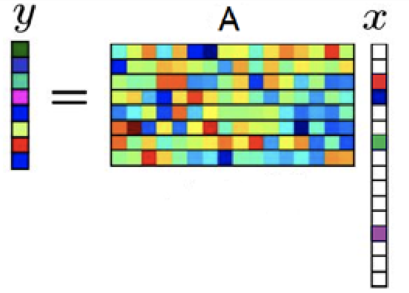
Approach:

$\min_{\mathbf{x} \in \mathbb{R}^N} \|\mathbf{x}\|_1$ s.t. $\|\mathbf{Ax} - \mathbf{y}\| \leq \epsilon$

or

$\min_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{m} \|\mathbf{Ax} - \mathbf{y}\|^2$ s.t. $\|\mathbf{x}\|_0 \leq s$

Sparse Recovery Problem



Sparse Recovery: reconstruct approximately sparse $\boldsymbol{x} \in \mathbb{R}^N$ from few nonadaptive, linear, and noisy measurements, $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e}$

$\boldsymbol{A} \in \mathbb{R}^{m \times N}$: measurement matrix
$\boldsymbol{e} \in \mathbb{R}^m$: noise

Applications:

- ▷ image reconstruction
- ▷ hyper spectral imaging
- ▷ wireless communications
- ▷ analog to digital conversion

Approach:

$\min_{\mathbf{x} \in \mathbb{R}^N} \|\mathbf{x}\|_1$ s.t. $\|\mathbf{A}\mathbf{x} - \mathbf{y}\| \leq \epsilon$

or

$\min_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{m} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2$ s.t. $\|\mathbf{x}\|_0 \leq s$

Algorithmic Approaches

Convex optimization:

- linear programming
- (proximal) gradient descent
- coordinate descent
- stochastic iterative methods (SGD)

Algorithmic Approaches

Convex optimization:

- linear programming
- (proximal) gradient descent
- coordinate descent
- stochastic iterative methods (SGD)

Greedy pursuits:

- orthogonal matching pursuit (OMP)
- regularized OMP (ROMP)
- compressive sampling matching pursuit (CoSaMP)
- iterative hard thresholding (IHT)

Algorithmic Approaches

Convex optimization:

- linear programming
- (proximal) gradient descent
- coordinate descent
- stochastic iterative methods (SGD)

Greedy pursuits:

- orthogonal matching pursuit (OMP)
- regularized OMP (ROMP)
- compressive sampling matching pursuit (CoSaMP)
- iterative hard thresholding (IHT)

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

StoIHT

Algorithm 1 StoIHT Algorithm [22]

input: s, γ, $p(i)$, and stopping criterion
initialize: x^1 and $t = 1$
repeat
  randomize: select $i_t \in [M]$ with probability $p(i_t)$
  proxy: $b^t = x^t + \frac{\gamma}{Mp(i_t)} A^*_{b_{i_t}} (y_{b_{i_t}} - A_{b_{i_t}} x^t)$
  identify: $\Gamma^t = \text{supp}_s(b^t)$
  estimate: $x^{t+1} = b^t_{\Gamma^t}$
  $t = t + 1$
until halting criterion *true*
output: $\hat{x} = x^t$

[1]

[1] Nguyen, Needell, Woolf, IEEE Transactions on Information Theory '17

Asynchronous Parallelization

Asynchronous approaches: popular when the objective functions are sparse in $\mathbf{x}$

Asynchronous Parallelization

Asynchronous approaches: popular when the objective functions are sparse in $\mathbf{x}$

- ▷ all cores run simultaneously accessing and updating shared memory as necessary

Asynchronous Parallelization

Asynchronous approaches: popular when the objective functions are sparse in $\mathbf{x}$

- ▷ all cores run simultaneously accessing and updating shared memory as necessary
- ▷ eliminates idle time of synchronous approaches

Asynchronous Parallelization

Asynchronous approaches: popular when the objective functions are sparse in $\mathbf{x}$

- ▷ all cores run simultaneously accessing and updating shared memory as necessary
- ▷ eliminates idle time of synchronous approaches

Challenge: objective of $\min_{\mathbf{x}\in\mathbb{R}^N} \frac{1}{m}\|\mathbf{A}\mathbf{x}-\mathbf{y}\|^2$ s.t. $\|\mathbf{x}\|_0 \leq s$ is dense in $\mathbf{x}$

Asynchronous Parallelization

Asynchronous approaches: popular when the objective functions are sparse in $\mathbf{x}$

- ▷ all cores run simultaneously accessing and updating shared memory as necessary
- ▷ eliminates idle time of synchronous approaches

Challenge: objective of $\min_{\mathbf{x}\in\mathbb{R}^N} \frac{1}{m}\|\mathbf{Ax}-\mathbf{y}\|^2$ s.t. $\|\mathbf{x}\|_0 \leq s$ is dense in $\mathbf{x}$

