CS 395T: Sublinear Algorithms, Fall 2020 October 8th 2020

Lecture 13: Compressed Sensing

Prof. Eric Price *Scribe: Joshua Prupes*

NOTE: THESE NOTES HAVE NOT BEEN EDITED OR CHECKED FOR CORRECTNESS

1 Overview

Goal of Compressed Sensing: want to estimate a structured signal $x \in \mathbb{R}^n$ from $m << n$ linear measurements

$$y = Ax \quad (+\text{ noise})$$

Examples for this model include

- x is image of bottom through murky water
- Single pixel camera, takes in an n x n dimensional array of inputs, applies some masking and convolution to give one output (could be less expensive then megapixel camera)
- MRIs: look at inner product of measurements with Fourier transform from magnets
- Strata of Earth: Thump the ground in certain locations and have microphones in other areas that determine strength of vibrations. Want to minimize number of measurements.
- Audio: High resolution estimation
- Spectrum sensing: want to find empty band that isn't in use for radio. Want $Sampling\ rate << 8GHz$

$$m < n \implies x \text{ not uniquely determined by y}$$

These images are compressible because x can be sparse in some basis (ie images are sparse in wavelet/DFT basis)

Definitions
x is "exactly" k-sparse iff x has k nonzero values
x is "approximately" k-sparse iff $||x - x_{(k)}||$ is "small"

If x is "exactly" k-sparse we can represent x with $\log(\binom{n}{k})$ bits for locations and k "words"

$\implies$ hope for $m = k$ or $m = k\log(n)$ to suffice

2 Basic Result of Candes, Romberg, Tao '06

$$A \text{ has iid } N(0,1/m) \text{ entries } \implies \text{ can reconstruct } \hat{x} \text{ from } y = Ax \text{ s.t. } \hat{x} = x \text{ if x is k-sparse}$$

$$m = O(k \log{(n/k)})$$

$$||\hat{x} - x||_1 \leq 3||x - x_{(k)}||_1 \quad \forall x$$

$$||\hat{x} - x||_2 \leq 3||x - x_{(k)}||_2 \text{ w.h.p.}$$

$$||\hat{x} - x||_2 \leq O(||e||_2) \quad \forall \text{ k-sparse x } \forall \text{ noise } e$$

L1 minimization: This holds for $\hat{x} = argmin||\hat{x}||_1$ s.t. $||y - A\hat{x}|| \leq \epsilon$ for appropriate ϵ
LASSO: This holds for $\hat{x} = argmin(\lambda||\hat{x}||_1 + ||y - A\hat{x}||_2)$ for appropriate λ
Iterative Hard Thresholding: Holds for

$$x_0 = 0$$

$$x_{i+1} = H_k(x_i + A^T(y - Ax_i))$$

H_k is the function which restricts to the largest k entries, and $y - Ax_i$ is the residual error at each step

The above methods work for any RIP matrix A. RIP includes

- iid subgaussian
- subsampled Fourier
- partial circulent
- incoherent

3 RIP Matrices

Definition of Restricted Isometry Property (RIP)

$A \in \mathbb{R}^{m*n}$ satisfies (k, ϵ)-RIP if

$$||Ax||_2 = (1 + \epsilon)||x||_2 \quad \forall \text{ k-sparse } x$$

$$(2k, 1/2)\text{-RIP} \implies ||Ax - Ax'|| \geq 1/2||x - x'|| \quad \forall \text{ k-sparse } x, x'$$

$$\implies Ax + e \neq Ax' + e' \quad \text{if } ||e||, ||e'|| << ||x - x'||$$

$$\implies \text{ can't confuse x and x'}$$

Alternative Definition (RIP)

$$||(A^TA - I)_{SxS}||_2 \leq \epsilon \quad \forall |S| \leq k$$

Claim A with iid $N(0, 1/m)$ satisfies RIP with $m = O_\epsilon(k \log{(n/k)}$

From last class: to be (n, ϵ) RIP, m = $1/\epsilon^2 \log(\#$ possible $x)$

$$T_k = \{x| \quad x \text{ is k-sparse, } ||x||_2 \leq 1\}$$

$$N_c(\epsilon, T_k, ||\cdot||_2) \leq \binom{n}{k} * (1 + 2/\epsilon)^k$$

Where N_c is the covering number

$$\implies \log(N_c) \leq (k \log(n/k) + k \log(1/\epsilon) = k \log(n/k\epsilon)$$

$$\implies \frac{1}{\epsilon^2} k \log(\frac{n}{k\epsilon}) \text{ suffices for RIP}$$

RIP Matrix examples

- Random (sub)-gaussian
  - $m \geq k \log(n/k)$ will have RIP with probability $1 - e^{-\Omega(n)}$
  - But: can't test if A has RIP and takes mn space to store
- Matrices with low coherence
  - columns $a_1, ..., a_n$ satisfy $\frac{|\langle a_i, a_j \rangle|}{\sqrt{||a_i|| * ||a_j||}} < 1/k$
  - above holds if iid gaussian $m > k^2 \log(n)$
  - Benefit is easy to check condition
  - Cost is $m > k^2$
- Random rows of Fourier matrix
  - $F_{ij} = e^{2\pi\sqrt{-1} * ij/n}$
  - $F_{\Omega \cdot \star} \quad \Omega \subset [n]$ uniformly $|\Omega| \geq k \log(n) \log^2(k)$
  - Pros
    - How MRI's work
    - Can multiply quickly which leads to faster algorithms
    - can be stored in $|\Omega| = o(m)$ space
  - Cons
    - have $\log^2(k) factor$
    - can't check if Ω is good
- Partial circulant
  - $a_1, a_2, ...a_n = \pm1, \pm1, ..., \pm1$
    $a_n, a_1, ...a_{n-1} = \pm1, \pm1, ..., \pm1$
    $a_{n-1}, a_n, ...a_{n-2} = \pm1, \pm1, ..., \pm1$
    for the first m rows
  - very similar to Random Fourier
  - $m = O(k \log(n) \log^3(k))$
  - explicit construction: $k^{2-\epsilon} \log(n)$ rows for small ϵ
- No sparse matrices satisfy RIP :(

4 Compressed Sensing vs Heavy Hitter Algorithms

Both: see $y = Ax$ and output $\hat{x} \approx x$ assuming $x \approx$ k-sparse

Compressed Sensing

- Dense
- w.h.p, for all x
- $||\hat{x} - x||_1 \leq C||x - x_k||_1$
- matrix are more restricted so algos are more general

Heavy Hitters

- Sparse (ie fast updates)
- for each x w.h.p.
- $||\hat{x} - x||_\infty \leq \frac{C}{k}||x - x_k||_1$
- matrix is specifically constructed so algos are tied to the matrix

References

[CandesRT06] Emmanuel Candes, Justin Romberg, Terence Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory*, 58(2):489–509, 2006.