Compressive Sensing

Based on Candes and Tao’s work

Presenter: Namrata Vaswani

Compressive Sensing

- Combine compression with sensing to improve sensing.
  The term “compression” is used here in the sense of dimension reduction.
- Examples
  - The gradient of a piecewise constant signal is sparse in the time domain.
  - Natural signals or images are “sparse” (or “compressible”) in the DCT domain, i.e. many DCT coefficients are zero (or small enough to be approx. by zero): exploited by JPEG.

Idea of CS

- **Assume discrete time signals throughout. Signal here is always an N-length vector.**
- Exploit sparsity or compressibility of natural signals (images) in a given basis: call it the "sparsity basis".
  - If a signal is sparse in the time domain, sparsity basis is I_N.
  - If a signal is sparse in Fourier or DCT domain, sparsity basis is F_N (N-DFT matrix) or D_N (N-DCT matrix).
  - In general any $N \times N$ orthogonal matrix may be the sparsity basis.
- Measure random linear projections of the signal (exploit the incoherence between the measurement matrix and the sparsity basis).

Sparsity and Compressibility: Definitions

- A $N \times 1$ vector x is S-sparse, if only $S < N$ elements are non-zero.
- x is compressible if only a small number of elements are significantly non-zero. One model is:

$$|x_{(k)}| < Rk^{1/p}, \;\; 1 < p < 2$$

  where $|x_{(1)}| \geq |x_{(2)}| \cdots \geq |x_{(k)}| \cdots \geq |x_{(N)}|$.
- Wavelet coefficients of many natural signals and images satisfy this.

Compressive Sensing

- Assume $x_{N \times 1}$ is an S-sparse vector.
- Measure $y = Ax$ where $A_{K \times N}$ is the measurement matrix, $K < N$.
- A: random DFT matrix (pick K random rows of F_N) or a random Gaussian matrix (each entry iid Gaussian) or random Bernoulli matrix.
- $K = O(S \log N)$
- Goal: reconstruct x from y and A.
- CS: compute $\hat{x}$ by solving P1

$$(P1) \quad \min_{\tilde{x} \in \mathbb{R}^N} ||x||_{\ell_1} \; s.t. \; Ax = y. \quad ||x||_{\ell_1} := \sum_{i=1}^{N} |x_i|$$

Exact Reconstruction Result for CS

$$(P1) \quad \min_{\tilde{x} \in \mathbb{R}^N} ||x||_{\ell_1} \; s.t. \; Ax = y$$

- Result 1: If x is S-sparse and A satisfies the "Uniform Uncertainty Principle (UUP) at about level $3S$", then the solution, $\hat{x}$ to P1 is unique and is equal to x.
- Result 2: If $K = O(S \log N)$ and A is random Gaussian, it satisfies UUP with high probability (w.p. $\geq 1 - O(e^{-\gamma N})$)
- Similar results exist for random Fourier and Bernoulli also.

Uniform Uncertainty Principle: Quantifying Incoherence

A $K \times N$ matrix, A, satisfies UUP at level S if it obeys the S-Restricted Isometry Property.

Let A_T, $T \subset \{1, 2, \ldots N\}$ be the $K \times |T|$ sub-matrix obtained by extracting the columns of A corresponding to the indices in T. Then RIP requires that there exists a $\delta_S < 1$ s.t.

$$(1 - \delta_S)||c||^2 \leq ||A_T c||^2 \leq (1 + \delta_S)||c||^2$$

for all subsets $T \subset \{1, 2, \ldots N\}$ of size $|T| \leq S$ and for all $c \in \mathbb{R}^N$.

- In other words, every set of S or less columns of A is approximately orthogonal (has eigenvalues b/w $1 \pm \delta_S$).
- Or that, A is approximately orthogonal for any S-sparse vector, c.
- Smaller δ_S (more incoherence) requires lesser measurements.

CS Theorem: Exact reconstruction of an S-sparse signal is possible by solving (P1) if $A_{K\times N}$ satisfies

$$\delta_S + \delta_{2S} + \delta_{3S} < 1$$

CS: General form

- Assume the signal of interest z is sparse in the basis ϕ, i.e. $z = \phi x$ and x is S-sparse. We sense $y = \psi z = \psi \phi x$.
- It is assumed that $\psi_{K \times N}$ is "incoherent w.r.t. ϕ", i.e. $A = \psi \phi$ satisfies UUP.
- A random Gaussian matrix, ψ, is incoherent w.r.t. any orthogonal basis, w.h.p. This is because if ψ is r-G, then $\psi \phi$ is also r-G (ϕ any orthonormal matrix).
- Same property for random Bernoulli.

