Combinatorial Algorithms for Compressed Sensing

Graham Cormode
cormode@bell-labs.com

S. Muthukrishnan
muthu@cs.rutgers.edu

Background

- Dictionary Ψ is orthonormal basis for $\mathbb{R}^n$, ie n vectors ψ_i so $<\psi_i, \psi_j> = 1$ iff $i=j$, 0 otherwise
- Representation of dimension n vector A under Ψ is $\theta = \Psi A$, and $A = \Psi^T\theta$
- R^k is representation of A with k coefficients under Ψ
- Define "error" of representation R^k as sum squared difference between R^k and A: $\|R^k - A\|_2^2$
- By Parseval's, $\|R^k - A\|_2^2 = \|\theta^k - \theta\|_2^2 = \sum_{j \in \{[n] - k\}} \theta_j^2$ so picking k *largest* coefficients minimizes error
- Denote this by R^k_{opt} and aim for error $\|R^k_{opt} - A\|_2^2$

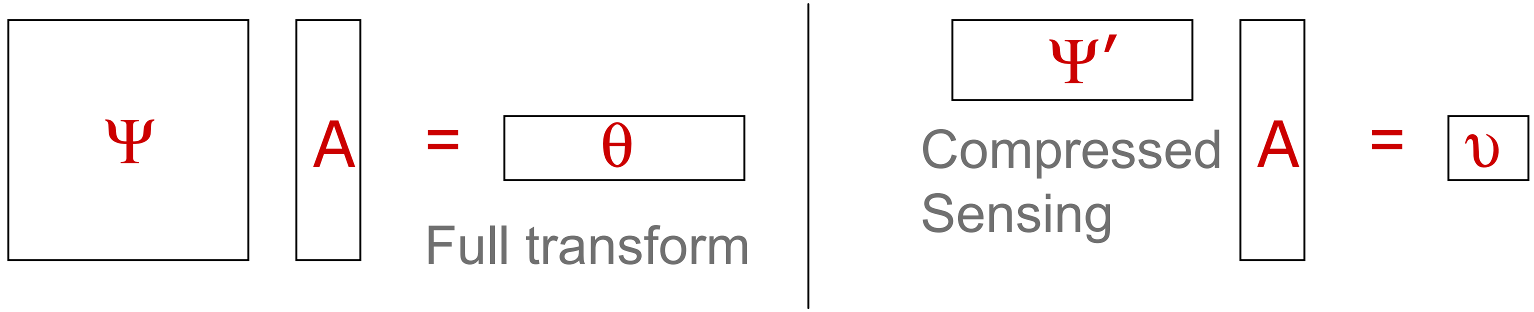
Sparse signals

How to model signals well-represented by k terms?

- k-support: signals that have k non-zero coefficients under Ψ. Hence $\|R^k_{opt} - A\|_2^2 = 0$
- p-compressible: coefficients (sorted by magnitude) display 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 $\|R^k_{opt} - A\|_2^2$.

Under an appropriate basis, many real signals are p-compressible or exponentially decaying. k-support is a simplification of this model.

Compressed Sensing



- Compressed Sensing approach: take $m \ll n$ (ie sublinear) measurements to build representation R
- Build Ψ' of m vectors from Ψ, compute Ψ'A and be able to recover good representation of A
- Developed by several groups: Donoho; Candes and Tao; Rudelson and Vershynin, and others, in frenetic burst of activity over last year or two.
- Results for p-compressible signals: randomly construct $O(k \log n)$ measurements, get error $O(k^{1-2/p})$ on any A (constant factor approx to best k term repn. of class)

Our Results

Can deterministically construct $O((k\varepsilon^p)^{4/(1-p)^2}\log^4 n)$ measurements in time polynomial in k and n.

For every p-compressible signal A, from these measurements of A, we can return a representation R for A of at most k coefficients θ' under Ψ such that

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

The time required to produce the coefficients from the measurements is $O((k\varepsilon^p)^{6/(1-p)^2} \log^6 n)$.