- ▷ likely that same non-zero entries are updated from one iteration to the next

Asynchronous Parallelization

Asynchronous approaches: popular when the objective functions are sparse in $\mathbf{x}$

- ▷ all cores run simultaneously accessing and updating shared memory as necessary
- ▷ eliminates idle time of synchronous approaches

Challenge: objective of $\min_{\mathbf{x}\in\mathbb{R}^N} \frac{1}{m}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2$ s.t. $\|\mathbf{x}\|_0 \leq s$ is dense in $\mathbf{x}$

- ▷ likely that same non-zero entries are updated from one iteration to the next
- ▷ a slow core could easily "undo" the progress of previous updates by faster cores

Asynchronous StoIHT

Algorithm 2 Asynchronous StoIHT Iteration

Each core performs the following at each iteration. The tally vector ϕ is available to each core.

randomize: select $i_t \in [M]$ with probability $p(i_t)$

proxy: $b^t = x^t + \frac{\gamma}{Mp(i_t)} A^{\star}_{b_{i_t}}(y_{b_{i_t}} - A_{b_{i_t}} x^t)$

identify: $\Gamma^t = \mathrm{supp}_s(b^t)$

$\widetilde{T}^t = \mathrm{supp}_s(\phi)$

estimate: $x^{t+1} = b^t_{\Gamma^t \cup \widetilde{T}^t}$

update tally: $\phi_{\Gamma^t} = \phi_{\Gamma^t} + t$

$\phi_{\Gamma^{t-1}} = \phi_{\Gamma^{t-1}} - (t-1)$

$t = t + 1$

[2]

[2] Needell, Woolf, Proc. Information Theory and Applications '17

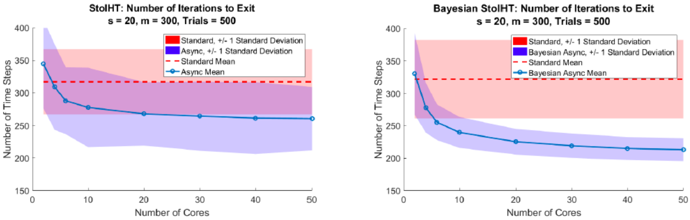
Bayesian Asynchronous StoIHT

Require: Number of subproblems, M, and probability of selection $p(B)$. The reliability score distribution parameters, $\hat{\beta}_i^1$ and $\hat{\beta}_i^0$, and the tally scores parameters, $\hat{a}_n^1$ and $\hat{a}_n^0$, are available to each processor.
Each processor performs the following at each iteration:

1: **randomize:** select $B_t \in [M]$ with probability $p(B_t)$
2: **proxy:** $\boldsymbol{b}^{(t)} = \boldsymbol{x}^{(t)} + \frac{\gamma}{Mp(B_t)} \boldsymbol{A}_{B_t}^* (\boldsymbol{y}_{B_t} - \boldsymbol{A}_{B_t} \boldsymbol{x}^{(t)})$
3: **identify:** $\hat{\mathcal{S}}^{(t)} = supp_s(\boldsymbol{b}^{(t)})$ and $\tilde{T}^{(t)} = supp_s(\phi)$
4: **estimate:** $\boldsymbol{x}^{(t+1)} = \boldsymbol{b}^{(t)}_{\hat{\mathcal{S}}^{(t)} \cup \tilde{T}^{(t)}}$
5: **repeat**
6: update $\mathbb{E}_{\mathbb{Q}\{u_{ni}\}}\{u_{ni}\} = \mathbb{Q}\{u_{ni} = 1\}$
7: update $\hat{\beta}_i^1$ and $\hat{\beta}_i^0$, $\hat{a}_n^1$ and $\hat{a}_n^0$
8: **until** convergence
9: update ϕ
10: $t = t + 1$

[2] Zaeemzadeh, H., Rahnavard, Needell, Proc. 49th Asilomar Conf. on Signals, Systems and Computers '18

Experimental Convergence



(a) StoIHT (b) Bayesian StoIHT

Figure 2: Number of time steps until convergence versus number of cores used in (a) asynchronous StoIHT method and (b) Bayesian asynchronous StoIHT. Half of the cores are *slow* and complete an iteration only once out of every four time steps.

