

Combining geometry and combinatorics

A unified approach to sparse signal recovery

Anna C. Gilbert

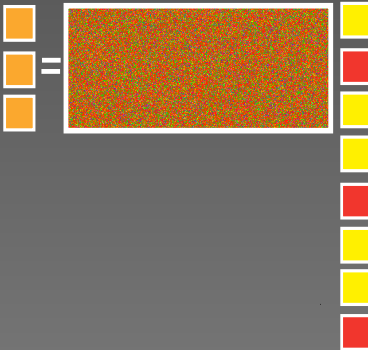
University of Michigan

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

Sparse signal recovery

measurements:

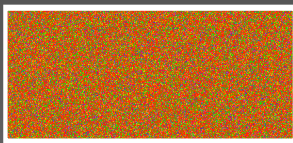
length $m = k \log(n)$



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

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

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

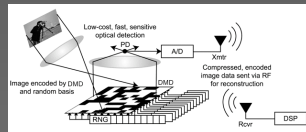
x_i = number of items with index i
can maintain Ax under increments to x
recover approximation to x

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)

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

Uniform guarantee: one matrix A that works for all x

Combinatorial [Gilbert-Guha-Indyk-Kotidis-Muthukrishnan-Strauss '02],
[Charikar-Chen-Farach-Colton '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

Scale: Excellent Good Fair Poor

(Recent) History

Paper	Rand. / Det.	Sketch length	Encode time	Update time	Recovery time	Approx
[CCF'02], [CM'06]	R	$k \log n$	$n \log n$	$\log n$	$n \log n$	I2 / I2
	R	$k \log^c n$	$n \log^c n$	$\log^c n$	$k \log^c n$	I2 / I2
[CM'04]	R	$k \log n$	$n \log n$	$\log n$	$n \log n$	I1 / I1
	R	$k \log^c n$	$n \log^c n$	$\log^c n$	$k \log^c n$	I1 / I1
[CRT'04] [RV'05]	D	$k \log(n/k)$	$nk \log(n/k)$	$k \log(n/k)$	n^c	I2 / I1
	D	$k \log^c n$	$n \log n$	$k \log^c n$	n^c	I2 / I1
[GSTV'06] [GSTV'07]	D	$k \log^c n$	$n \log^c n$	$\log^c n$	$k \log^c n$	I1 / I1
	D	$k \log^c n$	$n \log^c n$	$k \log^c n$	$k^2 \log^c n$	I2 / I1
[BGKS'08]	D	$k \log(n/k)$	$n \log(n/k)$	$\log(n/k)$	n^c	I1 / I1
[GLR'08]	D	$k \log^{\log \log \log n}$	kn^{1-a}	n^{1-a}	n^c	I2 / I1
[NV'07], [DM'08], [NT'08]	D	$k \log(n/k)$	$nk \log(n/k)$	$k \log(n/k)$	$nk \log(n/k) * T$	I2 / I1
	D	$k \log^c n$	$n \log n$	$k \log^c n$	$n \log n * T$	I2 / I1
[IR'08]	D	$k \log(n/k)$	$n \log(n/k)$	$\log(n/k)$	$n \log(n/k)$	I1 / I1

Unify these techniques

Achieve “best of both worlds”

- LP decoding using sparse matrices

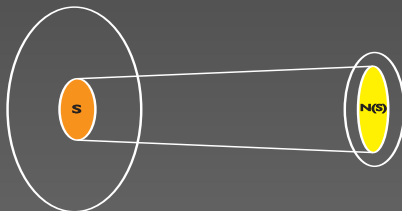
- combinatorial decoding (with augmented matrices)

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:

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

RIP(p)

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

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

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

Theorem

(k, ϵ) expansion implies

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

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

Theorem

$\text{RIP}(1)$ + 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 \implies Combinatorial decoding

$$B_1 s = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \begin{array}{l} \text{MSB} \\ \text{LSB} \end{array}$$

bit-test matrix \cdot signal = location in binary

Theorem

Ψ is $(k, 1/8)$ -expander. $\Phi = \Psi \otimes_{\mathbb{F}_2} B_1$ 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 $\text{polylog}(n)$ 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]

