Nearly Linear-Time Algorithms for Structured Sparsity

Chinmay Hegde
Piotr Indyk
Ludwig Schmidt
MIT
Compressive sensing

• Setup:
  – Data/signal in n-dimensional space : x
    E.g., x is an 256x256 image ⇒ n=65536
  – Goal: compress x into Ax,
    where A is a m x n “measurement” or “sketch” matrix,
    m << n

• Goal: want to recover an “approximation” x* of k-sparse x from Ax+e, i.e.,
    \[ ||x^*-x|| \leq C \ ||e|| \]

• Want:
  – Good compression (small m=m(k,n))
  – Efficient algorithms for encoding and recovery

\[
\begin{pmatrix}
A \\
\end{pmatrix}\begin{pmatrix}
x \\
\end{pmatrix} = \begin{pmatrix}
Ax \\
\end{pmatrix}
\]
Bounds

• Want:
  – Good compression (small $m=m(k,n)$)
    
    $m=O(k \log (n/k))$  [Candes-Romberg-Tao’04,....]
  – Efficient algorithms for encoding and recovery

    L1 minimization, CoSaMP, IHT, SMP, ....

• Issue?
  – $\log (n/k)$ penalty compared to non-linear compression
  – Unavoidable in general [Donoho’04, Do Ba-Indyk-Price-Woodruff’10]
Structured sparsity

• Some signals contain more structure than mere sparsity
• Less sparsity patterns to worry about
Model-based compressive sensing
[Baraniuk-Cevher-Duarte-Hegde’10]

• Idea: structure ⇔ restricted support
• Definition:
  A **structured sparsity model** $M$ is defined by a set of allowed supports $M = \{\Omega_1, \ldots, \Omega_{ap}\}$ where $\Omega_i \subseteq [n]$: 

  $$M = \{x \in \mathbb{R}^n \mid \exists \Omega_i \in \Omega : \text{supp}(x) \subseteq \Omega_i\}$$

• Only $O(k + \log |M|)$ measurements suffice if $|\Omega_i| \leq k$
• For all models considered in this talk $|M| = 2^{O(k)}$, so $m = O(k)$
Model specs

• Recovery algorithm depends on the model $M$
• Need **model-projection oracle**
  
  $$M(x) = \arg\min_{x' \in M} \|x' - x\|_2$$

• Requirements:
  – The oracle should run **fast** (e.g. in polynomial or linear time)
  – The oracle must find the **best** approximation in the model
• **Model-IHT:** iterate
  
  $$x^i \leftarrow M(x^{i-1} + A^T(y - Ax^{i-1}))$$

• **Summary:** If $|M| = 2^{O(k)}$ and the oracle $M()$ available, then Model-IHT recovers any $x$ s.t. $\text{supp}(x) \subseteq M$ from $O(k)$ measurements $Ax$
  – Stable generalizations exists as well
  – Theoretical and empirical improvement
Why not approximations?

• E.g., why not an approximate model-projection: given $x$ find $x' \in \mathcal{M}$ s.t.:

$$||x-x'||_2 \leq c \min_{x'' \in \mathcal{M}} ||x''-x||_2$$

• Turns out MIHT might not work if $c>1$!

  – While iterating $x^i \leftarrow \mathcal{M}(x^{i-1} + A^T(y - Ax^{i-1}))$ we can keep $x^i=0$ even though the optimal solution is 1-sparse
Our framework [HIS14]

• **Approximation-Tolerant** Model-Based Compressive Sensing

• Intuition:
  - Consider oracle $M(x) = \arg\min_{\Omega \in \mathbb{M}} \| x - x_\Omega \|_2$
    • **Minimizes** the norm of the “tail”
  - Equivalently: $M(x) = \arg\max_{\Omega \in \mathbb{M}} \| x_\Omega \|_2$
    • **Maximizes** the norm of the “head”
  - However, these two problems are **not** equivalent if approximation is allowed
  - Our approach: required **two** separate approximate head and tail oracles...
  - …and then things work 😊
Our framework ctd.

- **Tail-approximation oracle** $T(x, p)$
  - $T(x, p) = x_\Omega$ for some $\Omega \in M$
  - Tail approximation:
    \[
    || x - T(x, p) ||_2 \leq c_T \min_{\Omega \in M} || x - x_\Omega ||_2
    \]

- **Head-approximation oracle** $H(x, p)$
  - $H(x, p) = x_\Omega$ for some $\Omega \in M$
  - Head approximation:
    \[
    || H(x, p) ||_2 \geq c_H \max_{\Omega \in M} || x_\Omega ||_2
    \]

- **Approximation-Tolerant MIHT**:
  \[
  x^i \leftarrow T( x^{i-1} + H( A^T ( y - Ax^{i-1}) ) )
  \]
Approximation-Tolerant M-IHT

**Theorem [HIS14]:** The iterates of AM-IHT satisfy

$$\|x - x_{i+1}\|_2 \leq (1 + c_T) \left( \delta + \sqrt{1 - (c_H (1 - \delta) - \delta)^2} \right) \|x - x_i\|_2$$

if $A$ satisfies the model-RIP with constant $\delta$ and $c_T$ and $c_H$ are as on the previous slide.
## Prior work

<table>
<thead>
<tr>
<th></th>
<th>Oracle</th>
<th>Stable Recovery</th>
<th>Assumptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blumensath</td>
<td>Additive</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>Kyrillidis-Cevher</td>
<td>Head</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>Giryes-Elad</td>
<td>Tail</td>
<td>Yes</td>
<td>Bound on singular values of $A$</td>
</tr>
<tr>
<td>Davenport-Needell-Wakin</td>
<td>Head, Tail</td>
<td>Yes</td>
<td>Portion of optimal support identified</td>
</tr>
<tr>
<td>This work</td>
<td>Head, Tail</td>
<td>Yes</td>
<td>None</td>
</tr>
</tbody>
</table>

[Giryes, Needell]: similar approach, somewhat different results
OK, but what does it lead to?

