Combining geometry and combinatorics

A unified approach to sparse signal recovery

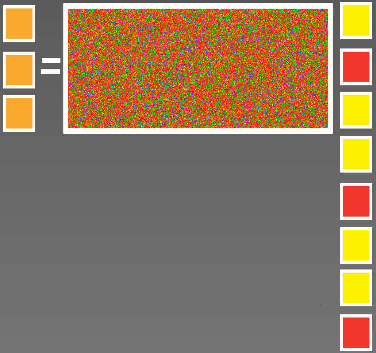
Martin J. Strauss

University of Michigan

joint work with R. Berinde (MIT), A. C. Gilbert (Univ. of Michigan), P. Indyk (MIT), H. Karloff (AT&T)

Sparse signal recovery

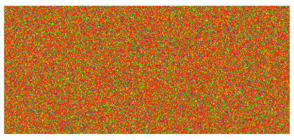
measurements:
length $m = k \log(n)$



k-sparse signal
length n

Problem statement

m as small as possible



Assume x has low complexity: x is k-sparse (with noise)

Construct matrix $A\colon \mathbb{R}^n \to \mathbb{R}^m$

Given Ax for any signal $x \in \mathbb{R}^n$, we can quickly recover $\widehat{x}$ with

$$\|x - \widehat{x}\|_p \leq C \min_{y\ k\text{-sparse}} \|x - y\|_q$$

Applications to Networks

E.g., x is indexed by IP (src,dst) and $x_{i,j}$ is number of bytes sent from i to j... (no new discussion)
Linear measurement paradigm: Measure Ax and Ay separately (time, space, ...). Get $A(x \pm cy)$.
Geometric aging: $A(x^{(0)} + \theta x^{(-1)} + \theta^2 x^{(-2)} + \cdots)$, computed as $y \leftarrow Ax^{\text{first}}$; Repeat: $y \leftarrow \theta y + Ax^{\text{latest}}$.
Sparse A: Only $\approx 1/k$ entries non-zero.

Parameters

Number of measurements m

Recovery time

Approximation guarantee (norms, mixed)

One matrix vs. distribution over matrices

Explicit construction

Universal matrix (for any basis, after measuring)

Tolerance to measurement noise

Two approaches

Geometric [Donoho '04],[Candes-Tao '04, '06],[Candes-Romberg-Tao '05], [Rudelson-Vershynin '06], [Cohen-Dahmen-DeVore '06], and many others...

Dense recovery matrices (e.g., Gaussian, Fourier)
Geometric recovery methods (ℓ_1 minimization, LP)

$$\widehat{x} = \mathrm{argmin} \|z\|_1 \text{ s.t. } \Phi z = \Phi x$$

Uniform guarantee: one matrix A that works for all x

Combinatorial [Gilbert-Guha-Indyk-Kotidis-Muthukrishnan-Strauss '02], [Charikar-Chen-FarachColton '02] [Cormode-Muthukrishnan '04], [Gilbert-Strauss-Tropp-Vershynin '06, '07]

Sparse random matrices (typically)
Combinatorial recovery methods or weak, greedy algorithms
Per-instance guarantees, later uniform guarantees

Paper	A/E	Sketch length	Encode time	Column sparsity/ Update time	Decode time	Approx. error	Noise
[CCFC02, CM06]	E	$k\log^c n$	$n\log^c n$	$\log^c n$	$k\log^c n$	$\ell_2 \le C\ell_2$	
	E	$k\log n$	$n\log n$	$\log n$	$n\log n$	$\ell_2 \le C\ell_2$	
[CM04]	E	$k\log^c n$	$n\log^c n$	$\log^c n$	$k\log^c n$	$\ell_1 \le C\ell_1$	
	E	$k\log n$	$n\log n$	$\log n$	$n\log n$	$\ell_1 \le C\ell_1$	
[CRT06]	A	$k\log(n/k)$	$nk\log(n/k)$	$k\log(n/k)$	LP	$\ell_2 \le \frac{C}{k^{1/2}}\ell_1$	Y
	A	$k\log^c n$	$n\log n$	$k\log^c n$	LP	$\ell_2 \le \frac{C}{k^{1/2}}\ell_1$	Y
[GSTV06]	A	$k\log^c n$	$n\log^c n$	$\log^c n$	$k\log^c n$	$\ell_1 \le C\log n\,\ell_1$	Y
[GSTV07]	A	$k\log^c n$	$n\log^c n$	$\log^c n$	$k^2\log^c n$	$\ell_2 \le \frac{C}{k^{1/2}}\ell_1$	
[NV07]	A	$k\log(n/k)$	$nk\log(n/k)$	$k\log(n/k)$	$nk^2\log^c n$	$\ell_2 \le \frac{C(\log n)^{1/2}}{k^{1/2}}\ell_1$	Y
	A	$k\log^c n$	$n\log n$	$k\log^c n$	$nk^2\log^c n$	$\ell_2 \le \frac{C(\log n)^{1/2}}{k^{1/2}}\ell_1$	Y
[GLR08] (k "large")	A	$k(\log n)^{c\log\log\log n}$	kn^{1-a}	n^{1-a}	LP	$\ell_2 \le \frac{C}{k^{1/2}}\ell_1$	
This paper	A	$k\log(n/k)$	$n\log(n/k)$	$\log(n/k)$	LP	$\ell_1 \le C\ell_1$	Y