For α-exponentially decaying and k-sparse signals, fewer measurements are needed: $O(k^2 \log^4 n)$. Time to reconstruct is also $O(k^2 \text{ polylog } n)$

Recapping CS

Formally define the Compressed Sensing problem:

1. Dictionary transform. From basis Ψ, build dictionary Ψ' (m vectors of dimension n)
2. Measurement. Vector A is measured by Ψ' to get $\upsilon = \langle \psi_i', A \rangle$
3. Reconstruction. Given υ, recover representation R^k of A under Ψ.

Study: cost of creating Ψ', size of Ψ', cost of decoding υ, etc.

Explicit Constructions

Build explicit constructions of sets of measurements with guaranteed error.

Constructions work for **all** possible signals in the class.

Size of constructions is poly(k,log n) measurements

Using a group testing approach, based on two parallel tests.

Fast to reconstruct the approximate representation R: also poly in k and sublinear in n

Building the transformation

Set $\Psi' = T\Psi$ for transformation matrix T

So $\Psi' A = T\Psi A = T\theta$. Hence we get a linear combination of coefficients θ.

Design T to let us recover k large coefficients θ_i approximately. Argue this gives good representation.

Our constructions of T are composed of two parts:

- separation: allow identification of i
- estimation: recover high quality estimate of θ_i

Combinatorial tools

We use following definitions:

- K-separating sets $S = \{S_1, \ldots S_l\}$. $l = O(k \log^2 n)$
  For $X \subset [n]$, $|X| \leq k$, $\exists S_i \in S$. $|S_i \cap X| = 1$
- K-strongly separating sets $S=\{S_1 \ldots S_m\}$ $m=O(k^2 \log^2 n)$
  For $X \subset [n]$, $|X| \leq k$, $\forall x \in X$. $\exists S_i \in S$. $S_i \cap X = \{x\}$
- For set S, χ_S is characteristic vector, $\chi_S[i] = 1 \Leftrightarrow i \in S$
- Hamming matrix H, is $1+\log n \times n$
  (H represents 2-separating sets)
- Combining: if V is $v \times n$, W is $w \times n$.
  Define $V \otimes W$ as $vw \times n$ matrix:
  $$(V \otimes W)_{iv+l,j} = V_{i,j} W_{l,j}$$

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

p-compressible signals

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

First, identify a superset containing the k′ largest coefficients by ensuring that the total "weight" of the remaining coefficients is so small that we can identify the k′ largest.

Then use more strongly separating sets to separate out this superset, and get a good estimate for each coefficient.

Argue that taking the k largest approximate coefficients is a good approximation to the true k largest.

p-compressible

Over whole class, worst case error is $C_p k^{1-2/p} = \|C_k^{opt}\|_2^2$

The tail sum after removing the top k' obeys

$$\sum_{i=k'+1}^{n} |\theta_i| \leq O(k^{1-1/p})$$

Picking $k' > (k\varepsilon^{-p})^{1/(1-p)^2}$ ensures that even if every coefficient after the k' largest is placed in the same set as θ_i, for i in top k, we will recover i.

Build a k' strongly separating set S, and measure $\chi_S \otimes H$ to identify a superset of the top-k.

Build a $k'' = (k' \log n)^2$ strongly separating set R, and measure χ_R to allow estimates to be made

Can show we estimate θ_i with θ'_i so

$$(\theta'_i - \theta_i)^2 \leq \varepsilon^2/(25k)\ \|C_k^{opt}\|_2^2$$

Picking k largest

Argue that the coefficients we do pick are good enough even if they are not the k largest.

Write estimates as ϕ_i so $|\phi'_1| \geq |\phi'_2| \geq \ldots \geq |\phi'_n| = 0$

We also label coefficients so $|\theta_1| \geq |\theta_2| \geq \ldots \geq |\theta_n|$

Let π be the mapping so that $\phi_i = \theta_{\pi(i)}$

Our representation has error

$$\|R^k - A\|_2^2 = \Sigma_{i=1}^k (\phi_i - \phi'_i)^2 + \Sigma_{i=k+1}^n \phi_i^2$$

$$= \Sigma_{i<k}\ \varepsilon/25k\ \|C_k^{opt}\|_2^2 + \Sigma_{i>k,\ \pi(i)\leq k}\ \phi_i^2 + \Sigma_{i>k,\ \pi(i)>k}\ \phi_i^2$$

Optimal would also miss these coefficients

Bounding error

Set up a bijection σ between the coefficients in top k that we missed ($i>k$ but $\pi(i)\leq k$) and the coefficients outside the top k that we selected ($i\leq k$ but $\pi(i)>k$).

