Tutorial: Sparse Signal Recovery

Anna C. Gilbert

Department of Mathematics
University of Michigan
(Sparse) Signal recovery problem

signal or
population
length $N$
k important
features

Under-determined linear system: $\Phi x = y$
Given $\Phi$ and $y$, recover information about $x$

Solving LS: $\arg \min ||\Phi x - y||_2$ returns a non-sparse $x$
Two main examples: group testing and compressed sensing

Group testing

Φ binary = pooling design
x binary, 1 → defective, 0 → acceptable
OUTPUT: defective set
success = find defective set exactly
Arithmetic: OR
Example: Combinatorial group testing

Rat dies only 1 week after drinking poisoned wine
Being good (computer) scientists, they do the following:

<table>
<thead>
<tr>
<th></th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
<th>B4</th>
<th>B5</th>
<th>B6</th>
<th>B7</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Unique encoding of each bottle
If bottle 5 were poison...
...after 1 week
Two main examples: group testing and compressed sensing

Compressed sensing

Φ = measurement matrix
x signal, sparse/compressible
OUTPUT: \( \hat{x} \) good approximation to x
success = \( \|x - \hat{x}\|_p \) versus \( \|x - x_k\|_q \)
Arithmetic: \( \mathbb{R} \) or \( \mathbb{C} \)
Biological applications are a combination of these examples

- Sparse (binary?) matrices
- Quantitative (non-Boolean) measurements
- Noisy or missing measurements
- Various signal models
Theoretical design problems: jointly design matrices and algorithms

Design/Characterize $\Phi$ with $m < N$ rows and recovery algorithm s.t. output $\hat{x}$ is close to good approximation of $x$:

$$\|x - \hat{x}\|_p \leq C\|x - x_k\|_q.$$
Algorithmic resources

- **Computational time**: return $\hat{x}$ in time proportional to $N$ or to $m$? (impacts number of items returned)

- **Computational memory**: use working space proportional to $N$ or to $m$?

- **Small-space random constructions**: can we generate rows/cols of $\Phi$ from collections of random vars with limited independence?

- **Explicit constructions**: how do we construct $\Phi$? Randomly or deterministically? Time to calculate each entry?

- **Matrix density**: number of non-zero entries in $\Phi$?
Signal models

- **Adversarial or “for all”** recover all $x$ with guarantee

$$\|x - \hat{x}\|_2 \leq \frac{C}{\sqrt{k}} \|x - x_k\|_1.$$  

When $x$ is compressible, or $\|x - x_k\|_1 \leq \sqrt{k} \|x - x_k\|_2$, obtain better result

$$\|x - \hat{x}\|_2 \leq C \|x - x_k\|_2.$$  

- **Probabilistic or “for each”**: recover all $x$ that satisfy a statistical constraint

  fixed signal $x$, recover with high probability over construction of $\Phi$

  uniform distribution over $k$-sparse signals

  i.e., random signal model
Empirical algorithmic performance

- **Polynomial time algorithms**: typically fewer measurements required for desired accuracy, by constant factor
  - Convex optimization: constrained minimization (e.g., Bregman iteration)
    \[
    \arg\min \|x\|_1 \quad \text{s.t.} \quad \|y - \Phi x\|_2 \leq \delta.
    \]
  - Unconstrained minimization (e.g., \(\ell_1\)-regularized LS)
    \[
    \minimize L(x; \gamma, y) = \frac{1}{2} \|y - \Phi x\|_2^2 + \gamma \|x\|_1.
    \]
  - Greedy algorithms: iterative thresholding, subspace iteration
    \[
    z = \Phi^* y \quad \text{then threshold} \quad z\text{, keep largest } T\text{ entries}
    \]

- **Sublinear time algorithms**: for sparse, binary matrices only (currently)
  Strive to work in space and time proportional to \(m\), specially designed matrices/distributions and algorithms
Sparse matrices: Expander graphs

- Adjacency matrix $\Phi$ of a $d$ regular $(1, \epsilon)$ expander graph
- Graph $G = (X, Y, E)$, $|X| = n$, $|Y| = m$
- For any $S \subset X$, $|S| \leq k$, the neighbor set
  $$|N(S)| \geq (1 - \epsilon)d|S|$$

- Probabilistic construction:
  $$d = O(\log(n/k)/\epsilon), m = O(k \log(n/k)/\epsilon^2)$$

- Deterministic construction:
  $$d = O(2^{O(\log^3(\log(n)/\epsilon))}), m = k/\epsilon 2^{O(\log^3(\log(n)/\epsilon))}$$
Bipartite graph

Adjacency matrix

Measurement matrix (larger example)
A measurement matrix $\Phi$ satisfies RIP($p$, $k$, $\delta$) property if for any $k$-sparse vector $x$,

$$(1 - \delta)\|x\|_p \leq \|\Phi x\|_p \leq (1 + \delta)\|x\|_p.$$ 

Examples:

- RIP(2): Gaussian random matrix, random rows of discrete Fourier matrix
- RIP(1): adjacency matrix of expander
RIP(p) ⇐⇒ expander

Theorem

\((k, \epsilon)\) expansion implies

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

for any k-sparse x. \(d = \) degree of expander.

