Sparse Approximation, List Decoding, and Uncertainty Principles

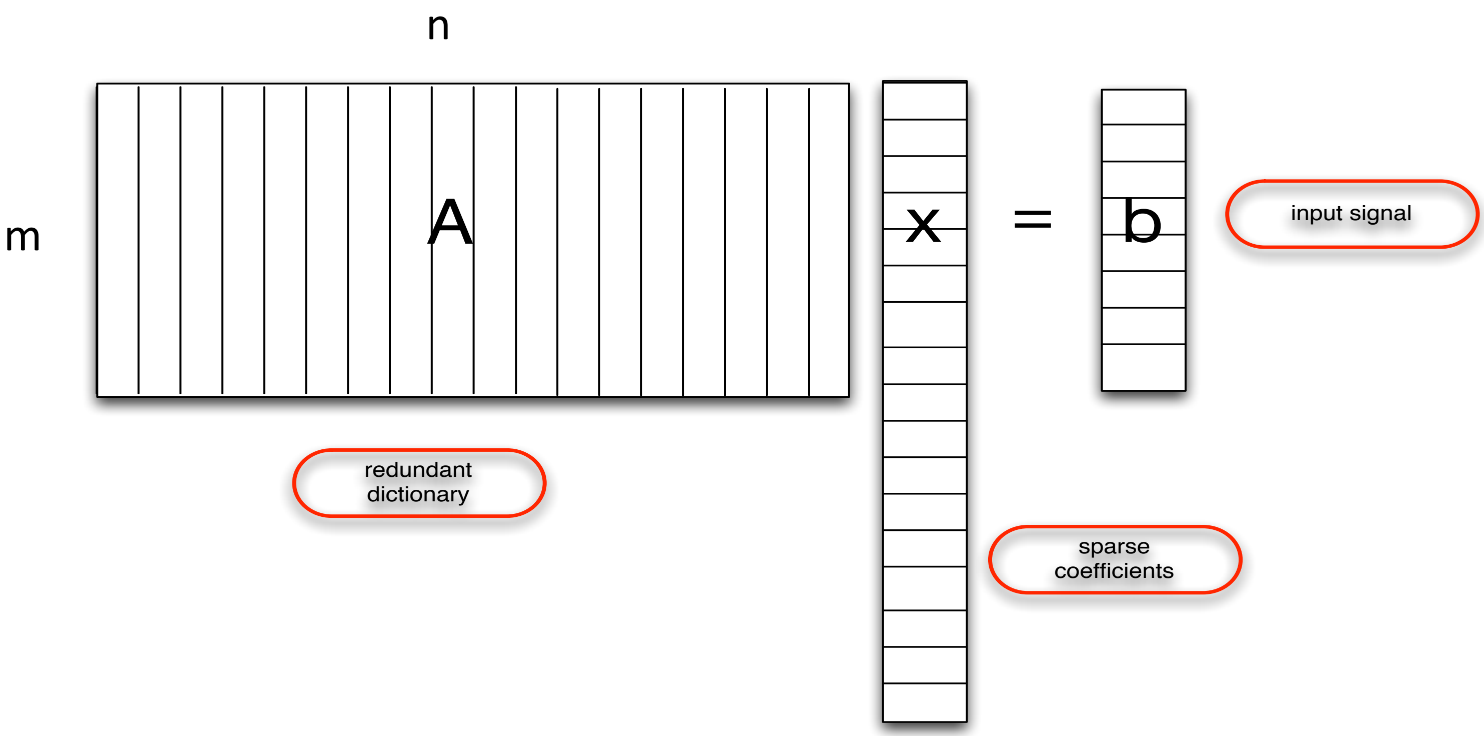
Anna C. Gilbert

Department of Mathematics

University of Michigan

Joint work with Mahmoud Abo Khamis, Hung Q. Ngo,
Atri Rudra (SUNY: University at Buffalo)

Sparse Approximation: Definitions



EXACT: Given *A*,*b*, find sparsest *x* s.t. *Ax* = *b*.

$$\hat{x} = \operatorname{argmin} \|x\|_0 \text{ s.t. } Ax = b$$

SPARSE: Given *A*,*b*,*k*, find best *k*-term approximation for *b*.

$$\hat{x} = \operatorname{argmin} \|Ax - b\|_2 \text{ s.t. } \|x\|_0 \leq k$$

Sparse Approximation: Unique repn. "barriers"