Prior work: summary

Paper	A/E	Sketch length	Encode time	Update time	Decode time	Approx. error	Noise
[DM08]	A	$k\log(n/k)$	$nk\log(n/k)$	$k\log(n/k)$	$nk\log(n/k)\log D$	$\ell_2 \le \frac{C}{k^{1/2}}\ell_1$	Y
[NT08]	A	$k\log(n/k)$	$nk\log(n/k)$	$k\log(n/k)$	$nk\log(n/k)\log D$	$\ell_2 \le \frac{C}{k^{1/2}}\ell_1$	Y
	A	$k\log^c n$	$n\log n$	$k\log^c n$	$n\log n\log D$	$\ell_2 \le \frac{C}{k^{1/2}}\ell_1$	Y
[IR08]	A	$k\log(n/k)$	$n\log(n/k)$	$\log(n/k)$	$n\log(n/k)$	$\ell_1 \le C\ell_1$	Y

Recent results: breaking news

Unify these techniques

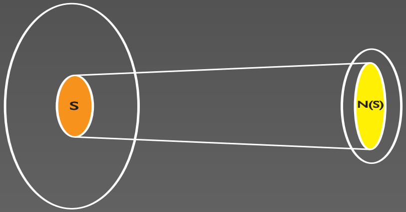
Achieve "best of both worlds"

- LP decoding using sparse matrices
- combinatorial decoding (with augmented matrices)

Deterministic (explicit) constructions

What do combinatorial and geometric approaches share?
What makes them work?

Sparse matrices: Expander graphs



Adjacency matrix A 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

$$|N(S)| \geq (1 - \epsilon) d |S|$$

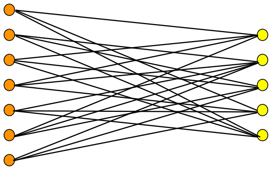
Probabilistic construction:

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

Deterministic construction:

$$d = O(2^{O(\log^3(\log(n)/\epsilon))}), m = k/\epsilon \, 2^{O(\log^3(\log(n)/\epsilon))}$$

Bipartite graph

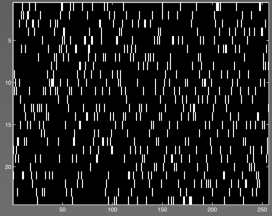


Adjacency matrix

1	0	1	1	0	1	1	0
0	1	0	1	1	1	0	1
0	1	1	0	0	1	1	1
1	0	0	1	1	0	0	1
1	1	1	0	1	0	1	0

Measurement matrix

(larger example)



RIP(p)

A measurement matrix A satisfies RIP(p, k, δ) property if for any k-sparse vector x,

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

RIP(p) $\iff$ expander

Theorem
(k, ϵ) expansion implies

$$(1-2\epsilon)d\|x\|_1 \leq \|Ax\|_1 \leq d\|x\|_1$$

for any k-sparse x. Get RIP(p) for $1 \leq p \leq 1 + 1/\log n$.

Theorem
RIP(1) + binary sparse matrix implies (k, ϵ) expander for

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

Expansion $\implies$ LP decoding

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 $\alpha(\epsilon) = (2\epsilon)/(1 - 2\epsilon)$.

- Guarantees that Linear Program recovers good sparse approximation
- Robust to noisy measurements too

Augmented expander $\Longrightarrow$ Combinatorial decoding

$$\boldsymbol{B}_1\boldsymbol{s} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \color{blue}{1} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \begin{array}{l} \text{MSB} \\ \\ \text{LSB} \end{array}$$

bit-test matrix $\cdot$ signal $=$ location in binary

Theorem
Ψ is $(k, 1/8)$-expander. $\Phi = \Psi \otimes_{\mathrm{r}} B_1$ with $m \log n$ rows. Then, for any k-sparse x, given Φx, we can recover x in time $O(m \log^2 n)$.

With additional hash matrix and polylog(n) more rows in structured matrices, can approximately recover all x in time $O(k^2 \log^{O(1)} n)$ with same error guarantees as LP decoding.

Expander central element in [Indyk '08], [Gilbert-Strauss-Tropp-Vershynin '06, '07]

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)

$$x^T = \begin{pmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix}$$

$$y^T = \begin{pmatrix} 1/k & \cdots & 1/k & 0 & \cdots & 0 \end{pmatrix}$$

$$\|x\|_1 = \|y\|_1 \qquad \text{but} \qquad \|Gx\|_1 \approx \sqrt{k}\|Gy\|_1$$

Expansion $\implies$ RIP(1)

Theorem

(k, ϵ) expansion implies

$$(1 - 2\epsilon)d\|x\|_1 \leq \|Ax\|_1 \leq d\|x\|_1$$

for any k-sparse x.

Proof.

Take any k-sparse x. Let S be the support of x.