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

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)

$$x^T = (0 \quad \cdots \quad 0 \quad 1 \quad 0 \quad \cdots \quad 0)$$

$$y^T = (1/k \quad \cdots \quad 1/k \quad 0 \quad \cdots \quad 0)$$

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

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:

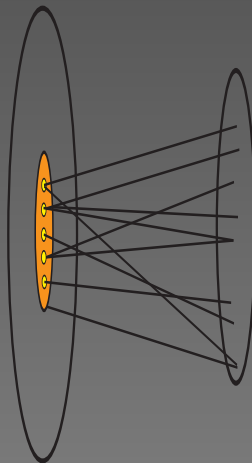
most right neighbors unique
if all neighbors unique, would have

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

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 $S_0 = S, S_1, \dots$ be coordinate sets of size k in decreasing order of magnitudes

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

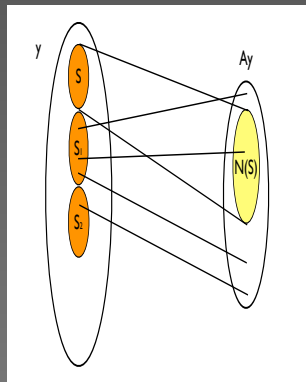
On the one hand

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

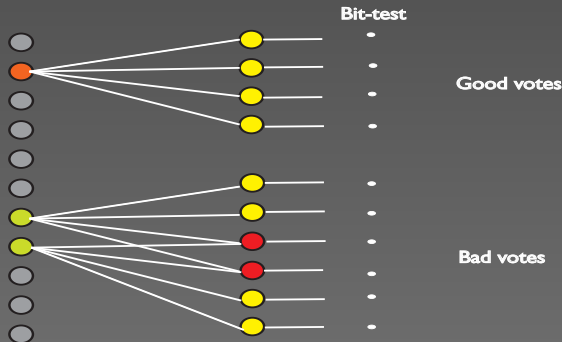
On the other

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

□



Combinatorial decoding

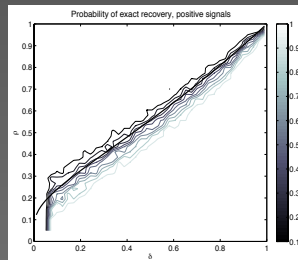
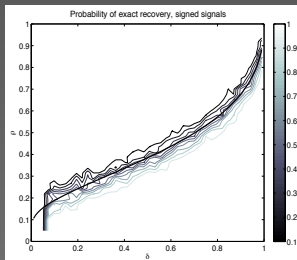


Retain $\{\text{index}, \text{val}\}$ if have $> d/2$ votes for index

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

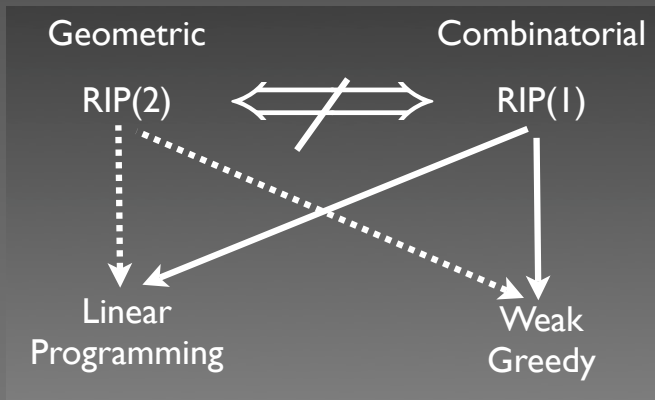
Decrease incorrect indices by factor 2 each iteration