Because of the accuracy in estimation, can show that these mistakes have bounded error:

$$\phi_i^2 - \phi_{\sigma(i)}^2 \leq (2|\phi_{\sigma(i)}| + \varepsilon/(5\sqrt{k})\ \|C_k^{opt}\|_2^2)(2\varepsilon/(5\sqrt{k})\|C_k^{opt}\|_2^2)$$

Substituting in, can show

$$\Sigma_{i>k,\ \pi(i)\leq k}\ \phi_i^2 \leq 22\varepsilon/25\ \|C_k^{opt}\|_2^2 + \Sigma_{i\leq k,\ \pi(i)>k}\ \phi_i^2$$

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

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.

Other signal models

For α-exponentially decaying and k-sparse signals, can use fewer measurements

Separation: Build a k-strongly separating collection of sets S, encode as a matrix χ_S

Combine with H as $(H \oplus \chi_S)$

Estimation: build a $(k^2 \log^2 n)$-separating collection of sets R, encode as a matrix χ_R

Stronger guarantee on decay of coefficient values means we can estimate and subtract them one by one, and total error will not accumulate.

Total number of measurements in T is $O(k^2 \text{ polylog } n)$

Instance Optimal Results

We also give a randomized construction of Ψ' that guarantees instance optimal representation recovery with high probability:

- With probability at least $1 - n^{-c}$, and in time $O(c^2\, k/\varepsilon^2 \log^3 n)$ we can find a representation R^k of A under Ψ such that $\|R^k - A\|_2^2 \le (1+\varepsilon)\, \|R^k_{opt} - A\|_2^2$ (instance optimal) and R has support k.
- Dictionary $\Psi' = T\Psi$ has $O(ck \log^3 n / \varepsilon^2)$ vectors, constructed in time $O(cn^2 \log n)$; T is represented with $O(c^2 \log n)$ bits.
- If A has support k under Ψ then with probability at least $1 - n^{-c}$ we find the exact representation R.
- Some resilience to error in measurements

Concluding Remarks

- Alternate approach to compressed sensing by using combinatorial tools and techniques.
- Core of problem is to build a sublinear set of measurements to estimate of k largest coefficients.
- Still open to show better bounds on the size of Ψ', reconstruction cost, error guarantee etc.
- Many variations of the problem to consider: eg, what if basis Ψ is specified after measurements are made? Can there be deterministic constructions under conditions on Ψ (coherence to measurement basis?)

References and Thanks

CT04: Candes & Tao *Near optimal signal recovery from random projections and universal encoding strategies*, 2004

CRT04: Candes, Romberg & Tao *Robust uncertainty principles and optimally sparse decompositions* 2004

Don04: Donoho *Compressed Sensing*, 2004

GGIKMS02: Gilbert, Guha, Indyk, Kotidis, Muthukrishnan & Strauss *Fast, small-space algorithms for approximate histogram maintenance*, 2002

GT05: Gilbert & Tropp *Signal recovery from partial information via orthogonal matching pursuit*, 2005

RV05: Rudelson and Vershynin *Geometric approach to error correcting codes and reconstruction of signals*, 2005

Thanks to: Ron Devore, Ingrid Daubechies, Anna Gilbert and Martin Strauss for explaining compressed sensing.

Extension – Error Resilience

Prior work has considered resilience to errors, where random measurements are replaced with noise.

If a fraction $\rho = O(\log^{-1} n)$ of measurements are corrupted in this way, we can still recover R^k with $\|R^k - A\|_2^2 \leq (1+\varepsilon) \|R^k_{opt} - A\|_2^2$

Basic intuition is that provided error avoids some set of measurements of θ_i we can recover it as before.

Estimation is also resilient to errors, due to taking median of several estimates.

Can improve error tolerance to $\rho = O(1)$ [can be as much as 1/10] by a modified algorithm with higher decoding cost ($\Omega(n)$).