A Compact Survey of Compressed Sensing

Graham Cormode

graham@research.att.com

Compressed Sensing In the News

Guardian Unlimited

Read today's paper · Jobs

Search: Go

Guardian Unlimited Web

technology

Home | Technology blog | Ask Jack blog | Gamesblog | Games | Opinion | Innovations | Inside IT

Inside IT

One pixel is plenty for pictures

Forget millions of pixels - two American researchers are working on a digital camera that has just one

Economist.com

advanced search »

Tuesday December 12th 2006

= requires subscription

LOG IN: E-mail address Password

PRINT EDITION

Science & Technology

Photography

A pixel worth a thousand words

Oct 26th 2006

From *The Economist* print edition

A new type of camera that could detect explosives and other hidden items

SALESMEN flogging digital cameras boast of the number of pixels in the images captured as an indicator of quality. The more pixels there are—and they are

Compressed Sensing on the Web

Discovery and Initial Papers

- Emmanuel Candès, Justin Romberg and Terence Tao, **Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information**. (IEEE Trans. on Information Theory, 52(2) pp. 489 - 509, Feb. 2006)
- Emmanuel Candès and Justin Romberg, **Quantitative Robust Uncertainty Principles and Optimally Sparse Decompositions**. (To appear in Foundations of Computational Mathematics)
- Emmanuel Candès and Terence Tao, **Near Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?** (To appear in IEEE Trans. on Information Theory)
- David Donoho, **Compressed Sensing**. (IEEE Trans. on Information Theory, 52(4), pp. 1289-1306, April 2006)

Compressed Sensing in Practice

Practical Signal Recovery

- Emmanuel Candès and Justin Romberg, **Practical Signal Recovery from Random Projections**. (Preprint, Jan. 2005)
- David Donoho and Yaakov Tsaig, **Extensions of Compressed Sensing**. (Signal Processing, 86(3), pp. 533-548, March 2006.)
- Joel Tropp and Anna Gilbert, **Signal Recovery From Partial Information Via Orthogonal Matching Pursuit**. (Preprint, 2005)
- Marco Duarte, Michael Wakin and Richard Baraniuk, **Fast Reconstruction of Piecewise Smooth Signals from Random Projections**. (Proc. SPARS Workshop, Nov. 2005)
- Chinh La and Minh Do, **Signal Reconstruction using Sparse Tree Representations**. (Proc. SPIE Wavelets XI, Sep. 2005)
- Gabriel Peyré, **Best Basis Compressed Sensing**. (Preprint, 2006) [See also related conference publication: **NeuroComp 2006**]
- Michael Elad, **Optimized Projections for Compressed Sensing**. (Preprint, 2006)

Compressed Sensing in Noise

- Jarvis Haupt and Rob Nowak, **Signal Reconstruction from Noisy Random Projections**. (IEEE Trans. on Information Theory, 52(9), pp. 4036-4048, Sep. 2006)
- Emmanuel Candès, Justin Romberg and Terence Tao, **Stable Signal Recovery from Incomplete and Inaccurate Measurements**. (Communications on Pure and Applied Mathematics, 59(8), pp. 1207-1223, Aug. 2006)
- Emmanuel Candès and Terence Tao, **The Dantzig Selector: Statistical Estimation When p is Much Larger Than n** (To appear in Annals of Statistics)
- Shriram Sarvotham, Dror Baron and Richard Baraniuk, **Measurements vs. Bits: Compressed Sensing Meets Information Theory**. (Proc. Forty-Fourth Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, Sep. 2006)
- Martin J. Wainwright, **Sharp Thresholds for High-Dimensional and Noisy Recovery of Sparsity** (Proc. Forty-Fourth Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, Sep. 2006)

Foundations and Connections

Coding and Information Theory

- Emmanuel Candès and Terence Tao, **Decoding by Linear Programming**. (IEEE Trans. on Information Theory, 51(12), pp. 4203-4215, Dec. 2005)
- Emmanuel Candès and Terence Tao, **Error Correction via Linear Programming**. (Preprint, 2005)