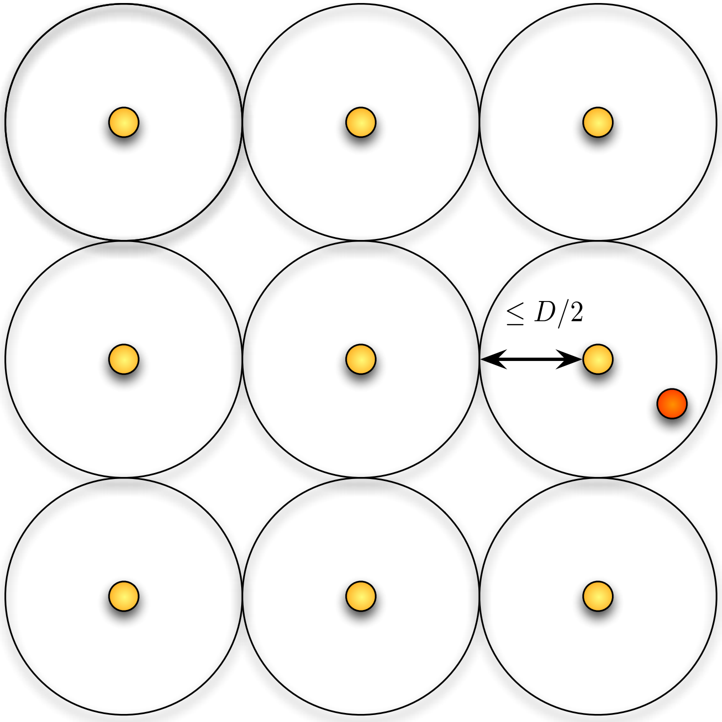
- **Spark(*A*)** = *σ(A)* = min number of linearly dependent cols
- **Unique EXACT** solution: If *k* < *σ/2*, then there's a unique sparsest solution wit[illegible] spa[illegible]ity *k*[illegible] [Donoho, Elad, 2003]

Unique sparse representations !?

- **Coherence(*A*)** = *μ(A)* = [illegible]en cols
- **Unique EXACT solution:** if k < *μ/2*, then there's a unique sparsest solution with sparsity *k*. [Donoho, Elad, 2003]
- *A* = spikes and sines, *μ(A) = 1/√n*
- Bounds = form of **Uncertainty Principle**

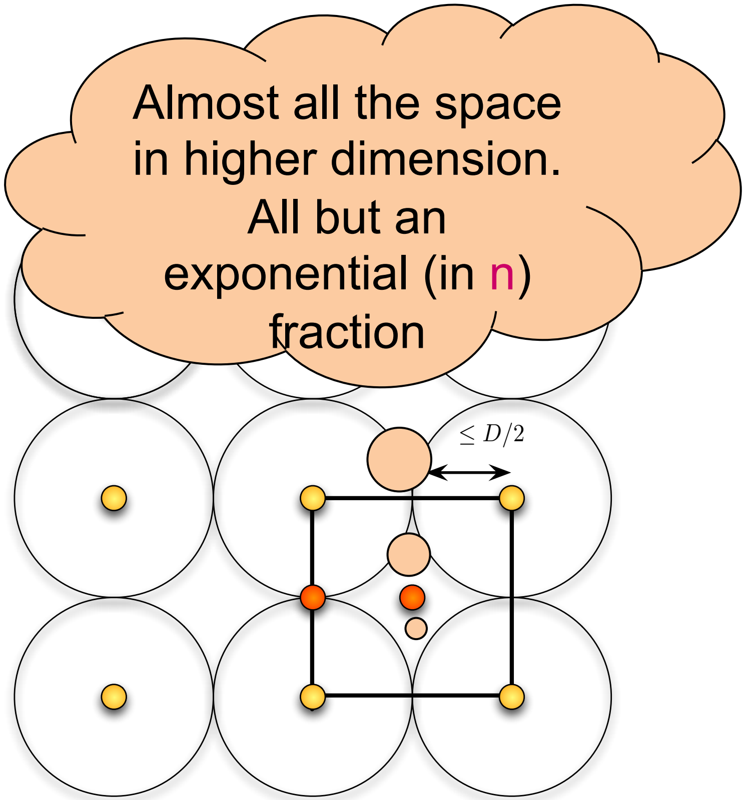
Error Correcting Codes: Unique decoding

- Distance of code = minimum distance between 2 codewords
- Receive corrupted codeword
- Return closest codeword
- Tolerate errors up to *D/2*



