Model Expander Iterative Hard Thresholding

Bubacarr Bah

Inst. of Computational Engineering and Sciences (ICES), and Dept. of Mathematics
The University of Texas at Austin

SPARS 2015
July 6-9
Cambridge, UK

Joint work with **Luca Baldassarre**, and **Volkan Cevher** at LIONS, EPFL

Algorithm is ...

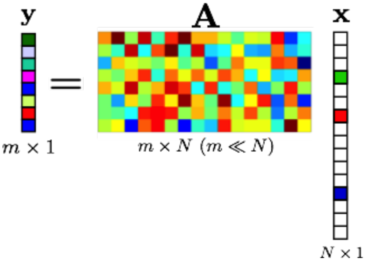
designed for structured sparse recovery using model expander sensing matrices, & it's particularly suited for linear sketching

Thus this talk will discuss ...

1. Structured sparse recovery
2. Model expander matrices
3. Algorithm and it's features
4. Convergence of algorithm
5. Experiements

Three key aspects of linear sketching

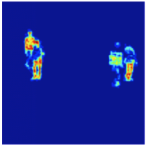
- Signal (vector) **x**
  sparse or compressible
- Projection **A**
  information preserving
  (stable embedding)
- Recovery algorithm Δ
  tractable & accurate

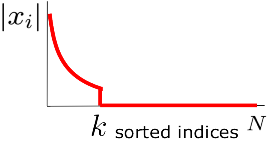


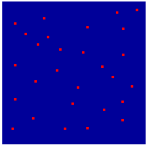
- Applications: Data streaming, compressive sensing (CS), graph sketching, machine learning, group testing, etc.

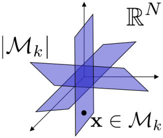
Sparsity and beyond

- Generic sparsity (or compressibility) not specific enough

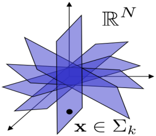








Note: $\mathcal{M}_k \subseteq \Sigma_k$

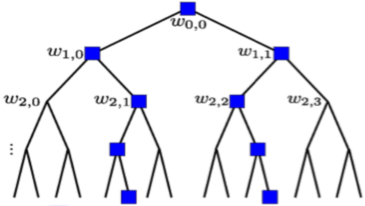


- Many applications exhibit some structure in sparsity pattern
- $\Rightarrow$ structured sparsity $\rightarrow$ model-based CS [Baraniuk et al. 2010]

Model-based CS

- Model-based CS exploits structure in sparsity model to
  - improve interpretability
  - reduce sketch length
  - increase speed of recovery
- Models of structured sparsity includes trees, blocks, groups, ...

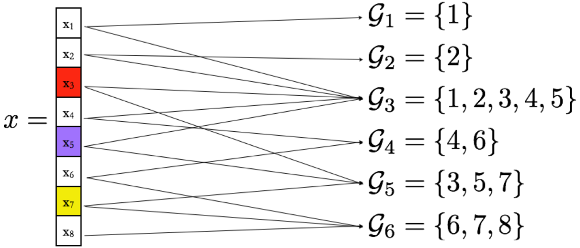
tree-sparse



Block-sparse

Overlapping Group Models

A natural generalization of sparsity



Group models application examples:

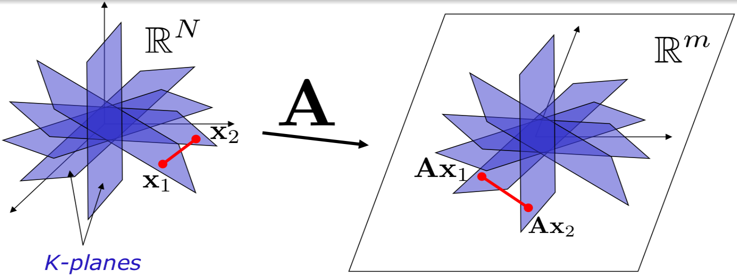
- Genetic Pathways in Microarray data analysis
- Wavelet models in image processing
- Brain regions in neuroimaging

Information preserving linear embeddings **A**

Definition (ℓ_p-norm Restricted Isometry Property (RIP-p))