www.dsp.ece.rice.edu/CS/

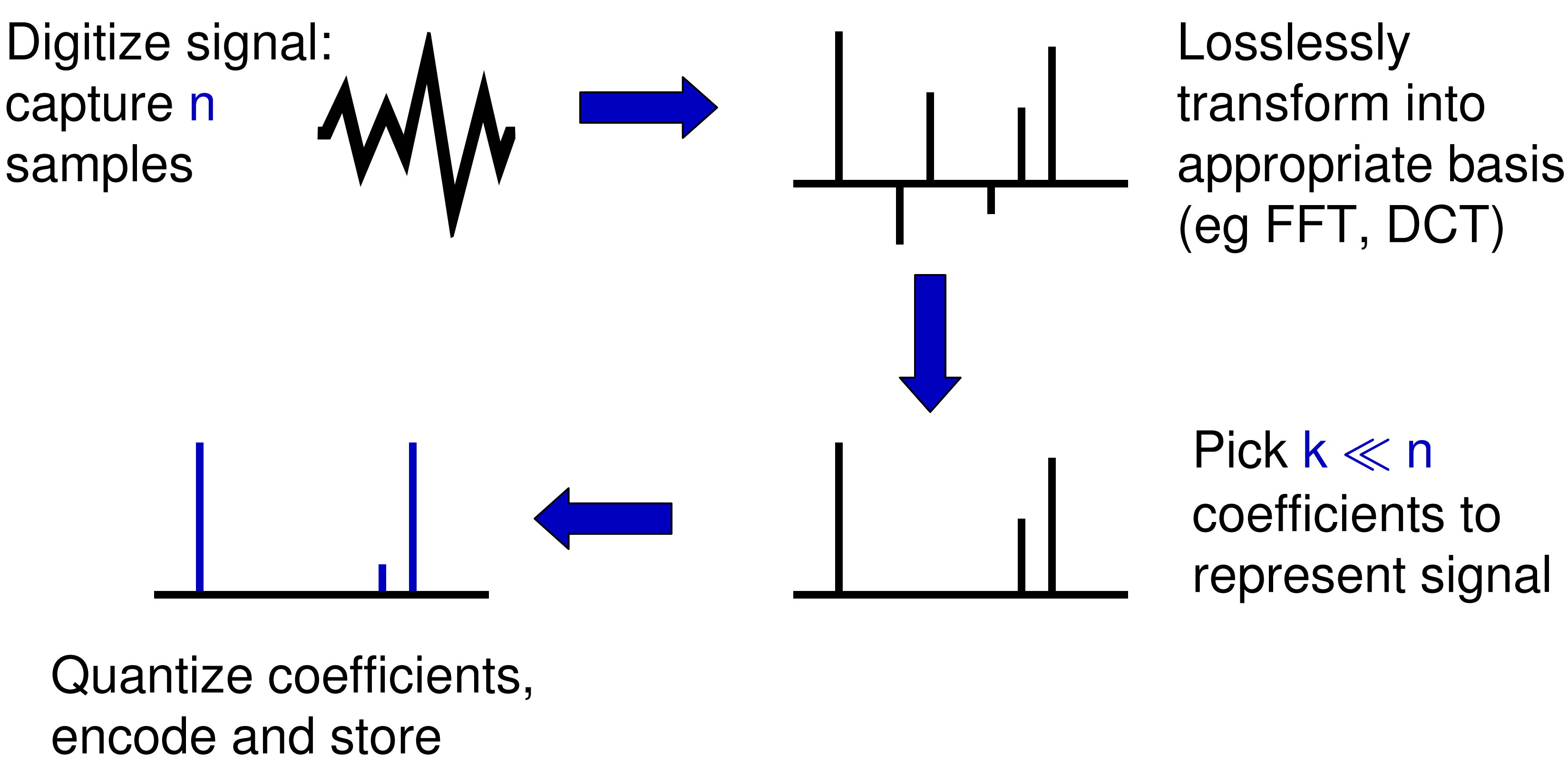
lists over 60 papers on "Compressed Sensing"…

So… what is Compressed Sensing?

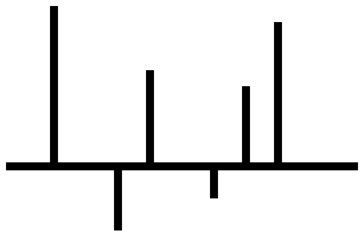
- Will introduce the CS problem and initial results
- Outline the (pre)history of Compressed Sensing
- Algorithmic/Combinatorial perspectives and new results
- Whither Compressed Sensing?

Signal Processing Background

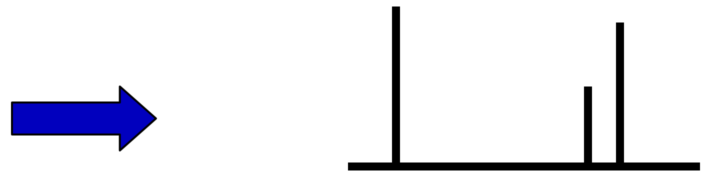
- Digital Signal Processing / Capture:



DSP Simplified



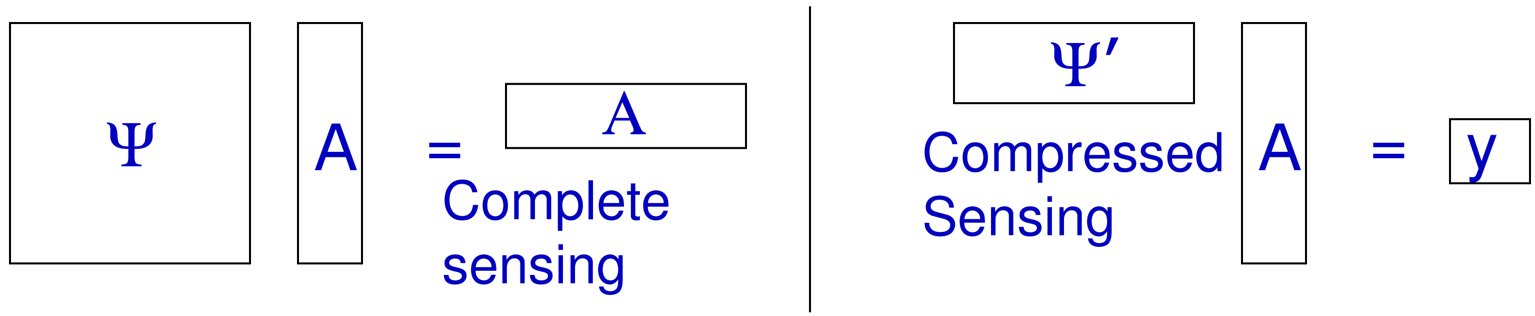
Discrete signal A
of dimension n



Select $k \ll n$ coefficients
to represent signal

- Observation: we make n measurements, but only end up storing k pieces of information
- What if measurements are very costly,
  - E.g. each one requires a separate hardware sensor
  - E.g. Medical imaging, patient is moved through scanner
- (Also, why do whole transform?, sometimes expensive)

The Compressed Sensing Credo



- Only measure (approximately) as much as is stored
- Measurement cost model:
  - Each measurement is a vector ψ_i of dimension n
  - Given ψ_i and signal (vector) A, measurement = $\psi_i \cdot A = y_i$
  - Only access to signal is by measuring
  - Cost is **number** of measurements
- Trivial solution: $\psi_i = 1$ at location i, 0 elsewhere
  - Gives exact recovery but needs n measurements

Error Metric

- Let R^k be a representation of A with k coefficients
- Define "error" of representation R^k as sum squared difference between R^k and A: $\|R^k - A\|_2^2$
- Picking k *largest* values minimizes error
  - Hence, goal is to find the "top-k"
- Denote this by R^k_{opt} and aim for error $\|R^k_{opt} - A\|_2^2$

“The” Compressed Sensing Result

Recover A “well” if A is “sparse” in few measurements

- “well” and “sparse” to be defined later

Only need $O(k \log n/k)$ measurements

- Each $\psi_i[j]$ is drawn randomly from iid Gaussian
- Set of solutions is all x such that $\psi x = y$
- Output $A' = \text{argmin } ||x||_1$ such that $\psi x = y$
  - Can solve by linear programming

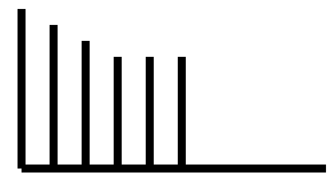
Why does it work?

[Donoho 04, Candes-Tao 04, Rudelson-Vershynin 04…]

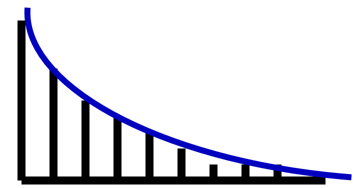
- Short answer: randomly chosen values ensure a set of properties of measurements ψ will work
  - The unexpected part: working in the L_1 metric optimizes error under L_2^2 with small support ("L_0 metric").
  - ψ works for any vector A (with high probability)
  - Other measurement regimes (eg Bernoulli ± 1)
- Long answer: read the papers for in-depth proofs that ψ has required properties (whp) and why they suffice
  - E.g. bounds on minimal singular value of each submatrix of ψ up to certain size

Sparse signals

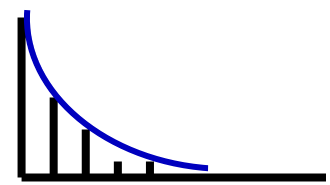
- How to model signals well-represented by k terms?
  - k-support: signals that have k non-zero coefficients under Ψ. So $\|R^k_{opt} - A\|_2^2 = 0$



  - p-compressible: sorted coefficients have a power-law like decay: $|\theta_i| = O(i^{-1/p})$.
    So $\|R^k_{opt} - A\|_2^2 = O(k^{1-2/p}) = \|C_k^{opt}\|_2^2$



  - α-exponentially decaying:
    even faster decay $|\theta_i| = O(2^{-\alpha i})$.

