Combining geometry and combinatorics A unified approach to sparse signal recovery

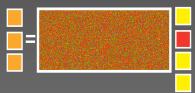
Anna C. Gilbert

University of Michigan

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

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 \hat{x} with

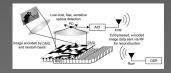
$$\|x - \widehat{x}\|_{p} \le C \min_{\substack{y \ k-sparse}} \|x - y\|_{q}$$

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

Applications

Data stream algorithms $x_i =$ number of items with index *i* can maintain Ax under increments to *x* recover approximation to *x* Efficient data sensing digital/analog cameras analog-to-digital converters Error-correcting codes code { $y \in \mathbb{R}^n | Ay = 0$ } x = error vector, Ax = syndrome



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)

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

Uniform guarantee: one matrix A that works for all x

Combinatorial (Giber: Gula Indy:-Kotelis-Muthalachian-Strauss: 02); [Charlia: Charl FarachColton: 02] [Cornect: Muthalachian: 04].

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 E	$k \log^{c} n$ $k \log n$	$n \log^{c} n$ $n \log n$	$\log^c n$ $\log n$	$k \log^{c} n$ $n \log n$	$\ell_2 \le C\ell_2$ $\ell_2 \le C\ell_2$	
[CM04]	E E	$k \log^{c} n$ $k \log n$	$n \log^c n$ $n \log n$	$\log^c n$ $\log n$	$k \log^{\circ} n$ $n \log n$	$\ell_1 \le C\ell_1$ $\ell_1 \le C\ell_1$	
[CRT06]	A	$k \log(n/k)$ $k \log^{\circ} n$	$nk \log(n/k)$ $n \log n$	$k \log(n/k)$ $k \log^{c} n$	LP LP	$\ell_2 \le \frac{C}{k^{1/2}} \ell_1$ $\ell_2 \le \frac{C}{k^{1/2}} \ell_1$	X
[GSTV06]	A	$k \log^c n$	$n \log^c n$	$\log^c n$	$k \log^{c} n$	$\ell_1 \leq C \log n \ell_1$	Ŷ
[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) (A)	$k \log(n/k) \\ k \log^e n$	$\frac{nk\log(n/k)}{n\log n}$	$k \log(n/k)$ $k \log^{c} n$	$nk^2 \log^c n$ $nk^2 \log^c n$	$\ell_2 \leq \frac{C(\log n)^{1/2}}{k^{1/2}} \ell_1$ $\ell_2 \leq \frac{C(\log n)^{1/2}}{k^{1/2}} \ell_1$	(Y) Y
[GLR08] (k "large")	A	$k(\log n)^{c\log\log\log n}$	kn^{1-a}	n^{1-a}	LP	$\ell_2 \leq \frac{C}{k^{1/2}}\ell_1$	
 This paper	A	$k \log(n/k)$	$n\log(n/k)$	$\log(n/k)$	LP	$\ell_1 \leq C\ell_1$	(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) (A)	$\frac{k \log(n/k)}{k \log^e n}$	$\frac{nk\log(n/k)}{n\log n}$	$\frac{k \log(n/k)}{k \log^c n}$	$\frac{nk\log(n/k)\log D}{n\log n\log D}$	$\ell_{2} \leq \frac{\frac{C}{k^{1/2}}}{\ell_{2}} \ell_{1}$ $\ell_{2} \leq \frac{\frac{C}{k^{1/2}}}{k^{1/2}} \ell_{1}$	Y Y
[IR08]	A	$k \log(n/k)$	$n\log(n/k)$	$\log(n/k)$	$n\log(n/k)$	$\ell_1 \leq C\ell_1$	Υ

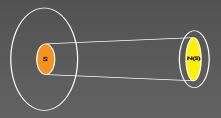
Recent results: breaking news

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| = mFor any $S \subset X$, $|S| \le k$, the neighbor set

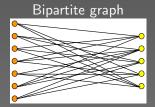
 $|N(S)| \ge (1-\epsilon)d|S|$

Probabilistic construction:

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

Deterministic construction:

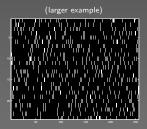
 $d = O(2^{O(\log^3(\log(n)/\epsilon))}), \overline{m = k/\epsilon \, 2^{O(\log^3(\log(n)/\epsilon))}}$