A matrix **A** has RIP-p of order k, if for all k-sparse **x**, it satisfies

$$(1-\delta_k)\|\mathbf{x}\|_p^p \leq \|\mathbf{A}\mathbf{x}\|_p^p \leq (1+\delta_k)\|\mathbf{x}\|_p^p$$



Information preserving linear embeddings **A**

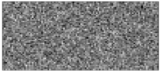
Definition (ℓ_p-norm Restricted Isometry Property (RIP-p))

A matrix **A** has RIP-p of order k, if for all k-sparse **x**, it satisfies

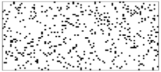
$$(1-\delta_k)\|\mathbf{x}\|_p^p \leq \|\mathbf{A}\mathbf{x}\|_p^p \leq (1+\delta_k)\|\mathbf{x}\|_p^p$$

- Subgaussian **A** $\in \mathbb{R}^{m\times N}$ (w.h.p) have RIP-2 with $m = \mathcal{O}(k\log(N/k))$, but sparse binary **A** does not have RIP-2 unless $m = \Omega\left(k^2\right)$
- Sparse adjacency matrices of **lossless expanders** satisfy RIP-1 with $m = \mathcal{O}(k\log(N/k))$
- Structured sparsity $\Rightarrow$ fewer m for model-RIP-2
  - $\mathcal{O}(k)$ for tree structure
  - $\mathcal{O}(k + \log(M))$ for block structure; M blocks

A



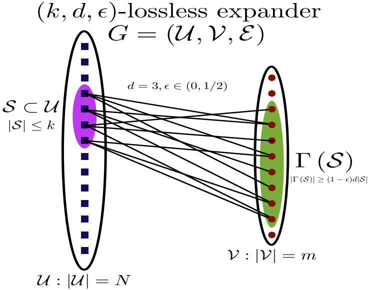
A



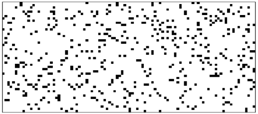
Sparse matrices from expanders

Definition (Lossless Expander Graphs)

$G = (\mathcal{U}, \mathcal{V}, \mathcal{E})$ is an (k, d, ϵ)-lossless expander if it is a bipartite graph with $|\mathcal{U}| = N$ left vertices, $|\mathcal{V}| = m$ right vertices & has a regular left degree d, s.t. any $\mathcal{S} \subset \mathcal{U}$ with $|\mathcal{S}| \leq k$ has $|\Gamma(\mathcal{S})| \geq (1 - \epsilon)\, d|\mathcal{S}|$ neighbors



A is sparse (d nonzeros per col.)



Computational benefits of **A**

- Low storage complexity
- Efficient application

Tractability of recovery

Nonlinear reconstruction

Given $\mathbf{A}$ & $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ with $\|\mathbf{e}\|_2 \leq \eta$, find k-sparse $\hat{\mathbf{x}}$ satisfying:

$$\hat{\mathbf{x}} = \min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \eta.$$

- Tractable recovery algorithms (Δ) with provable guarantees
  - Convex approach: ℓ_1-minimization

    $$\hat{\mathbf{x}} = \min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \eta$$

  - Discrete algorithms (OMP, IHT, CoSaMP, EIHT, ALPS, ...)

    $$\text{(IHT)} \quad \text{iterates} \quad \mathbf{x}^{n+1} = \mathcal{H}_k\left(\mathbf{x}^n + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^n)\right)$$

Accuracy of recovery

- Δ returns approximations with ℓ_p/ℓ_q-approximation error:

Definition (ℓ_p/ℓ_q-approximation error - instance optimality)

A Δ returns $\hat{\mathbf{x}} = \Delta(\mathbf{A}\mathbf{x} + \mathbf{e})$ with ℓ_p/ℓ_q-approximation error if

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_p \leq C_1 \sigma_k(\mathbf{x})_q + C_2 \|\mathbf{e}\|_p$$

for a noise vector $\mathbf{e}$, $C_1, C_2 > 0$, $1 \leq q \leq p \leq 2$, $\sigma_k(\mathbf{x})_q := \min_{k-\text{sparse } \mathbf{x}'} \|\mathbf{x} - \mathbf{x}'\|_q$