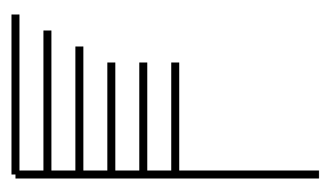


  - general: no assumptions on $\|R^k_{opt} - A\|_2^2$.
- (After an appropriate transform) many real signals are p-compressible or exponentially decaying.
  k-support is a simplification of this model.

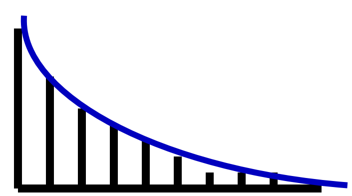
Sparse Signals

Original CS results apply principally to k-support and p-compressible signals.

- They guarantee exact recovery of k-support signals
- They guarantee "class-optimal" error on p-compressible
  - $\|R^k_{opt} - A\|_2^2 = O(k^{1-2/p}) = \|C_k^{opt}\|_2^2$
  - May not relate to the best possible error for that signal
  - (Algorithm does not take p as a parameter)



k-support

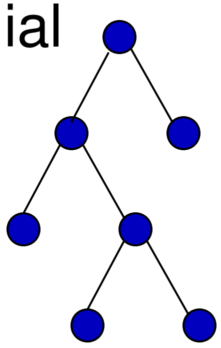


p-compressible

Prehistory of Compressed Sensing

Related ideas have been around for longer than 2 years…

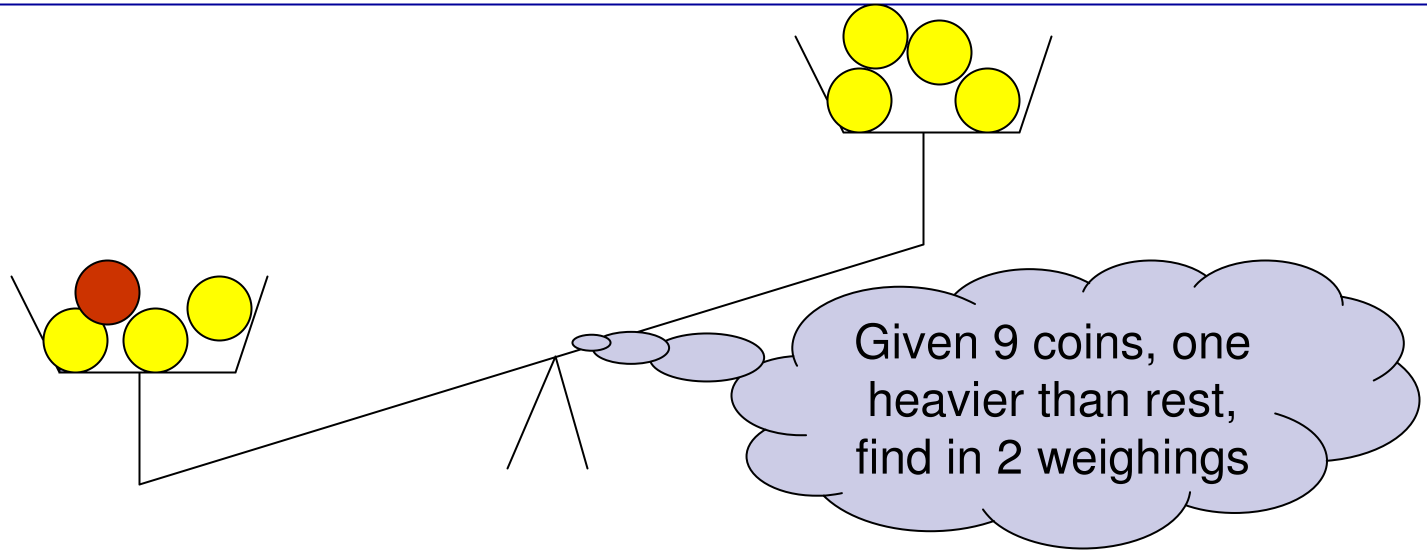
- Main results evolved through a series of papers on "a generalized uncertainty theorem" (Donoho/Candes-Tao…)
- Mansour 1992: "Randomized approximation and interpolation of sparse polynomials" by few evaluations of polynomial.
  - Evaluating a polynomial is dual of making a measurement
  - Algorithmic Idea: divide and conquer for the largest coefficient, remove it and recurse on new polynomial
  - Can be thought of as 'adaptive group testing', but scheme is actually non-adaptive



More Prehistory

- Gilbert, Guha, Indyk, Kotidis, Muthukrishnan, Strauss (and subsets thereof) worked on various fourier and wavelet representation problems in *data streams*
- Underlying problems closely related to Compressed Sensing: with restricted access to data, recover k out of n representatives to accurately recover signal (under L_2)
- Results are stronger (guarantees are instance-optimal) but also weaker (probabilistic guarantee per signal)
- Underlying technique is (non-adaptive) group testing.