Empirical results

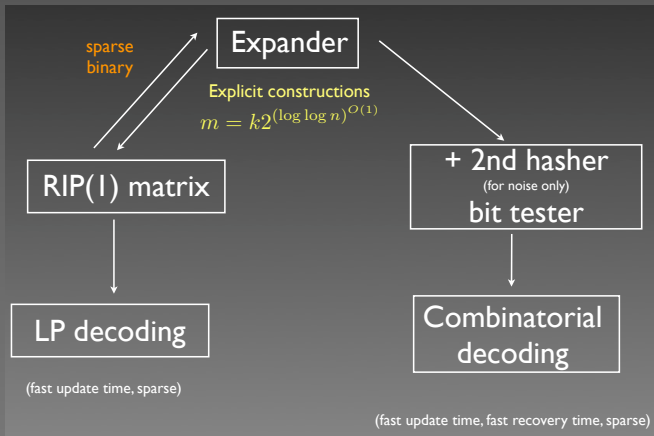


Performance comparable to dense LP decoding
Image reconstruction (TV/LP wavelets), running times, error
bounds available in [\[Berinde, Indyk '08\]](#)

Summary: Structural Results



More specifically,

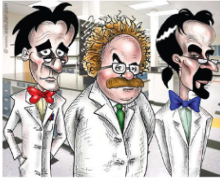


Application: combinatorial group testing

R. Kainkaryam (Univ. of Michigan)
and P. Woolf (Univ. of Michigan)

Application: combinatorial group testing

Scitentist-types



Big Party ~ 1 week



7 unlabelled
bottles



1 bottle

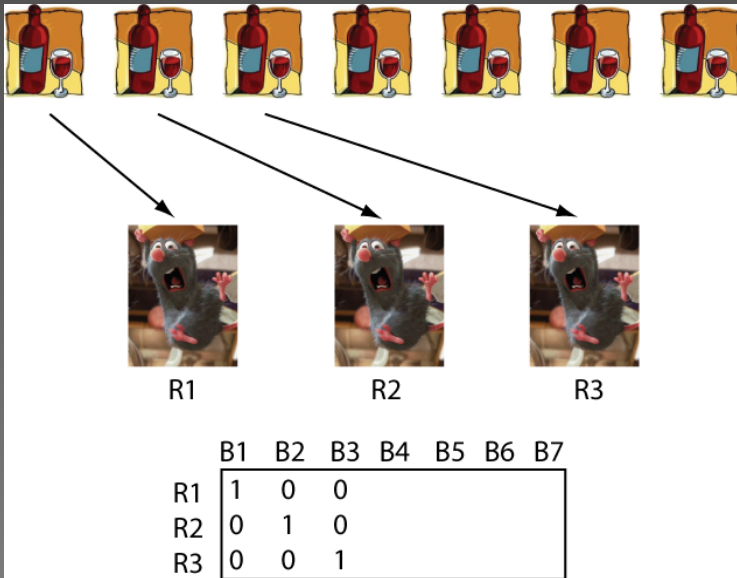


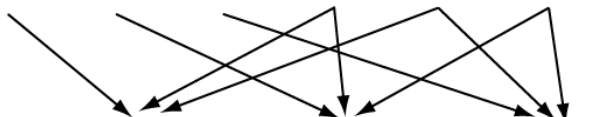
3 Rats



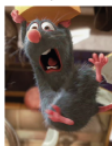
Rat dies only 1 week *after* drinking poisoned wine

Being good (computer) scientists, they do the following:





R1

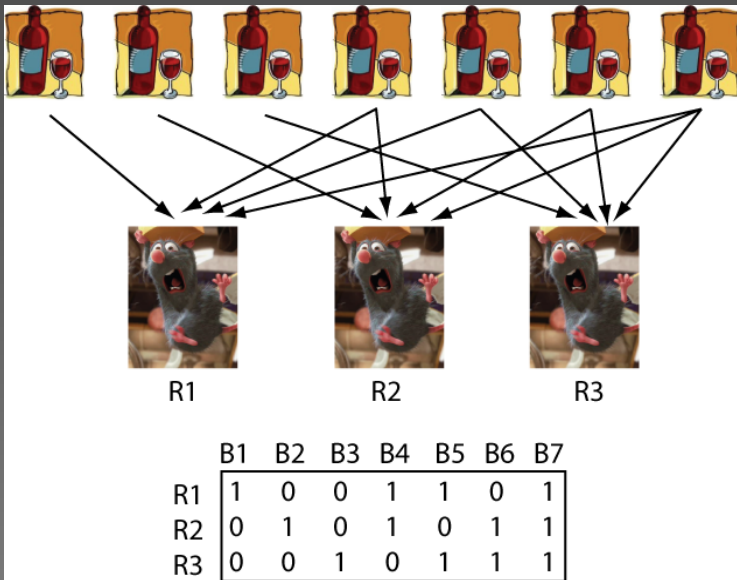


R2



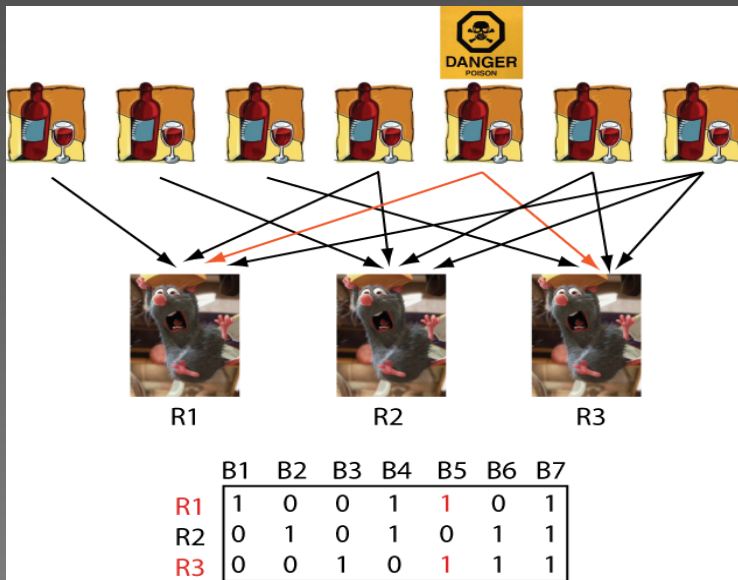
R3

	B1	B2	B3	B4	B5	B6	B7
R1	1	0	0	1	1	0	
R2	0	1	0	1	0	1	
R3	0	0	1	0	1	1	

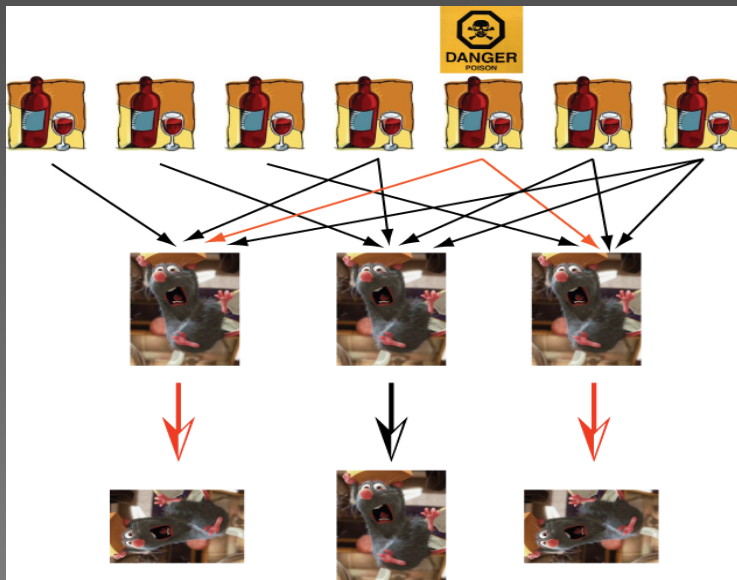


Unique encoding of each bottle

If bottle 5 were poison...