- The pair $(\mathbf{A}, \Delta)$ $\Rightarrow$ two types of error guarantees
  - *for each* - one pair $(\mathbf{A}, \Delta)$ for each given $\mathbf{x}$
  - *for all* - one pair $(\mathbf{A}, \Delta)$ for all $\mathbf{x}$

Goal of this work

To design an algorithm that makes it possible to efficiently exploit the benefits of combining the sparsity in $\mathbf{A}$ with structured sparsity in $\mathbf{x}$

- Prior work on model-based CS use dense $\mathbf{A}$
- Dense matrices: difficult to store, create computational bottlenecks, and not practical in real applications
- Sparse matrices: low storage complexity, efficient application, etc
- Existing recovery algorithm for such sparse matrices has exponential complexity

Contribution summary

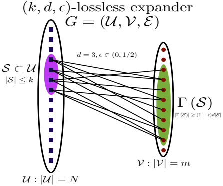
“Tractable” linear complexity algorithm with provable **for all** ℓ_1/ℓ_1 approximation guarantees

Definition (RIP-1 for (k, d, ϵ)-lossless expanders)

If $\mathbf{A}$ is an adjacency matrix of a (k, d, ϵ)-lossless expanders, then $\Phi = \mathbf{A}/d$ has RIP-1 of order k, if for all k-sparse $\mathbf{x}$, it satisfies

$$(1 - 2\epsilon)\|\mathbf{x}\|_1 \leq \|\Phi\mathbf{x}\|_1 \leq \|\mathbf{x}\|_1$$

- Probabilistic constructions of expanders achieve optimal $m = \mathcal{O}\left(k \log(N/k)\right)$
- But their deterministic constructions are sub-optimal $m = \mathcal{O}\left(k^{1+\alpha}\right)$ for $\alpha > 0$



Standard random construction of $G = ([N], [m], \mathcal{E})$

For every $u \in [N]$, uniformly sample a subset of $[m]$ of size d and connect u and all the vertices from this subset

Models everywhere

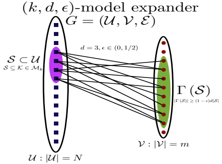
- $\mathcal{T}_k$ & $\mathfrak{G}_k$ denotes tree & loopless overlapping groups respectively, which are jointly denoted by $\mathcal{M}_k$

Definition (Model sparse vectors)

A vector $\mathbf{x}$ is $\mathcal{M}_k$-sparse if $\mathrm{supp}(\mathbf{x}) \subseteq \mathcal{K}$ for $\mathcal{K} \in \mathcal{M}_k$

Definition ((k, d, ϵ)-model expander graph)

Let $\mathcal{K} \in \mathcal{M}_k$, G is a model expander if for all $\mathcal{S} \subseteq \mathcal{K}$, we have $\left|\Gamma(\mathcal{S})\right| \geq (1 - \epsilon)d|\mathcal{S}|$



Definition (Model expander matrix)

A matrix $\mathbf{A}$ is a model expander if it is the adjacency matrix of a (k, d, ϵ)-model expander graph.

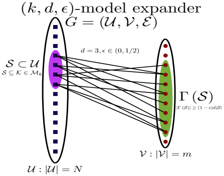
Model-Expander Iterative Hard Thresholding (MEIHT)

Initialize $\mathbf{x}^0 = \mathbf{0}$, iterate

$$\mathbf{x}^{n+1} = \mathcal{P}_{\mathcal{M}_k}\left[\mathbf{x}^n + \mathfrak{M}\left(\mathbf{y} - \mathbf{A}\mathbf{x}^n\right)\right]$$

- $\mathfrak{M}(\cdot)$ is the median operator which returns a vector $\mathfrak{M}(\mathbf{u}) \in \mathbb{R}^N$ for an input $\mathbf{u} \in \mathbb{R}^m$; defined elementwise

$$[\mathfrak{M}(\mathbf{u})]_i := \mathrm{med}[u_j, j \in \Gamma(i)], i \in [N]$$