Group Testing



- Break items (signal values) into groups
- Measure information on groups using binary vectors
  - Interpret results as positive or negative
- Recover identity of “heavy” items, and their values
- Continue (somehow) until all coefficients are found
  - General benefit: decoding tends to be much faster than LP

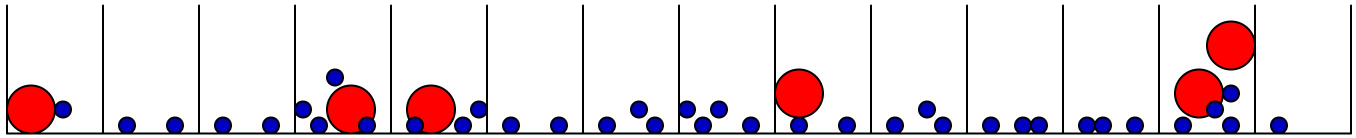
Trivial Group Testing

- Suppose A is 1-support signal (i.e. zero but for one place)
- Adaptive group testing: measure first half and second half, recurse on whichever is non-zero
- Non-adaptive: do in one pass using Hamming matrix H
  - log 2n x n matrix: log 2n measurements
  - The i'th column encodes i in binary
  - Measure A with H, read off location of the non-zero position, and its value

1	1	1	1	1	1	1	1
1	1	1	1	0	0	0	0
1	1	0	0	1	1	0	0
1	0	1	0	1	0	1	0

- Hamming matrix often used in group testing for CS
  - if a group has one large value and the rest "noise", using H on the group recovers item

Group Testing



From [C, Muthukrishnan 05], which specifically applies group testing to Compressed Sensing:

- From $O(c\ k/\varepsilon^2 \log^3 n)$ measurements, with probability at least $1 - n^{-c}$, and in time $O(c^2\ k/\varepsilon^2 \log^3 n)$ we find a representation R^k of A so $\|R^k - A\|_2^2 \leq (1+\varepsilon)\ \|R^k_{opt} - A\|_2^2$ (instance optimal) and R has support k.
- Randomly break into groups so not too many items fall in each group, encode as binary measurements using H
- Show good probability for recovering k largest values
- Repeat independently several times to improve probability

More Group Testing Results

- [Gilbert, Strauss, Tropp, Vershynin 06] develop new approaches with iterative recovery from measurements
  - Aiming for stronger “one set of measurements for all”
  - Must restate bounds on quality of representation
  - See next talk for full details!
- [Savotham, Baron, Baraniuk 06] use a more heuristic group testing approach, “sudocodes”
  - Make groups based on random divisions, no H
  - Use a greedy inference algorithm to recover
  - Seems to work pretty well in practice, needs strong assumptions on non-adversarial signals to analyze

Combinatorial Approaches

- A natural TCS question: if measurement sets exist which are good for all signals, can we construct them explicitly?
- Randomized Gaussian approach are expensive to verify – check complex spectral properties of all $\binom{N}{k}$ submatrices
- Do there exist combinatorial construction algorithms that explicitly generate measurement matrices for CS?
  - In n poly(log n,k) time, with efficient decoding algs.

K-support algorithms

- Achieve $O(k^2 \text{ poly}(\log n))$ measurements for k-support based on defining groups using residues modulo k log n primes > k [Muthukrishnan, Gasieniec 05]
  - Chinese remainder theorem ensures each non-zero value isolated in some group
  - Decode using Hamming matrix
- Apply k-set structure [Ganguly, Majumdar 06]
  - Leads to $O(k^2 \text{ poly}(\log n))$ measurements
  - Use matrix operations to recover
  - Decoding cost somewhat high, $O(k^3)$

More k-support algorithms

- Using "k-strongly separating sets" (from explicit constructions of expanders) [C, Muthukrishnan 06]
  - Similar isolation guarantees yield $O(k^2 \log^2 n)$ measurements
- [Indyk'06] More directly uses expanders to get $O(k2^{O(\log \log n)^2}) = O(kn^{\alpha})$ for $\alpha>0$ measurements
  - Bug Piotr to write up the full details…

Open question: seems closely related to coding theory on non-binary vectors, how can one area help the other