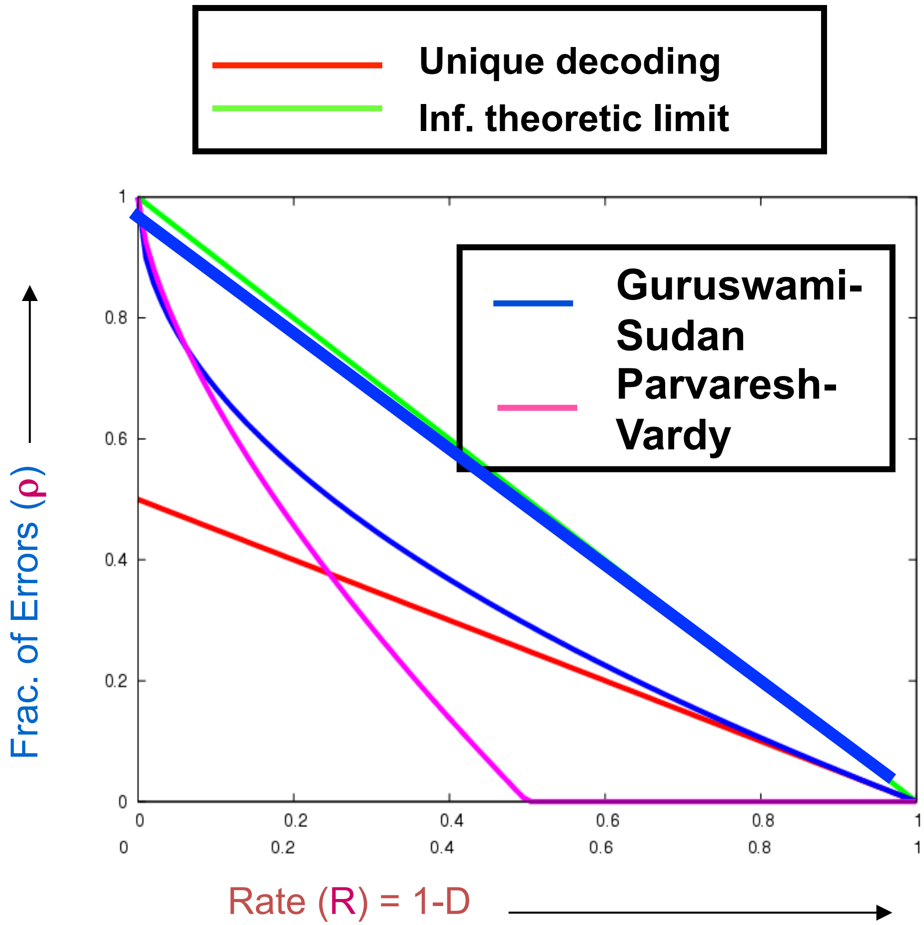
List Decoding ECC: definitions

- Return small list of codewords, with guarantee that transmitted codeword is in the list [Elias, 1957, Wozencraft, 1958]
- Formal defn: ρ = fraction of errors, list words differ from transmitted by no more than ρ



List Decoding ECC: implications

- Information Theory
  - Information theoretic limit $\rho < 1 - R$
  - Explicit constructions, efficient (?) algorithms: Folded RS codes
- Cryptography
  - Cryptanalysis of certain block-ciphers [Jakobsen, 1998]
  - Efficient traitor tracing scheme [Silverberg, Staddon, Walker 2003]
- Complexity Theory
  - Hardcore predicates from one way functions [Goldreich,Levin 1989; Impagliazzo 1997; Ta-Shama, Zuckerman 2001]
  - Worst-case vs. average-case hardness [Cai, Pavan, Sivakumar 1999; Goldreich, Ron, Sudan 1999; Sudan, Trevisan, Vadhan 1999; Impagliazzo, Jaiswal, Kabanets 2006]



List Decoding ECC: progression of results

1. Combinatorial bounds on list size (Johnson bound)
2. Algorithms for finding list
3. Explicit ECCs that achieve bounds + practical algorithms

Sparse Approximation <-> ECC

Sparse Approximation	ECC
Redundant dictionary	Codebook
Input signal	Received codeword + errors
Coherence	Distance
Redundancy	Rate
Spark	Spark
Best k-term approximation	Decoding
k=1	Closest codeword
k > 1	-------