- $\mathcal{P}_{\mathcal{M}_k}(\mathbf{u}) \in \mathrm{argmin}_{\mathbf{z} \in \mathcal{M}_k} \{\|\mathbf{u} - \mathbf{z}\|_1\}$ is the ℓ_1 projection of $\mathbf{u}$ onto $\mathcal{M}_k$
- MEIHT is a fusion (with adaptation) of various works:
  - SMP of [Berinde et al. 2008]
  - EIHT of [Foucart & Rauhut 2013]
  - Tractable group projections of [Baldassare et al. 2013]

Tractability of structured sparse models

- The projection is equivalent to Weighted Max Cover (WMC) for group-sparse problems

$$\mathcal{P}_{\mathcal{M}}(\mathbf{u}) = \min_{\mathbf{z}:\mathrm{supp}(\mathbf{z})\in\mathcal{M}} \|\mathbf{z}-\mathbf{u}\|_1 = \max_{\mathcal{S}\in\mathcal{M}} \|\mathbf{u}_{\mathcal{S}}\|_1 \equiv \text{WMC}$$

- So all WMC instances can be formulated as $\mathcal{P}_{\mathcal{M}}(\cdot)$
- Caveat: WMC is NP-hard $\Rightarrow$ $\mathcal{P}_{\mathcal{M}}(\cdot)$ is NP-hard too
- But for some models, $\mathcal{M}_k$ (i.e. $\mathcal{T}_k$ & $\mathfrak{G}_k$) in particular, there exist linear time algorithms
- Like dynamic programs that recursively compute the optimal solution via the model graph [Baldassarre et al. 2013]

Runtime: *polynomial* in N for all tractable models

- Due to the sparsity of $\mathbf{A}$, the projection onto the model is the **dominant** operation in MEIHT
- Based on the projection complexity from [Baldassarre et al. 2013], for fixed iterations, n, MEIHT achieves linear runtime of:
  - $\mathcal{O}(knN)$ for the $\mathcal{T}_k$ model
  - $\mathcal{O}(M^2 kn + nN)$ for the $\mathfrak{G}_k$ model; M groups

Error guarantees: ℓ_1/ℓ_1 in the *for all* case

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq C_1 \sigma_{\mathcal{M}_k}(\mathbf{x})_1 + C_2 \|\mathbf{e}\|_1$$

where $C_1, C_2 > 0$ and $\sigma_{\mathcal{M}_k}(\mathbf{x})_1 := \min_{\mathbf{x}' \in \mathcal{M}_k} \|\mathbf{x} - \mathbf{x}'\|_1$

- Approximate solutions are in the model, $\mathcal{M}_k$; this is very useful for some applications

Lemma (Key ingredient of proof)

Let $\mathbf{A} \in \{0,1\}^{m\times N}$ *be a* $(k, d, \epsilon_{\mathcal{M}_k})$*-model expander. If* $\mathcal{S} \subset [N]$ *is* $\mathcal{M}_k$*-sparse, then for all* $\mathbf{x} \in \mathbb{R}^N$ *and* $\mathbf{e} \in \mathbb{R}^m$,

$$\| \left[\mathfrak{M}\left(\mathbf{A}\mathbf{x}_{\mathcal{S}} + \mathbf{e}\right) - \mathbf{x}\right]_{\mathcal{S}} \|_1 \leq \frac{4\epsilon_{\mathcal{M}_k}}{1-4\epsilon_{\mathcal{M}_k}}\|\mathbf{x}_{\mathcal{S}}\|_1 + \frac{2}{(1-4\epsilon_{\mathcal{M}_k})\,d}\|\mathbf{e}_{\Gamma(\mathcal{S})}\|_1$$

- For $\mathcal{Q}^{n+1} := \mathcal{S} \cup \operatorname{supp}(\mathbf{x}^n) \cup \operatorname{supp}\left(\mathbf{x}^{n+1}\right)$, the triangle inequality yields

$$\|\mathbf{x}^{n+1} - \mathbf{x}_{\mathcal{S}}\|_1 \leq 2\| \left[\mathbf{x}_{\mathcal{S}} - \mathbf{x}^n - \mathfrak{M}\left(\mathbf{A}\left(\mathbf{x}_{\mathcal{S}} - \mathbf{x}^n\right) + \mathbf{A}\mathbf{x}_{\bar{\mathcal{S}}} + \mathbf{e}\right)\right]_{\mathcal{Q}^{n+1}} \|_1$$