- Problem seems easier if restricted to non-negative signals

p-Compressible Signals

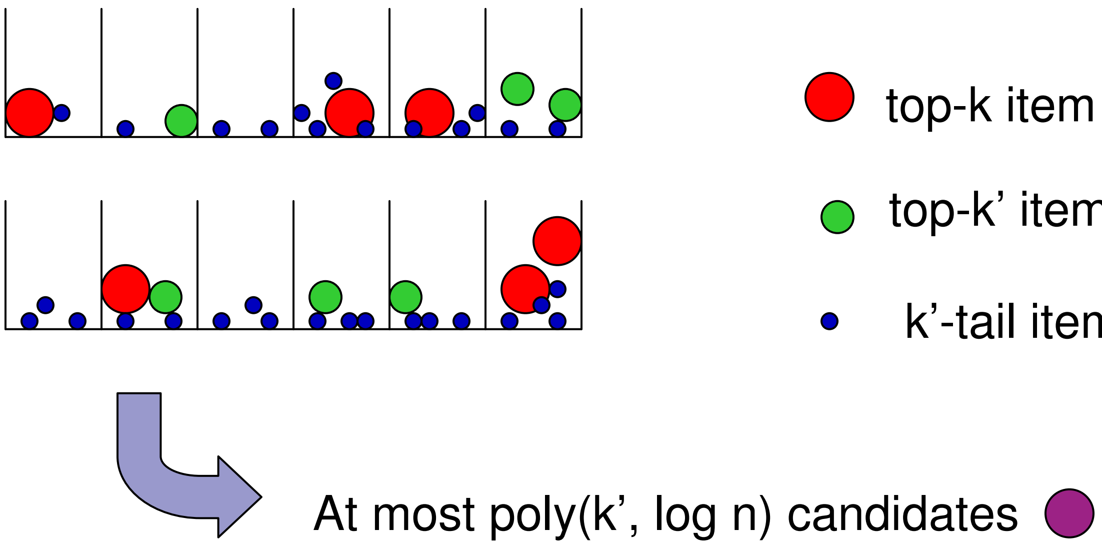
Explicit construction for p-compressible signals based on group testing [C, Muthukrishnan 06]

Approach: use two **parallel** rounds of group testing to find $k' > k$ large coefficients, and separate these to allow accurate estimation.

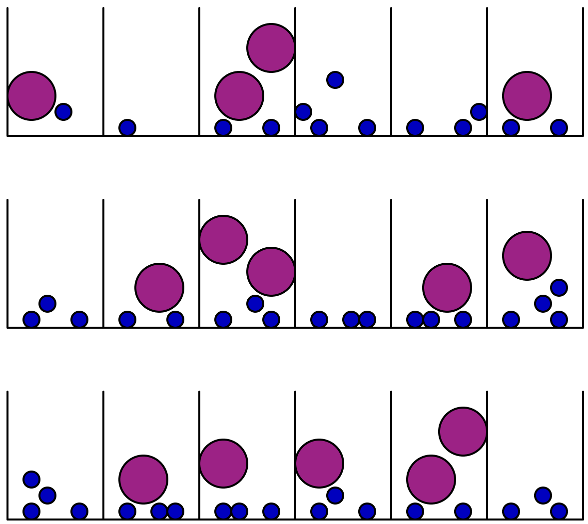
- Make use of K-strongly separating sets:
  - $S=\{S_1 \ldots S_m\}$ $m=O(k^2 \log^2 n)$
    For $X \subset [n], |X| \leq k, \forall x \in X. \exists S_i \in S. S_i \cap X = \{x\}$
  - Any subset of k items has each member isolated from $k-1$ others in some set

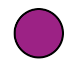
First Round

- Use k’ strongly separating sets to identify superset of k’ largest coefficients.
- k’ chosen based on p to ensure total “weight” of tail is so small that we can identify the k largest
- Combine groups with matrix H to find candidates



Second Round

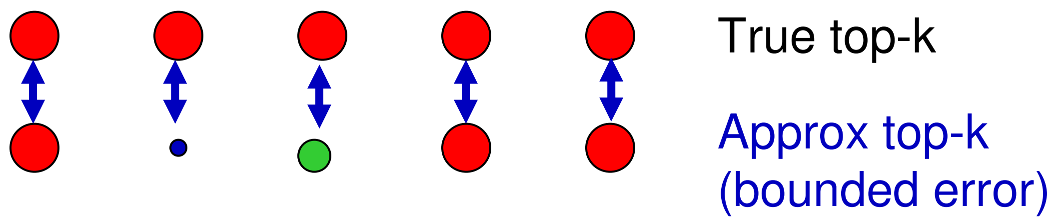


 At most C = poly(k', log n) candidates

- Use more strongly separating sets to separate out the candidates. (only need to know bound on C in advance)
- Get a good estimate for each coefficient: find a group it is isolated in, and use measurement of that group
  - can bound error in terms of ε, k, $||C_k^{opt}||_2^2$

Picking k largest

- Pick approximate k largest, and argue that coefficients we pick are good enough even if not the true k largest.
- Set up a bijection between the true top-k and the approx top-k, argue that the error cannot be too large.