- **Constrained Earth-Mover Distance** [SAMPTA’13, SODA’14]
  - Columns with similar sparsity patterns
  - Polytime recovery from $O(k)$ measurements

- **Tree-sparsity** [ISIT’14, ICALP’14]
  - Supports form connected tree components
  - Near-linear time recovery from $O(k)$ measurements

- **Clustered sparsity** [???’15]
  - Supports form connected components in graphs
  - Near-linear time recovery from $O(k)$ measurements
Tree-Sparsity
Tree-sparsity

- Sparse signals whose large coefficients can be arranged in the form of a rooted, connected tree

Applications: data streams, imaging, genomics, ...

Piecewise const. signals

Natural images
An optimization problem

• Given a signal $x$, compute the optimal (exact) **tree-sparse projection** of $x$, i.e., solve

$$\min_{|\Omega| \leq k, \Omega \text{ tree}} \| x - x_\Omega \|_2^2$$
An optimization problem

- Given a signal $x$, compute the optimal (exact) tree-sparse projection of $x$, i.e., solve

$$\min_{|\Omega| \leq k, \Omega \text{ tree}} ||x-x_\Omega||_2^2$$
### Summary of results

<table>
<thead>
<tr>
<th></th>
<th>Runtime</th>
<th>Guarantee</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baraniuk-Jones ‘94</td>
<td>$O(n \log n)$</td>
<td>?</td>
</tr>
<tr>
<td>Donoho ‘97</td>
<td>$O(n)$</td>
<td>?</td>
</tr>
<tr>
<td>Bohanec-Bratko ‘94</td>
<td>$O(n^2)$</td>
<td>Exact</td>
</tr>
<tr>
<td>Cartis-Thompson ‘13</td>
<td>$O(nk)$</td>
<td>Exact</td>
</tr>
<tr>
<td>This work</td>
<td>$O(n \log n)$</td>
<td>Approximate Head</td>
</tr>
<tr>
<td>This work</td>
<td>$O(n \log n + k \log^2 n)$</td>
<td>Approximate Tail</td>
</tr>
</tbody>
</table>

Implication: stable recovery of tree-sparse signals from $O(k)$ measurements in time

#iterations * ($n \log n + k \log^2 n + \text{matrix-vector-mult-time}$)
Why nlogn is better than nk

• Consider a ‘moderate’ problem size
  – e.g. $n = 10^6$, $k = 5\%$ of $n$

• Then, $nk \sim 50 \times 10^9$ while nlog $n \sim 20 \times 10^6$

• Really need near-linear time
Tail Approximation

• We want to approximate:

$$\min_{|\Omega| \leq k, \, \Omega \text{ tree}} \|x - x_\Omega\|_2^2$$

• Perform a Lagrange relaxation of the sparsity constraint:

$$\min_{\Omega \text{ tree}} \|x - x_\Omega\|_2^2 + \lambda |\Omega|$$

• Can be solved using a simple dynamic program (DP) using $O(n)$ time, $O(n)$ space
  – Similar approach as [Donoho ’97]
How to choose \( \lambda \) ?

• Pareto analysis: examine the solution curve as a function of

\[
\text{tail}(\lambda) \quad \quad \text{sparsity}(\lambda)
\]

• Via convexity of the curve, we can always obtain
  – Either: \( (\leq 2k, \leq \text{OPT}) \)
  – Or: \( (\leq k, \leq 2\text{OPT}) \)
• Altogether: we get a tail guarantee in time roughly \( O(n \log n) \)
Head approximation

• Want to approximate:
  \[ \max_{|\Omega| \leq k, \Omega \text{ tree}} \|x_\Omega\|_2^2 \]

• Lemma: The *optimal* tree-projection can be always decomposed into *disjoint* pieces of size \(O(\log n)\)

• **Greedy** approach: compute exact projections with sparsity \(O(\log n)\), assemble pieces, merge

• Total running time:
  \(O(k \log^2 n + n\log n)\)
Experiments: 1D signals

\[ n = 1024 \]
\[ k = 40 \]
\[ m = 140 \]

Gaussian measurements

Original signal \( x \) is tree-sparse in the wavelet basis

Both Tree-CS and AM-CoSaMP exactly recover signal
Experiments: 2D images

Experiments:

- \( n = 512 \times 512 \)
- \( k \approx 10,000 \)
- \( m \approx 35,000 \)

Fourier measurements

Image by AniRaptor2001
Experiments: Monte Carlo

Test signals:
piecewise polynomial
(tree-sparse in the wavelet basis)

Trials randomized over choice of measurement matrix

Successful recovery:
Error is within 5% of the norm of the original signal
Experiments: Speed

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Exact</th>
<th>TailApprox</th>
<th>2 Matlab FFTs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Runtime</td>
<td>13.54 sec</td>
<td>0.0161 sec</td>
<td>0.0136 sec</td>
</tr>
</tbody>
</table>

TailApprox offers **840x speedup** over the exact-tree projection method

(In fact, runtime comparable to Matlab’s FFT)

*Code available at:*
http://people.csail.mit.edu/ludwigs/code.html
Conclusions

• Can make model-based compressive sensing approximation tolerant

• This lets us use approximation algorithms tools to make recovery tractable/faster
  – Network flows, Dynamic programming, Greedy, Steiner tree,…

• Open questions:
  – Other models?