- Using the nestedness property of $\mathcal{M}_k$ and the lemma gives:

$$\|\mathbf{x}^{n+1} - \mathbf{x}_{\mathcal{S}}\|_1 \leq \frac{8\epsilon_{\mathcal{M}_{3k}}}{1-4\epsilon_{\mathcal{M}_{3k}}}\|\mathbf{x}_{\mathcal{S}} - \mathbf{x}^n\|_1 + \frac{4}{\left(1-4\epsilon_{\mathcal{M}_{3k}}\right)d}\|\mathbf{A}\mathbf{x}_{\bar{\mathcal{S}}} + \mathbf{e}\|_1$$

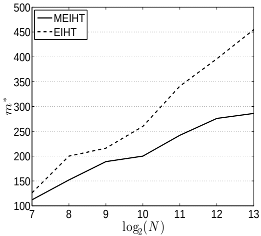
- Taking $\lim_{n\to\infty} \mathbf{x}^n = \hat{\mathbf{x}}$, using the RIP-1 property of $\mathbf{A}$ and the triangle inequality with the condition $\epsilon_{\mathcal{M}_{3k}} < 1/12$, we have:

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_1 \leq C_1\sigma_{\mathcal{M}_k}(\mathbf{x})_1 + C_2\|\mathbf{e}\|_1, \quad C_2 = \beta = 4\left(\left(1-12\epsilon_{\mathcal{M}_{3k}}\right)d\right)^{-1},\ C_1 = 1 + \beta d$$

■

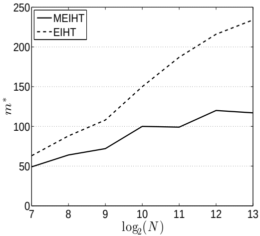
- Simulations, with different N, on group and tree models
- The median over different realizations of the minimum no. of samples for which $\frac{\|\hat{\mathbf{x}}-\mathbf{x}\|_1}{\|\mathbf{x}\|_1} \leq 10^{-5}$ is plotted for MEIHT & EIHT

Group sparse



$M = \lfloor N/\log_2(N) \rfloor$, $g = \lfloor N/M \rfloor$,
$k = 5$, $d = \lfloor 2\log(N)/\log(kg) \rfloor$

Tree sparse



$m \in [2k, 10\log_2(N)]$, $k = \lfloor 2\log_2(N) \rfloor$,
$d = \lfloor 5\log(N/k)/(2\log\log(N/k)) \rfloor$

- MEIHT requires fewer measurements than EIHT as expected

Summary

- MEIHT for model-based sketching with sparse matrices
- MEIHT has linear runtime & achieves ℓ_1/ℓ_1 error in the for all case
- MEIHT in proper perspective

	Price 2011	I. & R. 2013[1]	this work
Structures (models)	block & tree	tree	tree & groups
Error guarantees	ℓ_2/ℓ_2	ℓ_1/ℓ_1	ℓ_1/ℓ_1
Guarantee types	for each	for all	for all
Runtime complexity	sublinear	exponential	linear

[1] Indyk and Razenshteyn 2013

Possible extensions

- Implementation of MEIHT in lower level languages like C/C++
- Using MEIHT in real-life sketching & CS applications

References

[1] B. Bah, L. Baldassarre, & V. Cevher. *Model-based sketching and recovery with expanders*. ACM-SIAM Symposium on Discrete Algorithms (SODA '14)

[2] L. Baldassarre, N. Bhan, V. Cevher, & A. Kyrillidis, *Group-sparse model selection: Hardness and relaxations*, arXiv, (2013)

[3] R. Baraniuk, V. Cevher, M. Duarte, & C. Hegde, *Model-based compressive sensing*, IEEE IT. on, 56 (2010), pp. 1982-2001

[4] S. Foucart & H. Rauhut, *A mathematical introduction to compressive sensing*, Applied Numerical Harmonic Analysis Birkhäuser, Boston, (2013)

[5] P. Indyk & I. Razenshteyn, *On model-based RIP-1 matrices*, arXiv:1304.3604, (2013)

[6] E. Price, *Efficient sketches for the set query problem*, in Proceedings of the 22nd Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, 2011, pp. 41-56

THANK YOU