- Careful choice of k' and k'' gives error that is

$$\|R^k - A\|_2^2 < \|R^k_{opt} - A\|_2^2 + \varepsilon\|C_k^{opt}\|_2^2$$

- Thus, explicit construction using $O((k\varepsilon^p)^{4/(1-p)^2}\log^4 n)$ ($poly(k,\log n)$ for constant $0 < p < 1$) measurements.

Open problem: Improve bounds, remove dependency on p

New Directions

- Universality
- Error Resilience
- Distributed Compressed Sensing
- Continuous Distributed CS
- Functional Compressed Sensing
- Links to Dimensionality Reduction
- Lower Bounds

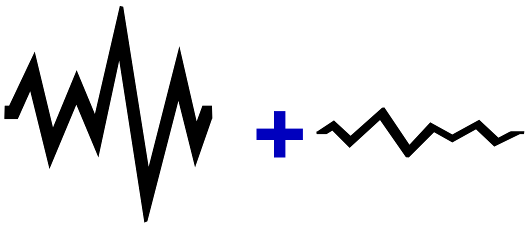
Universality

- Often want to first transform the signal with T
- So we compute $(\psi T)A = \psi(TA)$
- What if we don't know T till after measuring?
- If ψ is all Gaussians, we can write $\psi = \psi' T$, where ψ' is also distributed Gaussian
- We can solve to find ψ' and hence decode (probably)
- Only works for LP-based methods with Gaussians.

Open question: is there any way to use the group testing approach and obtain (weaker) universality?

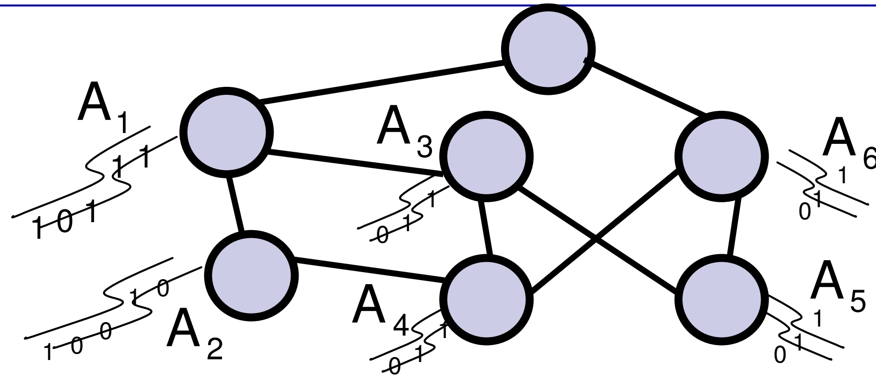
Error Resilience

- Various models of (random) errors:
  - signal is distorted by additive noise
  - certain measurements distorted by noise
  - certain measurements lost (erased) entirely
- LP techniques and group testing techniques both naturally and easily incorporate various error models



Open problem: extend to other models of error. More explicitly link CS with Coding theory.

