Model-based Sketching and Recovery with Expanders

Bubacarr Bah, Luca Baldassarre, and Volkan Cevher

Institute of Computational Engineering and Sciences (ICES), UT Austin

AMS Special Session on Frames and Their Applications

JMM 2015
Three key aspects of linear sketching

- Sparse or compressible $\mathbf{x}$
  - not sufficient alone

- Projection $\mathbf{A}$
  - information preserving
    - (stable embedding)

- Recovery algorithm $\Delta$
  - tractable & correct

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

- **Generic sparsity** (or compressibility) not specific enough

![Diagram](attachment:image.png)

- **Structured sparsity** $\Rightarrow$ **model-based CS** [Baraniuk, Cevher, Duarte, Hegde, IEEE Transactions on Information Theory 2010]:

Note: $\mathcal{M}_k \subseteq \Sigma_k$
Model-based CS

- Model-based CS exploits **structure** in sparsity model
  - improves interpretability
  - reduces sketch length
  - increases speed of recovery

- tree-sparse
- Block-sparse
Overlapping Group Models

A natural generalization of sparsity

\[ x = \begin{align*}
& x_1 \\
& x_2 \\
& x_3 \quad \mathcal{G}_1 = \{1\} \\
& x_4 \quad \mathcal{G}_2 = \{2\} \\
& x_5 \quad \mathcal{G}_3 = \{1, 2, 3, 4, 5\} \\
& x_6 \quad \mathcal{G}_4 = \{4, 6\} \\
& x_7 \quad \mathcal{G}_5 = \{3, 5, 7\} \\
& x_8 \quad \mathcal{G}_6 = \{6, 7, 8\}
\end{align*} \]

Group models application examples:

- Genetic Pathways in Microarray data analysis
- Wavelet models in image processing
- Brain regions in neuroimaging
Definition ($\ell_p$-norm Restricted Isometry Property (RIP-$p$))

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

$$(1 - \delta_k)\|x\|_p^p \leq \|Ax\|_p^p \leq (1 + \delta_k)\|x\|_p^p$$

$\mathbb{R}^N$ and $A$ transforming to $\mathbb{R}^m$.
Information preserving linear embeddings $A$

**Definition ($\ell_p$-norm Restricted Isometry Property (RIP-$p$))**

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

$$(1 - \delta_k)\|x\|_p^p \leq \|Ax\|_p^p \leq (1 + \delta_k)\|x\|_p^p$$

- **Subgaussian** $A \in \mathbb{R}^{m \times N}$ (w.h.p) have RIP-2 with $m = O(k \log(N/k))$, but sparse binary $A$ does not have RIP-2 unless $m = \Omega(k^2)$
- Model sparsity requires fewer $m$ for RIP-2
  - $O(k)$ for tree structure
  - $O(k + \log(M))$ for block structure with $M$ blocks [Baraniuk et al. ’10]
- Scaled adjacency mat. of lossless expanders have RIP-1 with $m = O(k \log(N/k))$
Recovery algorithms

- Tractable recovery algorithms ($\Delta$) with provable guarantees
  - Convex $\ell_1$-minimization approaches, and
  - Discrete algorithms (OMP, IHT, CoSaMP, ALPS)

- $\Delta$ returns approximations with $\ell_p/\ell_q$-approximation error:

**Definition ($\ell_p/\ell_q$-approximation error - instance optimality)**

$A \Delta$ returns $\hat{x} = \Delta(Ax + e)$ with $\ell_p/\ell_q$-approximation error if

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

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

- The pair $(A, \Delta)$ ⇒ two types of error guarantees
  - *for each* - one pair $(A, \Delta)$ for each given $x$
  - *for all* - one pair $(A, \Delta)$ for all $x$
Goal of this work

To combine benefits of sparsity in $A$ and benefits of model-based CS

- Prior work on model-based CS use dense $A$
- Difficult to store, creates computational bottlenecks, and not practical in real applications
Our results in perspective

<table>
<thead>
<tr>
<th></th>
<th>Price 2011</th>
<th>L. &amp; R. 2013</th>
<th>this work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Models (structures)</td>
<td>standard</td>
<td>tree(^1)</td>
<td>tree(^2) &amp; groups(^3)</td>
</tr>
<tr>
<td>Error guarantees</td>
<td>(\ell_2/\ell_2)</td>
<td>(\ell_1/\ell_1)</td>
<td>(\ell_1/\ell_1)</td>
</tr>
<tr>
<td>Guarantee types</td>
<td>for each</td>
<td>for all</td>
<td>for all</td>
</tr>
<tr>
<td>Recovery algorithm</td>
<td>sublinear</td>
<td>exponential</td>
<td>polynomial</td>
</tr>
</tbody>
</table>

\(^1\)binary trees, \(^2\)\(D\)-ary trees for \(D \geq 2\), \(^3\)Loopless overlapping groups