List Sparse Approximation:
Definitions

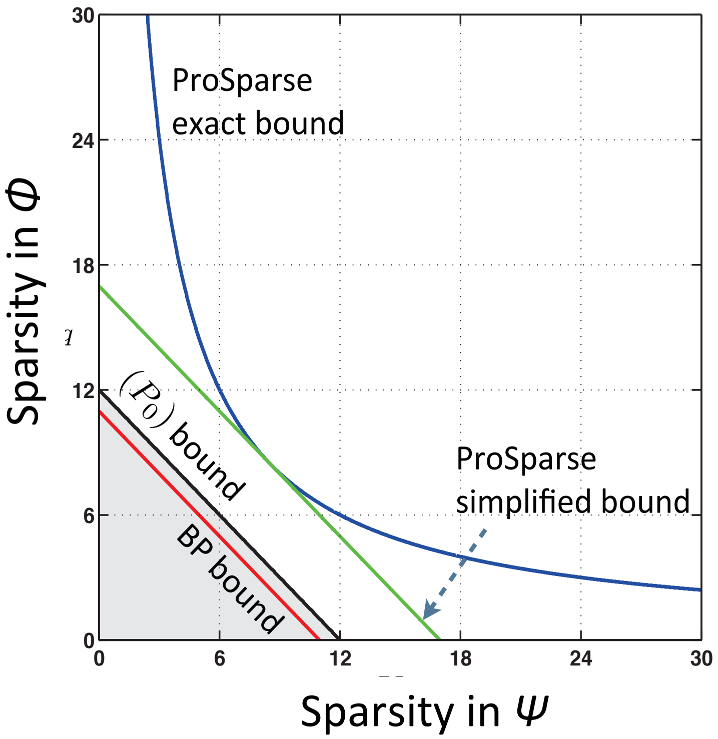
List SPARSE: Given *A,b,* and *k*, list *all k*-sparse *x* such that $\|Ax - b\|_2$ is minimized.

Exact analogy with List Decoding ECCs

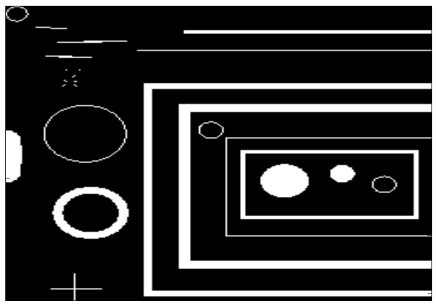
List APPROX: Given *A,b,k*,and *ε*, list *all k*-sparse *x* such that $\|Ax - b\|_2 \leq \epsilon$

List Sparse Approximation: Implications?

- **Algorithms' achievability**
  [Dragotti, Lu 2013]
  - *A = [Ψ, Φ]* union of ONBs (can be generalized)
  - **ProSparse:** proto-list sparse approximation algorithm
  - Returns list of *exact* representations "beyond" convex relaxation bound and unique repn. bound

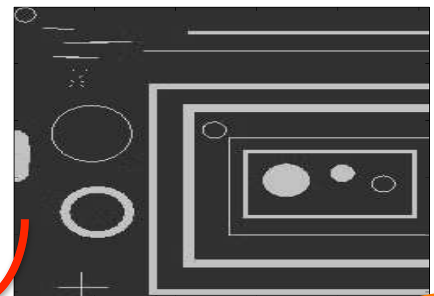


List Sparse Approximation: Implications?

