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 $n \times n$ dimensional array of inputs, applies some masking and convolution to give one output (could be less expensive than 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 $\log\left(\binom{n}{k}\right)$ bits for locations and k “words”

$$\implies \text{hope for } m = k \text{ or } m = k \log(n) \text{ to suffice}$$
2 Basic Result of Candes, Romberg, Tao '06

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

\[
m = O(k \log (n/k))
\]

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

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

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

**L1 minimization:** This holds for \( \hat{x} = \text{argmin} ||\hat{x}||_1 \) s.t. \( ||y - A\hat{x}|| \leq \epsilon \) for appropriate \( \epsilon \)

**LASSO:** This holds for \( \hat{x} = \text{argmin}(\lambda||\hat{x}||_1 + ||y - A\hat{x}||_2) \) for appropriate \( \lambda \)

**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 circulant
- incoherent

3 RIP Matrices

**Definition of Restricted Isometry Property (RIP)**

\( A \in \mathbb{R}^{m \times n} \) satisfies \( (k, \epsilon) \)-RIP if

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

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

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

\( \implies \text{can't confuse} \ x \ \text{and} \ x' \)

**Alternative Definition (RIP)**

\[
|| (A^T A - I)_{S \times S} ||_2 \leq \epsilon \quad \forall |S| \leq k
\]

**Claim** \( A \) with iid \( N(0,1/m) \) satisfies RIP with \( m = O_\epsilon(k \log (n/k)) \)
From last class: to be \((n, \epsilon)\) RIP, \(m = \frac{1}{\epsilon^2 \log(\# \text{ possible } x)}\)

\[ T_k = \{ x \mid x \text{ is } k\text{-sparse, } \|x\|_2 \leq 1 \} \]

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

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 \(a_1, \ldots, a_n\) 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
  - \(F_{ij} = e^{2\pi \sqrt{-1} * ij/n}\)
  - \(F_{\Omega}\) \(\Omega \subset [n]\) 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 \(\Omega\) is good

- Partial circulant
  - \(a_1, a_2, \ldots, a_n = \pm 1, \pm 1, \ldots, \pm 1\)
  - \(a_1, a_2, \ldots, a_{n-1} = \pm 1, \pm 1, \ldots, \pm 1\)
  - \(a_{n-1}, a_n, \ldots, a_{n-2} = \pm 1, \pm 1, \ldots, \pm 1\)
  - 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 \(\epsilon\)

- 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