Upper bound: $\|Ax\|_1 \leq d\|x\|_1$ for any x

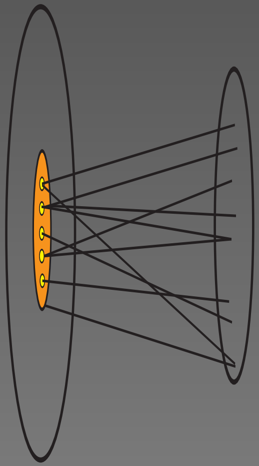
Lower bound:

most right neighbors unique
if all neighbors unique, would have

$$\|Ax\|_1 = d\|x\|_1$$

can make argument robust

Generalization to RIP(p) similar but upper bound not trivial.



□

RIP(1) $\implies$ LP decoding

ℓ_1 uncertainty principle

Lemma

Let y satisfy $Ay = 0$. 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

$$\|y\|_1 \leq \frac{2}{1 - 2\alpha(\epsilon)}\|u_{S^c}\|_1.$$

ℓ_1 uncertainty principle

Proof.

(Sketch): Let $S_0 = S$, S_1, ... be coordinate sets of size k in decreasing order of magnitudes

$$A' = A \text{ restricted to } N(S).$$

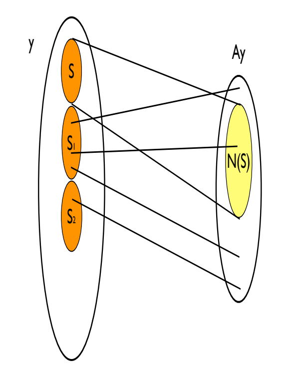
On the one hand

$$\|A' y_S\|_1 = \|A y_S\|_1 \geq (1 - 2\epsilon) d \|y\|_1.$$

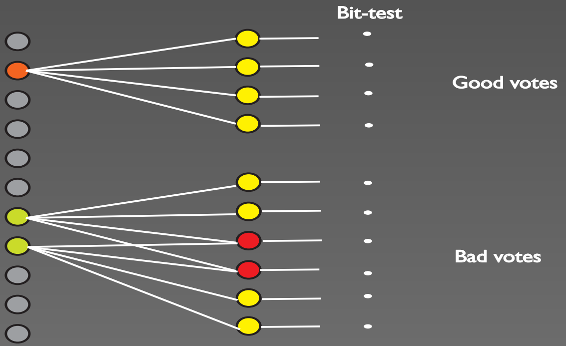
On the other

$$\begin{aligned}
0 = \|A' y\|_1 &= \|A' y_S\|_1 - \sum_{l \geq 1} \sum_{(i,j) \in E[S_l : N(S)]} |y_i| \\
&\geq (1 - 2\epsilon) d \|y_S\|_1 - \sum_{l} |E[S_l : N(S)]| 1/k \|y_{S_{l-1}}\|_1 \\
&\geq (1 - 2\epsilon) d \|y_S\|_1 - 2\epsilon d k \sum_{l \geq 1} 1/k \|y_{S_{l-1}}\|_1 \\
&\geq (1 - 2\epsilon) d \|y_S\|_1 - 2\epsilon d \|y\|_1
\end{aligned}$$

□



Combinatorial decoding

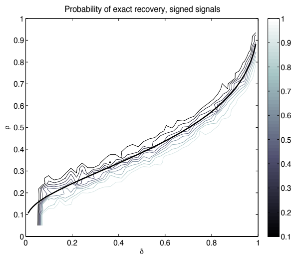


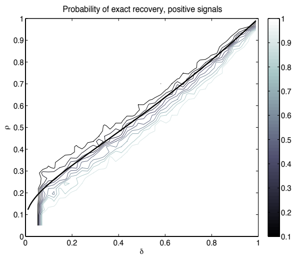
Retain $\{\texttt{index},\ \texttt{val}\}$ if have $> d/2$ votes for $\texttt{index}$

$d/2 + d/2 + d/2 = 3d/2$ violates expander $\implies$ each set of $d/2$ incorrect votes gives at most 2 incorrect indices

Decrease incorrect indices by factor 2 each iteration

Empirical results

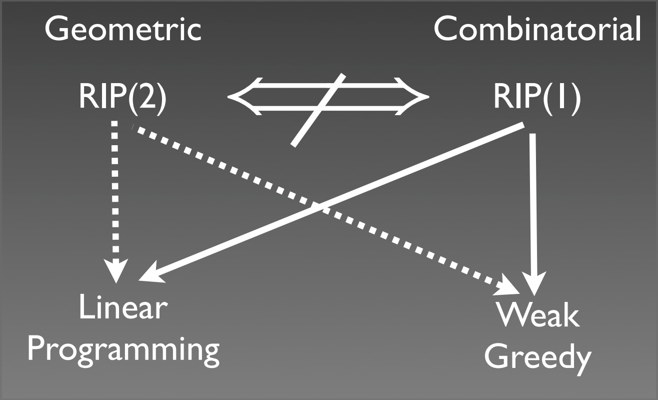




Performance comparable to dense LP decoding

Image reconstruction (TV/LP wavelets), running times, error bounds available in [Berinde, Indyk '08]

Summary: Structural Results



More specifically,