Distributed Compressed Sensing

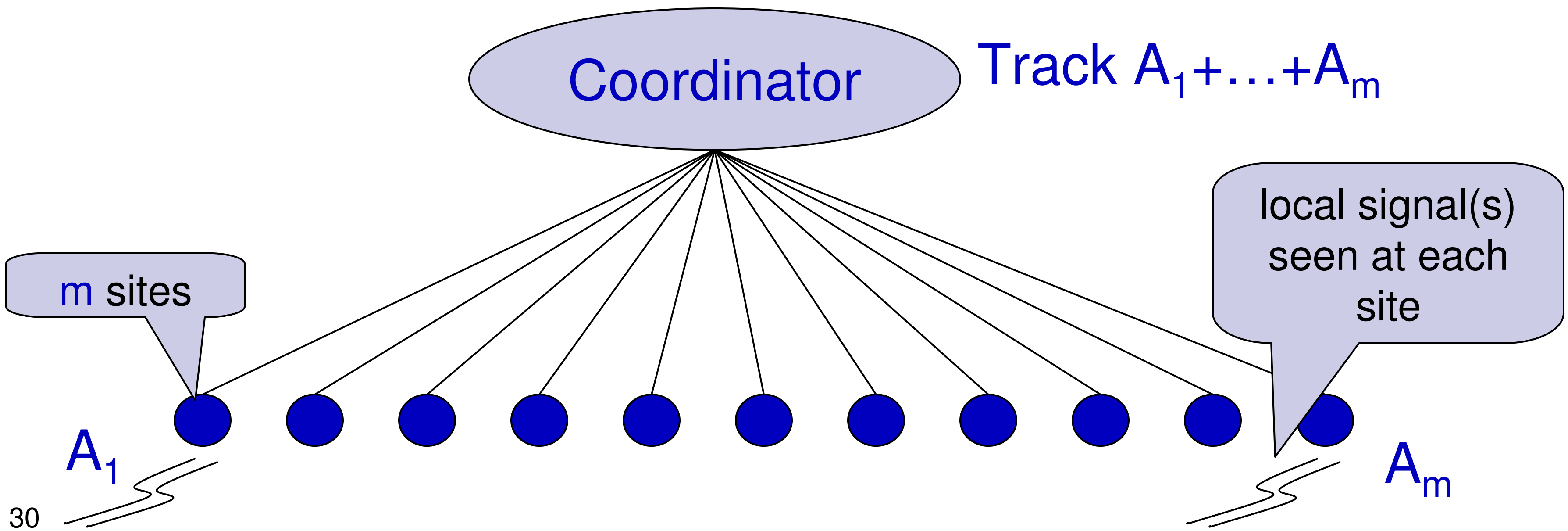


- Slepian-Wolf theorem: two correlated sources can be coded to use a total bandwidth proportional to their joint entropy without direct communication between two
- Apply to CS: consider correlated signals seen by multiple observers, they send measurements to a referee
  - Aim for communication proportional to CS bound
  - Different correlations: sparse common signal plus sparse/dense variations, etc Initial results in [Baraniuk+ 05]

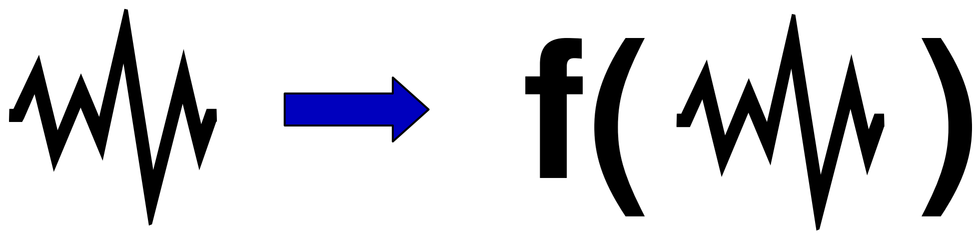
Open Problem: other arbitrary network graphs?

Continuous Distributed CS

- Different setting: each site sees part of a signal, want to compute on sum of the signals
- These signals vary "smoothly" over time, efficiently approximate the signal at coordinator site
- Statement and initial result in [Muthukrishnan 06]

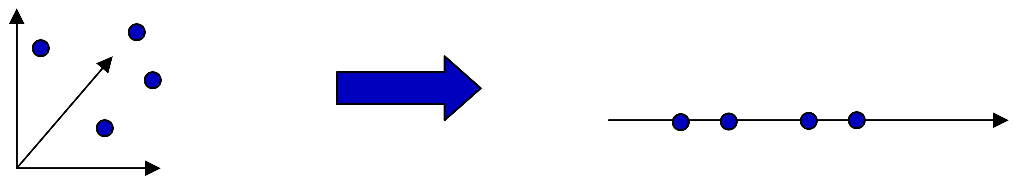


Functional Compressed Sensing



- In “traditional” CS, goal is accurate reconstruction of A
- Often, this is then used for other purposes
- Remember CS credo: measure for final goal
  - E.g. suppose we want to compute equidepth histograms, why represent A then compute histogram?
  - Instead, design measurements to directly compute function
- Initial results: quantiles on $A[i]^2$ [Muthukrishnan 06]
  - Different to previous sublinear work: need “for all” properties
  - Results in [Ganguly, Majumder 06] also apply here

Links to dimensionality reduction



- Johnson-Lindenstrauss lemma [JL 84]: Given a set of m points in n-dimensional Euclidean space, project to O(log m) dimensions and approximately preserve distances
  - Projections often via Gaussian random vectors
  - Intuitively related to CS somehow?
- [Baraniak et al 06] use JL-lemma to prove the “Restricted Isometry Property” needed to show existence of CS measurements

Open problem: further simplify CS proofs, use tools such as JL lemma and other embedding-like results

Lower Bounds

- Upper bounds are based on precise measurements
- But real measurements are discrete (encoded in bits)

Open Problems:

- What is true bit complexity needed by these algorithms?
- What is a lower bound on measurements needed?
  - $\Omega(k)$ or $\Omega(k \log k/n)$?
- How to relate to DSP-lower bounds: Nyquist bound etc.?
- LP formulation is over-constrained, can it be solved faster?

Conclusions

- A simple problem with a deep mathematical foundation
- Many variations and extensions to study
- Touches on Computer Science, Mathematics, EE, DSP…
- May have practical implications soon (according to the press)