Contribution summary

- **Primary**: “Tractable” algorithm with provable *for all* \(\ell_1/\ell_1\) error
- **Secondary**: Existence of model expander (model-RIP-1) \(A\), consistent with known sampling bounds, for *more general* models
PART I: Existence of Model Expanders
Definition (RIP-1 for \((k, d, \epsilon)\)-lossless expanders)

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

\[
(1 - 2\epsilon)\|x\|_1 \leq \|\Phi x\|_1 \leq \|x\|_1
\]

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

Standard random construction of \(G = ([N], [m], E)\)

For every \(u \in [N]\), sample a subset of \([m]\) of size \(d\) and connect \(u\) and all the vertices from this subset
Definition ((Nested) Model sparse vectors)

A vector $x$ is $M_k$-sparse if $\text{supp}(x) \subseteq K$ for $K \in M_k$.

Definition ($(k, d, \epsilon)$-model expander graph)

Let $K \in M_k$, $G$ is a model expander if for all $S \subseteq K$, we have $|\Gamma(S)| \geq (1 - \epsilon)d|S|$.

Definition (Model expander matrix)

A matrix $A$ is a model expander if it is the adjacency matrix of a $(k, d, \epsilon)$-model expander graph.
Randomized model RIP-1 constructions

Theorem \(((k, d, \epsilon)\)-model expanders for \(D\)-ary \((D \geq 2)\) tree models)

These exist with \(d = O\left(\frac{\log(N/k)}{\epsilon \log \log(N/k)}\right)\) and \(m = O\left(\frac{dk}{\epsilon}\right)\).

- **Note:** \(D\) is subsumed \((\text{as } \log(D))\) in the order constant for \(m\)
- This matches bounds for binary tree models by [I. & R. ’13]

Theorem \(((k, d, \epsilon)\)-model expanders for overlapping group models)

For \(M > 2\) number of groups of maximum size \(g_{\text{max}} = \omega(\log N)\) such that \(N \geq kg_{\text{max}}\), these exist with \(d = O\left(\frac{\log(N)}{\epsilon \log(kg_{\text{max}})}\right)\) and \(m = O\left(\frac{dkg_{\text{max}}}{\epsilon}\right)\).

- This matches bounds for block sparsity models by [I. & R. ’13]
- **Note:** Block sparse models are a subset of the loopless overlapping group sparsity models
Our approach-I

- **Proof technique** similar to those of [Indyk & Razenshteyn ’13]
- Key ingredient of the proof is the **standard tail inequality**

**Lemma** (For \( G = ([N], [m], E) \), a variant proven in [Buhrman et al. 2002])

There exist \( C > 1 \) and \( \mu > 0 \) such that, whenever \( m \geq Cdt/\epsilon \), for any \( T \subseteq [N] \) with \( |T| = t \) we have:

\[
\text{Prob}[\{j \in [m] : \exists i \in T, e_{ij} \in E\}] < (1 - \epsilon)dt \leq \left(\frac{\mu \epsilon m}{dt}\right)^{-\epsilon dt}
\]

- Then a **union bound** over all \( M_k \)-sparse sets of sparsity \( t \)
- The **enumeration** of the cardinality of these sets involves
  - Pfaff-Fuss-Catalan or \( k \)-Raney numbers for \( T_k \)
  - a careful **counting** of such groups in \( \mathcal{G}_k \)
Our approach-II

Lemma ([Bah, Baldassarre, and Cevher 2014])

Let $T_k$-sparse & $\mathcal{G}_k$-sparse sets with sparsity $t$ be $T_{k,t}$ & $\mathcal{G}_{k,t}$ respectively & the Catalan no. be $T_k$, then $|T_{k,t}| \leq \min\left[ T_k\left(\frac{k}{t}\right), \left(\frac{N}{t}\right) \right]$, $|\mathcal{G}_{k,t}| \leq \min\left[ \left(\frac{M}{k}\right)\left(\frac{kg_{\text{max}}}{t}\right), \left(\frac{N}{t}\right) \right]$

- It suffice to show that the following holds
  $$|M_{k,t}| \cdot \left(\frac{\mu t}{d}\right)^{-\epsilon dt} \leq f(N)$$
  where $f(N) = \text{decaying function of } N$, we used $f(N) = 1/N$

- First, bound $|M_{k,t}|$ using the fact that $\binom{n}{s} \leq \left(\frac{en}{s}\right)^s$

- Substitute for $d$ & $m$ as given with arbitrary order constants

- Finally, show that for different values of $t \in [1, k]$ for $T_{k,t}$ or $t \in [1, kg_{\text{max}}]$ for $\mathcal{G}_{k,t}$, this bound holds
PART II: Model Expander Algorithm
Model-Expander Iterative Hard Thresholding (MEIHT)

Initialize $x^0 = 0$, iterate

$$x^{n+1} = P_{M_k} \left[ x^n + M (y - Ax^n) \right]$$

- $M(\cdot)$ is the median operator which returns a vector $M(u) \in \mathbb{R}^N$ for an input $u \in \mathbb{R}^m$; defined elementwise $[M(u)]_i := \text{med}[u_j, j \in \Gamma(i)], i \in [N]$.
- **Note**: $M$ operates like an adjoint.