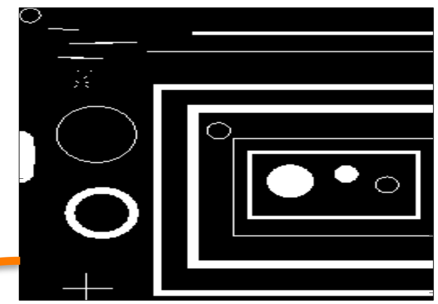


Original image

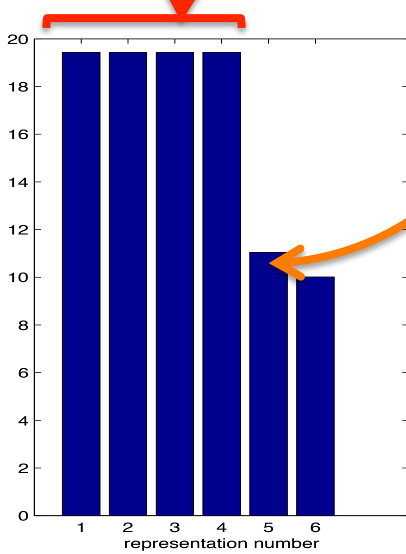
Random medium size coeffs

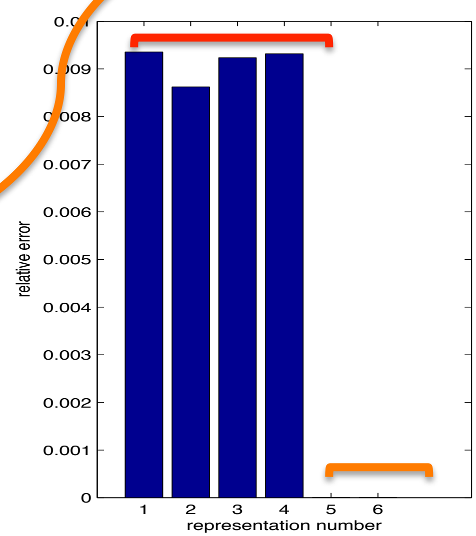


Truncated Best Basis



Equivalent error guarantees in compressed repn. versus ease of computation





Goals

1. Combinatorial bounds on list size

2. Dictionaries that achieve bounds

3. Practical algorithms + dictionaries

List Size: Clarifying definitions

- $A = I_n$, $b = (1,0,\ldots,0)$, $\varepsilon=\sqrt{2}$
  - Choose $x_1 \in (1-\sqrt{2}, 1+\sqrt{2})$ $x_j = 0$
  - NOT a meaningful list!

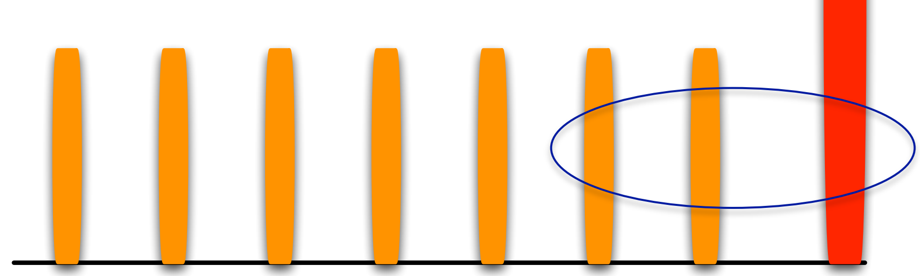
- **List Approx:** Given *A,b,k*,and ε, *L(A,b,k, ε)* is number of distinct support sets of *k*-sparse solutions *x* such that $\|Ax - b\|_2 \leq \epsilon$

List Size: Clarifying definitions

- A = I_n, $\epsilon < \sqrt{\frac{m-k}{m}}$

coherence μ = 0

$$b_i = \frac{\epsilon^2}{m-k} \qquad b_1 = 1 - (m-1)\frac{\epsilon^2}{m-k}$$



$$b = (b_1, b_2, \ldots, b_m)$$

- Any support set of size k that contains 1 has same error, $\binom{m-1}{k-1}$ different support sets
- *L(A, k, ε, R)* denotes the worst case bound on L over all b with the restriction that no atom appears in the support of more than *R* out of the *L* solutions.

List-Approx: Combinatorial bounds

- **Theorem:** (disjoint solutions)

If $\mu($ … $A)$ $]$

As long as the error
$\varepsilon \leq 1 - \Omega(\mu k)$, the number of
disjoint solutions is
$O(1/(1 - \varepsilon^2))$.

- Extends Uncertainty Principle to more than two disjoint solutions and allows some approximation error.
- Unique decoding results as corollaries.

List-Approx: Combinatorial bounds

- **Theorem:**

Let $0 < \gamma < 1$. As long as

If we consider only solutions where each atom appears only *o(L)* times in the output list of size *L*, then *L* is bounded only by *ε* (and is independent of *k* and *n*).

b derived from vector that demonstrates tightness of UP

- **Lemma:** Let A = Kerdock code dictionary, $\mu(A) = 1/\sqrt{n}$. For every $s < \sqrt{n}$, there is an input vector b s.t.
  - there are $\binom{n}{s}$ vectors x with sparsity $s\sqrt{n}$ and $Ax = b$
  - each atom appears in exactly s/n fraction of solns
- So,
  - For $\mu < 1/k$, $s = 1$, we have dictionary+input vector with $L(A,k,0,1) > n$.
  - For $s = \omega(1)$, $L(A,k,0,o(L))$ can be super-poly in n.
- ➔ Coherence bound is tight!

Conclusions

1. Combinatorial bounds on list sizes ✔
2. Dictionaries that achieve bounds ✔
3. Practical algorithms + dictionaries **?**