## 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 $A: \mathbb{R}^{n} \rightarrow \mathbb{R}^{m}$

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

$$
\|x-\widehat{x}\|_{p} \leq C \min _{y k-s p a r s e}\|x-y\|_{q}
$$

## 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 $A x$ under increments to $x$ recover approximation to $x$
Efficient data sensing
digital/analog cameras
analog-to-digital converters


Error-correcting codes
code $\left\{y \in \mathbb{R}^{n} \mid A y=0\right\}$
$x=$ error vector, $A x=$ syndrome

## Two approaches

## Geometric [Donoho '044]. Candes-Tao $\left.{ }^{\circ} 04,{ }^{\prime} 06\right]_{\mid, ~[C a n d e s-R o m b e r g-T a o ~ ' 05], ~}^{\text {a }}$

[Rudelson-Vershynin '06], [Cohen-Dahmen-DeVore '06], and many others.
Dense recovery matrices (e.g., Gaussian, Fourier) Geometric recovery methods ( $\ell_{1}$ 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 [Gilbert-Guha-Indyk-Kotidis-Muthukrishnan-Strauss '02],

[Charikar-Chen-FarachColton '02] [Cormode-Muthukrishnan '04],
[Gibert-Strauss-Tropp-Vershynin '06, '07]
Sparse random matrices (typically)
Combinatorial recovery methods or weak, greedy algorithms
Per-instance guarantees, later uniform guarantees


Prior work: summary

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

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

Deterministic construction:

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

## Measurement matrix

Adjacency matrix

| 1 | 0 | 1 | 1 | 0 | 1 | I | 0 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 0 | I | 0 | 1 | 1 | 1 | 0 | I |
| 0 | I | 1 | 0 | 0 | 1 | I | 1 |
| 1 | 0 | 0 | 1 | 1 | 0 | 0 | I |
| 1 | 1 | 1 | 0 | 1 | 0 | I | 0 |

(larger example)


## RIP(p)

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

$$
(1-\delta)\|x\|_{p} \leq\|A x\|_{p} \leq(1+\delta)\|x\|_{p}
$$

## $\operatorname{RIP}(\mathrm{p}) \Longleftrightarrow$ expander

Theorem
$(k, \epsilon)$ expansion implies

$$
(1-2 \epsilon) d\|x\|_{1} \leq\|A x\|_{1} \leq d\|x\|_{1}
$$

for any $k$-sparse $x . \operatorname{Get} R I P(p)$ for $1 \leq p \leq 1+1 / \log n$.
Theorem
RIP(1) + binary sparse matrix implies $(k, \epsilon)$ expander for

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

## Expansion $\Longrightarrow$ LP decoding

## Theorem

$\Phi$ adjacency matrix of $(2 k, \epsilon)$ expander. Consider two vectors $x, x_{*}$ such that $\Phi x=\Phi x_{*}$ and $\left\|x_{*}\right\|_{1} \leq\|x\|_{1}$. Then

$$
\left\|x-x_{*}\right\|_{1} \leq \frac{2}{1-2 \alpha(\epsilon)}\left\|x-x_{k}\right\|_{1}
$$

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


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

## Theorem

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

With additional hash matrix and polylog( $n$ ) more rows in structured matrices, can approximately recover all $x$ in time $O\left(k^{2} \log ^{O(1)} n\right)$ with same error guarantees as LP decoding.
Expander central element in [lndyk '08], [Gilbert-Strauss- Tropp-Vershynin 06 , ${ }^{\circ}$ or]

## $\operatorname{RIP}(1) \neq \operatorname{RIP}(2)$

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

Gaussian random matrix $m=O(k \log (n / k))$ (scaled) satisfies RIP(2) but not RIP(1)

$$
\begin{gathered}
x^{T}=\left(\begin{array}{lllllll}
0 & \cdots & 0 & 1 & 0 & \cdots & 0
\end{array}\right) \\
y^{T}=\left(\begin{array}{llllll}
1 / k & \cdots & 1 / k & 0 & \cdots & 0
\end{array}\right) \\
\|x\|_{1}=\|y\|_{1} \\
\text { but }
\end{gathered} \quad\|G x\|_{1} \approx \sqrt{k}\|G y\|_{1} .
$$

## Expansion $\Longrightarrow \operatorname{RIP}(1)$

## Theorem

( $k, \epsilon$ ) expansion implies

$$
(1-2 \epsilon) d\|x\|_{1} \leq\|A x\|_{1} \leq d\|x\|_{1}
$$

for any $k$-sparse $x$.

## Proof.

Take any $k$-sparse $x$. Let $S$ be the support of $x$.
Upper bound: $\|A x\|_{1} \leq d\|x\|_{1}$ for any $x$
Lower bound:
most right neighbors unique
if all neighbors unique, would have

$$
\|A x\|_{1}=d\|x\|_{1}
$$

can make argument robust
Generalization to RIP $(\mathrm{p})$ similar but upper bound not trivial.

## RIP $(1) \Longrightarrow$ LP decoding

## $\ell_{1}$ uncertainty principle

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

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

## LP guarantee

Theorem
Consider any two vectors $u, v$ such that for $y=u-v$ we have $A y=0,\|v\|_{1} \leq\|u\|_{1}$. Set of $k$ largest entries of $u$. Then

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

## $\ell_{1}$ uncertainty principle

## Proof.

(Sketch): Let $S_{0}=S, S_{1}, \ldots$ be coordinate sets of size $k$ in decreasing order of magnitudes

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

On the one hand

$$
\left\|A^{\prime} y_{S}\right\|_{1}=\left\|A y_{S}\right\|_{1} \geq(1-2 \epsilon) d\|y\|_{1} .
$$

On the other

$$
\begin{aligned}
0=\left\|A^{\prime} y\right\|_{1} & =\left\|A^{\prime} y_{S}\right\|_{1}-\sum_{I \geq 1} \sum_{(i, j) \in E\left[S_{l}: N(S)\right]}\left|y_{i}\right| \\
& \geq(1-2 \epsilon) d\left\|y_{S}\right\|_{1}-\sum_{l}\left|E\left[S_{l}: N(S)\right]\right| 1 / k\left\|y_{S_{I-1}}\right\|_{1} \\
& \geq(1-2 \epsilon) d\left\|y_{S}\right\|_{1}-2 \epsilon d k \sum_{l \geq 1} 1 / k\left\|y_{S_{I-1}}\right\|_{1} \\
& \geq(1-2 \epsilon) d\left\|y_{S}\right\|_{1}-2 \epsilon d\|y\|_{1}
\end{aligned}
$$



## Combinatorial decoding



## Good votes

Bad votes

Retain $\{$ index, val\} if have $>d / 2$ votes for index $d / 2+d / 2+d / 2=3 d / 2$ violates expander $\Longrightarrow$ 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,


(fast update time, sparse)