- $P_{M_k}(u) \in \arg\min_{z \in M_k} \{ ||u - z||_1 \}$ is the projection of $u$ onto $M_k$.

- MEIHT is a fusion of various works [Berinde & Indyk 2008; Foucart & Rauhut 2013; Baldassare, Bhan, and Cevher 2013; Baraniuk, Cevher, Duarte, and Hegde 2010].
Tractability of structured sparse models

\[ \min_{\mathbf{z}: \text{supp}(\mathbf{z}) \in M} \| \mathbf{z} - \mathbf{u} \|_1 = \max_{\text{supp}(\mathbf{z}) \subseteq S \in M} \| \mathbf{u}_S \|_1 \equiv \text{Weighted Max Cover (WMC) for group-sparse problems} \]

- All WMC instances can be formulated as \( P_M(\cdot) \)
- Caveat: WMC is NP-hard \( \Rightarrow P_M(\cdot) \) is NP-hard too
- But: for some \( M, M_k \) (i.e. \( T_k \) & \( G_k \)) in particular, \( \exists \) linear time algorithms
- These include dynamic programs that recursively compute the optimal solution via the model graph [Baldassarre, Bhan, Cevher 2013]
Runtime: *polynomial* in $N$ for all tractable models

- Thanks to the sparsity of $A$, the model projections are the dominant operation in MEIHT.
- Thus, using projection complexity from [Baldassarre et al. 2013], for a fixed $n$ MEIHT achieves linear runtime of:
  - $O(knN)$ for the $T_k$ model
  - $O(M^2kn + nN)$ for the $G_k$ model; $M$ groups

Error guarantees: $\ell_1/\ell_1$ in the *for all* case

\[
\|x - \hat{x}\|_1 \leq C_1\sigma_{M_k}(x)_1 + C_2\|e\|_1
\]

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

- **Approximate solutions** are in the model, $M_k$; this is very useful for some applications.
Lemma (Key ingredient of proof)

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

$$\| \left[ \mathcal{M} (Ax_S + e) - x \right]_S \|_1 \leq \frac{4\epsilon_{M_k}}{1 - 4\epsilon_{M_k}} \|x_S\|_1 + \frac{2}{(1 - 4\epsilon_{M_k})d} \|e_{\Gamma(S)}\|_1$$

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

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

- Using the nestedness property of $M_k$ and the lemma gives:

$$\|x^{n+1} - x_S\|_1 \leq \frac{8\epsilon_{M_{3k}}}{1 - 4\epsilon_{M_{3k}}} \|x_S - x^n\|_1 + \frac{4}{(1 - 4\epsilon_{M_{3k}})d} \|Ax_{\bar{S}} + e\|_1$$

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

$$\|\hat{x} - x\|_1 \leq C_1 \sigma_{M_k}(x)_1 + C_2 \|e\|_1, \quad C_2 = \beta = 4 \left((1 - 12\epsilon_{M_{3k}})d\right)^{-1}, \quad C_1 = 1 + \beta d$$
Simulations, with different $N$, on group and tree models

The median over different realizations of the minimum no. of samples for which $\frac{||\hat{x} - x||_1}{||x||_1} \leq 10^{-5}$ is plotted for MEIHT & EIHT

Group sparse

$$M = \lceil \frac{N}{\log_2(N)} \rceil, \quad g = \lceil \frac{N}{M} \rceil, \quad k = 5, \quad d = \lceil \frac{2 \log(N)}{\log(kg)} \rceil$$

Tree sparse

$$m \in [2k, 10 \log_2(N)], \quad k = \lceil 2 \log_2(N) \rceil, \quad d = \lceil 5 \log(N/k)/(2 \log \log(N/k)) \rceil$$

MEIHT requires fewer measurements than EIHT as expected
A surprising result

*Constant node degree, $d = 16$*
Summary

- Model expanders = model-based sketching + sparse matrices; ⇒ improvement in sampling and recovery
- Proposed an efficient algorithm with linear runtime for models considered & achieves $\ell_1/\ell_1$ guarantees in the for all case
- Random construction of model expanders for more a general class of models provably possible

Extensions

- Basis adaptivity for when the $\mathbf{x}$ is sparse in a basis not canonical
- Explicit construction of model expanders
- Application of model expanders to real-life sketching & compressed sensing applications
References