Adjacency matrix

I	0	I	I	0	I	I	0
0	I	0	I	I	I	0	Ι
0	I	I	0	0	I	I	Ι
I	0	0	I	I	0	0	Ι
I	I	I	0	I	0	I	0

Measurement matrix



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

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

$\mathsf{RIP}(\mathsf{p}) \iff \mathsf{expander}$

Theorem (k, ϵ) expansion implies

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

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

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

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

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

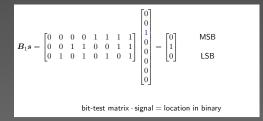
$$\|x - x_*\|_1 \le \frac{2}{1 - 2\alpha(\epsilon)} \|x - x_k\|_1$$

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 \implies Combinatorial decoding



Theorem

 Ψ is (k, 1/8)-expander. $\Phi = \Psi \otimes_r B_1$ 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 [note or] (where Strange Trape Versions (0, 0)]

$\mathsf{RIP}(1) \neq \mathsf{RIP}(2)$

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)

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

 $y^{T} = \begin{pmatrix} 1/k & \cdots & 1/k & 0 & \cdots & 0 \end{pmatrix}$
 $\|x\|_{1} = \|y\|_{1}$ but $\|Gx\|_{1} \approx \sqrt{k} \|Gy\|_{1}$

Expansion \implies RIP(1)

Theorem (k, ϵ) expansion implies

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

for any k-sparse x.

Proof.

Take any *k*-sparse *x*. Let *S* be the support of *x*. Upper bound: $||Ax||_1 \le d||x||_1$ for any *x*

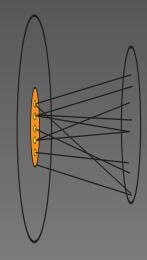
Lower bound:

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

 $\|y_{\mathcal{S}}\|_{1} \leq \alpha(\epsilon) \|y\|_{1}.$

LP guarantee

Theorem

Consider any two vectors u, v such that for y = u - v we have Ay = 0, $||v||_1 \le ||u||_1$. 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 $S_0 = S, S_1, \ldots$ be coordinate sets of size k in decreasing order of magnitudes

$$A' = A$$
 restricted to $N(S)$.

On the one hand

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

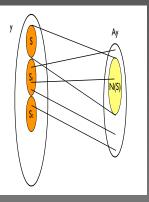
On the other

$$0 = ||A'y||_1 = ||A'y_S||_1 - \sum_{l \ge 1} \sum_{(i,j) \in E[S_l:N(S)]} |y_l|$$

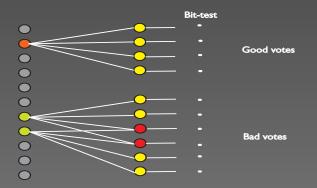
$$\ge (1 - 2\epsilon)d||y_S||_1 - \sum_l |E[S_l:N(S)]|1/k||y_{S_{l-1}}||_1$$

$$\ge (1 - 2\epsilon)d||y_S||_1 - 2\epsilon dk \sum_{l \ge 1} 1/k||y_{S_{l-1}}||_1$$

$$\ge (1 - 2\epsilon)d||y_S||_1 - 2\epsilon d||y||_1$$

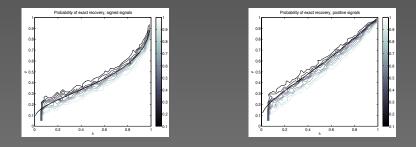


Combinatorial decoding



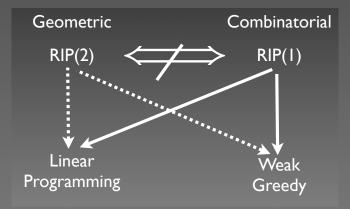
Retain {index, val} if have > d/2 votes for index d/2 + d/2 + d/2 = 3d/2 violates expander \implies each set of d/2 incorrect votes gives at most 2 incorrect indices Decrease incorrect indices by factor 2 each iteration

Empirical results



Performance comparable to dense LP decoding Image reconstruction (TV/LP wavelets), running times, error bounds available in (Bernule, index (D))

Summary: Structural Results



More specifically,

