The interplay of analysis and algorithms
(or, Computational Harmonic Analysis)

Anna Gilbert
University of Michigan

supported by DARPA-ONR, NSF, and Sloan Foundation
Two themes
Sparse representation

Represent or approximate signal, function by a linear combination of a few atomic elements
Compressed Sensing

Noisy, sparse signals can be approximately reconstructed from a small number of linear measurements
Recovery = find significant entries

Sparse representation = signal recovery

different input models
How to compute?

Analysis and algorithms are both key components
Signal space: dimension $d$

Dictionary: finite collection of unit norm atoms

$$\mathcal{D} = \{ \phi_\omega : \omega \in \Omega \}, \quad |\Omega| = N > d$$

Representation: linear combination of atoms

$$s = \sum_{\lambda \in \Lambda} c_\lambda \phi_\lambda$$

Find best $m$-term representation
Applications

Approximation theory
Signal/Image compression
Scientific computing, numerics
Data mining, massive data sets
Generalized decoding
Modern, hyperspectral imaging systems
Medical imaging
SPARSE is NP-HARD
SPARSE is NP-COMPLETE
If dictionary is ONB, then SPARSE is easy (in polynomial time)
Incoherent dictionaries
(a basic result)

$\mu$-coherent dictionary, $\mu = \text{smallest angle between vectors}$

$m = \text{number of terms in sparse representation}$

$$m < \frac{1}{2\mu}$$

Algorithm returns $m$-term approx. with error

$$\|x - a_m\| \leq \sqrt{1 + \frac{2\mu m^2}{(1 - 2\mu m)^2}} \|x - a_{OPT}\|$$

Two-phase greedy pursuit

Joint work with Tropp, Muthukrishnan, and Strauss
Future for sparse approximation

Hardness of approximation is related to hardness of SET COVER

Approximability of SET COVER well-studied (Feige, etc.)

Need insight from previous work in TCS

Geometry is critical in sparse approximation

Need a way to describe better geometry of dictionary and its relation to sparse approximation: VC dimension?

Methods for constructing “good” redundant dictionaries (data dependent?)

Watch the practitioners!
Exponential time $O(2^d)$

Polynomial time $O(d^2)$

Linear time $O(d)$

Logarithmic time $O(\log d)$

General SPARSE

SPARSE, geometry

Matrix multiplication

FFT

Chaining, HHS Pursuit

AAFFT

Streaming wavelets, etc.
Computational Resources

Time
Space
Randomness
Communication
Models: Sampling

measurements:
length $N = m \log d$

$=\quad$  

$m$-sparse signal,
length $d$
Models: linear measurements

measurements: length $N = m \log d$

$m$-sparse signal, length $d$
Models: Dictionary

- Orthonormal bases
  - Fourier
  - Wavelets
  - Spikes
- Redundant dictionaries
  - Piecewise constants
  - Wavelet packets
  - Chirps
Results: Fourier

**Theorem:** On signal $s$ with length $d$, AAFFT builds $m$-term Fourier representation $r$ in time $m\text{poly}(\log d/\epsilon)$ using $m\text{poly}(\log d/\epsilon)$ samples with error

$$\|s - r\|_2 \leq (1 + \epsilon)\|s - s_m\|_2$$

On each signal, succeed with high probability.

G., Muthukrishnan, and Strauss 2005
Why sublinear resources?
Sparsogram
Extensions, applications

Generalize Fourier sampling algorithm to sublinear algorithm for linear chirps

Multi-user detection for wireless comm.

Radar detection and identification

Calderbank, G., and Strauss 2006
Lepak, Strauss, and G.
Results: Wavelets

**Theorem:** On signal $s$ with length $d$, streaming algorithm builds $m$-term wavelet representation in time $\text{poly}(m \log d/\epsilon)$ using $\text{poly}(m \log d/\epsilon)$ linear measurements with error

$$\|s - r\|_2 \leq (1 + \epsilon)\|s - s_m\|_2$$

On each signal, succeed with high probability.

G., Guha, Indyk, Kotidis, Muthukrishnan, and Strauss 2001
Results: Chaining

**Theorem:** With probability at least \(1 - d^{-3}\), the random measurement matrix \(\Phi\) has the following property. Suppose that \(s\) is a \(d\)-dimensional signal whose best \(m\)-term approximation with respect to \(\ell_1\) norm is \(s_m\). Given the sketch \(v = \Phi s\) of size \(O(m \log^2 d)\) and the number \(m\), the Chaining Pursuit algorithm produces a signal \(\hat{s}\) with at most \(O(m)\) nonzero entries. This signal estimate satisfies

\[
\|s - \hat{s}\|_1 \leq C \log m \|s - s_m\|_1
\]

The time cost of the algorithm is \(O(m \log^2 (m) \log^2 (d))\)

G., Strauss, Tropp, and Vershynin 2006
Algorithmic linear dimension reduction in $\ell_1$

**Theorem:** Let $Y$ be a set of points in $\mathbb{R}^d$ endowed with the $\ell_1$ norm. Assume that each point has at most $m$ non-zero coordinates. These points can be linearly embedded in $\ell_1$ with distortion $O(\log^3(m) \log^2(d))$, using only $O(m \log^2 d)$ dimensions. Moreover, we can reconstruct a point from its low-dimensional sketch in time $O(m \log^2(m) \log^2(d))$.
Theorem: With probability at least $1 - d^{-3}$, the random measurement matrix $\Phi$ has the following property. Suppose that $s$ is a $d$-dimensional signal whose $m$ largest entries are given by $s_m$. Given the sketch $v = \Phi s$ of size $m\text{polylog}(d)/\epsilon^2$

and the number $m$, the HHS Pursuit algorithm produces a signal $\hat{s}$ with $m$ nonzero entries. This signal estimate satisfies

$$\|s - \hat{s}\|_2 \leq \|s - s_m\|_2 + \frac{\epsilon}{\sqrt{m}}\|s - s_m\|_1$$

The time cost of the algorithm is $m^2\text{polylog}(d)/\epsilon^4$
Desiderata

**Uniformity:** Sketch works for *all signals* simultaneously

**Optimal Size:** $mpolylog(d)$ measurements

**Optimal Speed:** Update and output times are $mpolylog(d)$

Must have high quality: answer to query has near-optimal error
less information → measure less → compute less
Remark: Numerous contributions in area are not strictly comparable

Candes-(Romberg)-Tao 2004, 2005; Donoho 2004, 2005....
More formally....
Signal Information Recovery

Golomb-Weinberger 1959

[Diagram showing signal space, statistic map, statistic space, information space, and recovery algorithm]
More Formal Framework...

What signal class are we interested in?
What statistic are we trying to compute?
How much nonadaptive information is necessary to do so?
What type of information? Point samples? Inner products?
Deterministic or random information?
How much storage does the measurement operator require?
How much computation time, space does the algorithm use?
How much communication is necessary?
Computational Harmonic Analysis?
Algorithmic Harmonic Analysis $= \text{AHA!}$
http://www.math.lsa.umich.edu/~annacg
annacg@umich.edu
Isolation = Approximate Group Testing
Approximate group testing

Want to find $m$ spikes at height $1/m$, $\|\text{noise}\|_1 = 1$

Assign $d$ positions into $n = m \log d$ groups by $\Phi$

$\geq c_1 m$ of $m$ spikes isolated

$\leq c_2 m$ groups have noise $\geq 1/(2m)$

$\geq (c_1 - c_2)m$ groups have single spike and low noise except with probability $e^{-m \log d}$

Union bound over all spike configurations