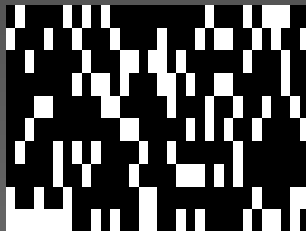


...after 1 week



Problem statement: CGT

m as small
as possible



Assume x has
low complexity:
 x has k -defects
the rest are zero

Construct matrix $A: \mathbb{B}^n \rightarrow \mathbb{B}^m$

Given Ax for any signal $x \in \mathbb{B}^n$, we can quickly recover k defects present in x . Note: arithmetic is boolean and result from pooled test is $\{0, 1\}$.

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

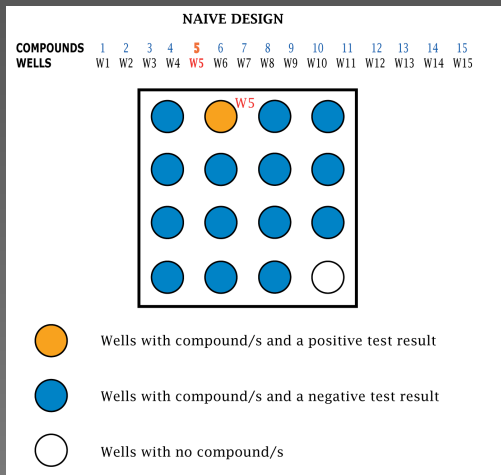
Usually a small number of compounds found

One-at-a-time screening: automation and
miniaturization

Noisy assays with false positives and
negative errors

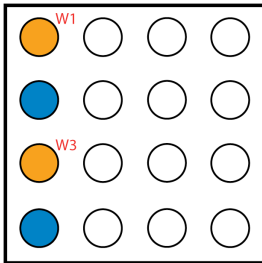


Current HTS uses one-at-a-time testing scheme (with repeated trials).



Pooled HTS design

POOLING DESIGN



WELLS

W1

W2

W3

W4

COMPOUNDS

1	3	5	7	9	11	13	15
2	3	6	7	10	11	14	15
4	5	6	7	12	13	14	15
8	9	10	11	12	13	14	15

Propose using pooled testing of compounds

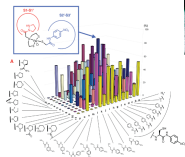
Uses fewer tests

Work moved from testing (costly) to computational analysis (cheap)

Handles errors in testing better due to built-in replication

Additional quantitative information

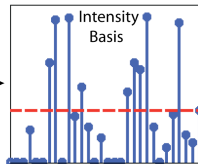
HTS and signal recovery



Million Compounds



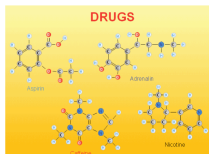
~ 1 Million
Measurements
(N)



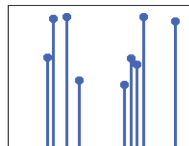
Processing :
Threshold, Z-score,
FDR etc.

Take **K** major
compounds
forward

$N \gg K$



Secondary
Screening



Quantitative analysis of pooling in HTS

Constraints

linearity: measured quantities map linearly to compound activities

sparsity: most compounds inactive

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 & 1 \\ 0 & 1 & \dots & 0 & 1 \\ & \vdots & & \vdots & \\ 1 & 0 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix}$$

Challenges

choosing a good mixing scheme

enforcing a mixing constraint

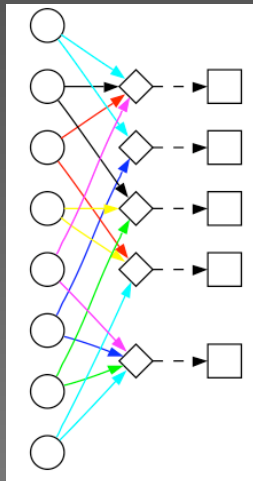
recovery algorithm tolerant to measurement noise + errors