Tools for Analysis

First step: analyze IHT variant running on each node of parallel system

Tools for Analysis

First step: analyze IHT variant running on each node of parallel system

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

Tools for Analysis

First step: analyze IHT variant running on each node of parallel system

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

Non-Symmetric Isometry Property:

$(1-\beta_k)\|\boldsymbol{z}\|_2^2 \leq \|\mathbf{A}\boldsymbol{z}\|_2^2 \leq \|\boldsymbol{z}\|_2^2$ for all k-sparse $\boldsymbol{z}$

Convergence of IHT$_{k,\tilde{k}}$

Theorem (H., Needell, Zaeemzadeh, Rahnavard '19+)

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

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

An Improved Scenario

Theorem (H., Needell, Zaeemzadeh, Rahnavard '19+)

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

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

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

Experimental Convergence of $\text{IHT}_{k,\tilde{k}}$

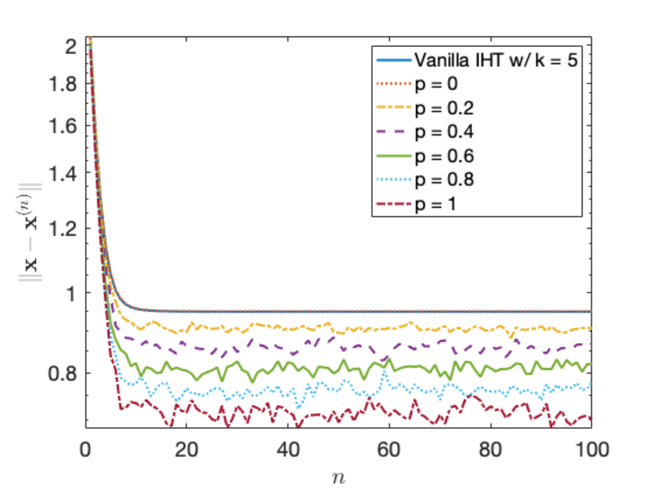
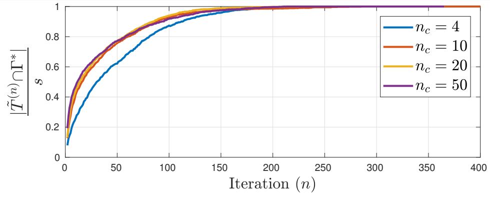
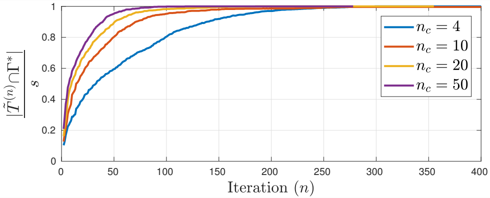


Figure 1: Plot of error $\|\mathbf{x} - \mathbf{x}^{(n)}\|$ vs. iteration for 100 iterations of $\text{IHT}_{k,\tilde{k}}$ with various probabilities p that the $\tilde{k}$ indices lie in $\text{supp}(\mathbf{x})$.

Rate of Support Intersection



(a)



(b)

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

Conclusions and Future Work

Conclusions and Future Work

- provided a convergence analysis for an IHT variant

Conclusions and Future Work

- provided a convergence analysis for an IHT variant
- identified scenario when IHT variant has potentially faster convergence

Conclusions and Future Work

- provided a convergence analysis for an IHT variant
- identified scenario when IHT variant has potentially faster convergence
- provided heuristic for why asynchronous versions of StoIHT converge faster than non-parallelized version

Conclusions and Future Work

- provided a convergence analysis for an IHT variant
- identified scenario when IHT variant has potentially faster convergence
- provided heuristic for why asynchronous versions of StoIHT converge faster than non-parallelized version

- analyze $\text{StoIHT}_{k,\tilde{k}}$

Conclusions and Future Work

- provided a convergence analysis for an IHT variant
- identified scenario when IHT variant has potentially faster convergence
- provided heuristic for why asynchronous versions of StoIHT converge faster than non-parallelized version

- analyze StoIHT$_{k,\tilde{k}}$
- extend to non-heuristic analysis of Asynchronous StoIHT

Thanks for listening!

Questions?

[1] J. Haddock, D. Needell, N. Rahnavard, and A. Zaeemzadeh. **Convergence of iterative hard thresholding variants with application to asynchronous parallel methods for sparse recovery.** In Proc. Asilomar Conf. Sig. Sys. Comp., 2019.

[2] Deanna Needell and Tina Woolf. **An asynchronous parallel approach to sparse recovery.** In 2017 Information Theory and Applications Workshop (ITA), pages 1–5. IEEE, 2 2017.

[3] Nam Nguyen, Deanna Needell, and Tina Woolf. **Linear Convergence of Stochastic Iterative Greedy Algorithms With Sparse Constraints.** IEEE Transactions on Information Theory, 63(11):6869–6895, 11 2017.

[4] A. Zaeemzadeh, J. Haddock, N. Rahnavard, and D. Needell. **A Bayesian approach for asynchronous parallel sparse recovery.** In Proc. Asilomar Conf. Sig. Sys. Comp., 2018.