Theorem

RIP(1) + binary sparse matrix implies \((k, \epsilon)\) expander for

\[\epsilon = \frac{1 - 1/(1 + \delta)}{2 - \sqrt{2}}.\]

Joint work with Berinde, Indyk, Karloff, Strauss
Let $\Phi$ be an RIP(2) matrix and collect measurements $\Phi x = y$. Then LP recovers a good approximation to the top $k$ entries of $x$:

$$\arg \min \|x\|_1 \quad \text{s.t.} \quad \Phi x = y$$

LP returns $\hat{x}$ with $\|x - \hat{x}\|_2 \leq \frac{C}{\sqrt{k}} \|x - x_k\|_1$. And, noise-resilient version too. [Candés, Tao, + Romberg; Donoho]
Theorem
Φ adjacency matrix of $(2k, \epsilon)$ expander. Consider two vectors $x, x^*$ such that $\Phi x = \Phi x^*$ and $\|x^*\|_1 \leq \|x\|_1$. Then

$$\|x - x^*\|_1 \leq \frac{2}{1 - 2\alpha(\epsilon)} \|x - x_k\|_1$$

where $x_k$ is the optimal $k$-term representation for $x$ and $\alpha(\epsilon) = \frac{(2\epsilon)}{(1 - 2\epsilon)}$.

- Guarantees that Linear Program recovers good sparse approximation
- Noise-resilient version too
RIP(1) $\neq$ RIP(2)

- Any binary sparse matrix which satisfies RIP(2) must have $\Omega(k^2)$ rows [Chandar '07]

- Gaussian random matrix $m = O(k \log(n/k))$ (scaled) satisfies RIP(2) but not RIP(1)
  
  $$x^T = (0 \cdots 0 1 0 \cdots 0)$$
  
  $$y^T = (1/k \cdots 1/k 0 \cdots 0)$$
  
  $$\|x\|_1 = \|y\|_1 \text{ but } \|Gx\|_1 \approx \sqrt{k}\|Gy\|_1$$

- For biological applications, it may be useful to use sparse, binary RIP(1)+RIP(2) matrices with many rows.
RIP(1) \implies \text{LP decoding}

- $\ell_1$ uncertainty principle

Lemma
Let $y$ satisfy $\Phi y = 0$. Let $S$ the set of $k$ largest coordinates of $y$. Then

$$\|y_S\|_1 \leq \alpha(\epsilon)\|y\|_1.$$ 

- LP guarantee

Theorem
Consider any two vectors $u, v$ such that for $y = u - v$ we have $Ay = 0$, $\|v\|_1 \leq \|u\|_1$. $S$ set of $k$ largest entries of $u$. Then

$$\|y\|_1 \leq \frac{2}{1 - 2\alpha(\epsilon)}\|u_S^c\|_1.$$
Greedy, iterative algorithms for sparse matrices

• Polynomial time algorithms: several variants, similar to IHT
  [Berinde, Indyk, Ruszic]

• Sublinear time algorithms: several variants, depend on norms, signal models,
  [Gilbert, Li, Porat, Strauss; Porat, Strauss; Ngo, Porat, Rudra; Price, Woodruff]
High Throughput Screening (HTS)

HTS is an essential step in drug discovery (and elsewhere in biology).

- Large chemical libraries screened on a biological target for activity
- Basic \(\{0, 1\}\) type biological assays to find active compounds
- Usually a small number of compounds found
- One-at-a-time screening: automation and miniaturization
- Noisy assays with false positives and negative errors
Current HTS uses one-at-a-time testing scheme (with repeated trials).
Mathematical set-up

- Binary measurement matrix: adjacency matrix of unbalanced expander graph
- Appropriate linear biochemical model
- Decoding via linear programming

Berinde et. al. (2008) – Combining geometry and combinatorics: A unified approach to sparse signal recovery
Potential Advantages

- Pooled testing of compounds
- Uses fewer tests
- Work moved from testing (costly) to computational analysis (cheap)
- Handles errors in testing better due to built-in replication
- **Additional quantitative information**
To use compressed sensing approach we need a linear model.
Experimental set-up

- Deterministic binary matrix = Shifted Transversal Design (STD)
- Time-resolved FRET screen to identify inhibitors of RGS4C
- Singleplex: 316 compounds screened twice
- Multiplex: 316 compounds screened 4 per well using 316 pooled wells
Experimental results

- Multiplex results match Singleplex 1 very closely
- Decoding in the presence of large errors
Theoretical → Practical challenges in HTS applications

- Accuracy of output *not* speed of algorithm
- Joint matrix and algorithm design, maximize recovered information within *budget*
- Incorporate prior information
- Incorporate noise
- Incorporate measurement device configuration