Our approach

Binary measurement matrix: adjacency matrix of unbalanced expander graph

Appropriate linear biochemical model

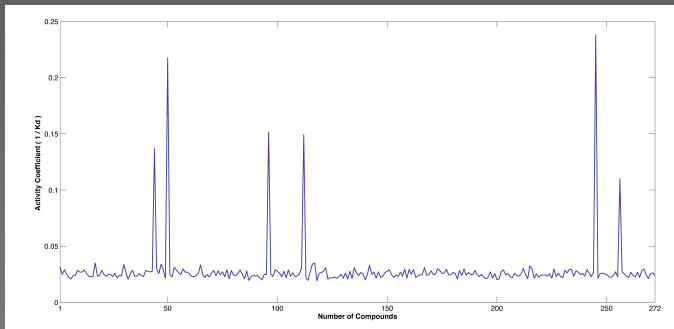
Decoding via linear programming



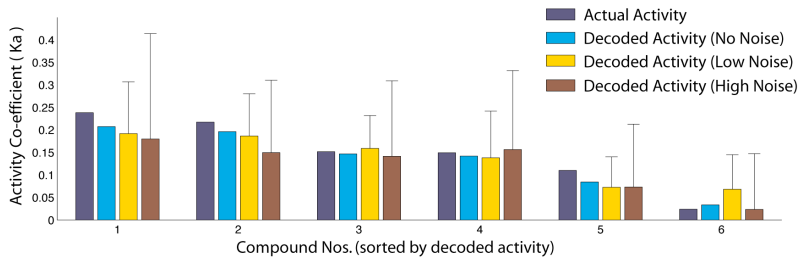
Small library

Synthetic screen: small molecule ligands for formylpeptide receptor, 6 active [Edwards, et al., Nature Protocols '06]

$n = 272$, $k = 6$, using deterministic STD matrix, $m = 116$



In silico

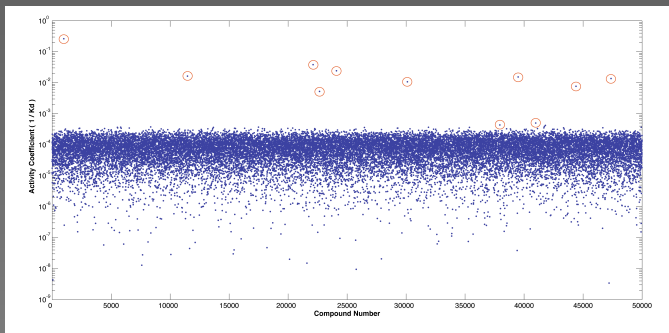


Large library

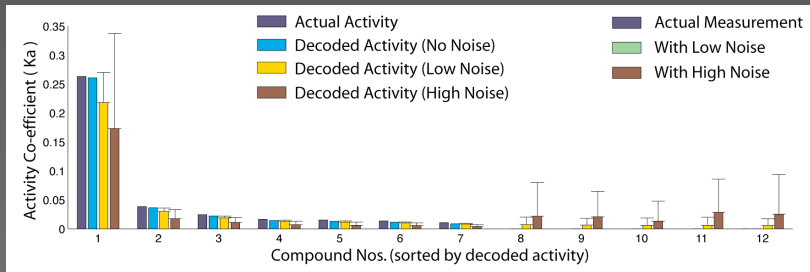
Actual screen: 50,000 compounds screened against *E. coli* dihydrofolate reductase (DHFR), 12 active [McMaster HTS Lab Data

Mining and Docking Competition '05]

$n = 50,000$, $k = 12$ screened in 122 blocks of 410 compounds using STD deterministic matrix, $m = 10,004$



In silico



Current/Future work

Computer Science:

- ℓ_2/ℓ_1 error bounds for sparse matrices?
- optimal number of measurements for Fourier matrix, sublinear algorithms
- refined decoding with noise + missing measurements
- refined error analysis
- decoding algorithms to rank compounds

Chemical Engineering:

- good/best explicit constructions which meet experimental constraints
- refine error analysis, algorithm output for cultural interpretations of biologists
- design and implementation of several in vitro experiments (HTS, differential gene expression)