Time-Frequency Examples

Sparse in frequency: Consider a periodic signal made up of only 6 sinusoids ($S = 12$ frequencies in the DFT domain). Assume its period is $T_0 = 1s$ and bandwidth is 499Hz.

- Traditional method: Sense this every $T_s = 1ms$ for one time period $T_0 = 1s$, i.e. obtain $N = 1000$ samples. Since it is periodic, it can be **exactly** reconstructed by sinc interpolation using these N samples.
- CS: know that the signal is sparse in Fourier domain (1000-DFT has only $S = 12$ non-zero coefficients).
- If sense the signal at only $K = O(S \log N)$ **randomly chosen time instants** chosen from the set $\{0, T_s, 2T_s, \dots (N-1)T_s\}$, w.h.p., it is possibly to exactly reconstruct it by solving (P1).

Sparse spikes in the time , e.g. neuronal spikes. Assume there are at most N spikes, spaced $i\Delta T$ apart, i.e. $x_c(t) = \sum_{i=0}^{N-1} x_i \delta(t - i\Delta T)$. But usually very sparse, i.e. only $S << N$ spikes occur, i.e. x is sparse. If can sense the Fourier transform of this signal at $K = O(S \log N)$ random frequencies chosen from $\{2\pi j/(N\Delta T), j = 0, 1, \ldots N-1\}$, can reconstruct it exactly using (P1) w.h.p.

- Another example is the gradient of a p.w. constant image. MR imaging systems sense the DFT of the image along 22 radial lines. If 256-DFT sensed, then have only K =22x256 DFT observations to reconstruct a N =256x256 image.
- Similar problem in tomography: sense the radon transform of the image: from which can compute 22 radial DFT's.

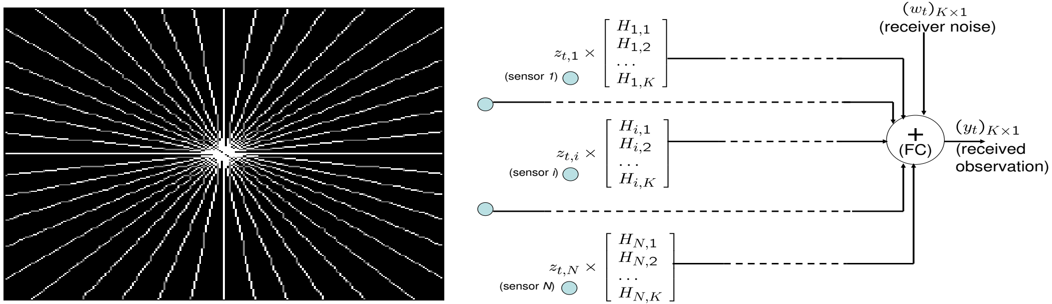
Applications

- Biomedical imaging: MR, tomography. Current methods either assume the unknown DFT coefficients to be zero or try to interpolate (sensitive in DFT domain) or regularize the problem by using priors for piecewise smooth or p.w. constant images. CS does not "assume any prior" only assumes knowledge of sparsity in a given basis.
- One pixel camera: senses K random linear projections of the image (image multiplied by a random Bernoulli matrix), reconstructs by using the fact that natural images are sparse in the wavelet domain.
- Decoding "sparse" channel transmission errors.
- Other A-to-D converters.

Comparison with Traditional Sensing Mechanisms

- Higher resolution cameras without increasing CCD sensor density to get more resolution: one-pixel camera (works by capturing K random projections of the image one at a time, current prototype takes 5 minutes to capture enough projections: valid if completely static image).
- Of course need the resolution in the DMD array.
- Do not need temporary storage of high volume data (required in traditional cameras before performing compression).
- High processing power needed only at the reconstruction end, only need enough power to compute K random projections (much less computation than that needed to compress, still tx much less than the uncompressed signal): useful to create a cheap sensors.

- While still transmitting only $K = O(S \log N)$ coefficients, not the entire N (in case of uncompressed).



(a) CT or MRI (projection) geometry [8]

(b) Sensing protocol of Haupt-Nowak [30]

Figure 2: Practically sensing linear incoherent measurements. Fig. 2(a): MR systems in projection geometry mode, or CT systems, measure the 2D DFT of the image along 22 radial lines [33, 8]. Fig. 2(b): An efficient protocol to receive K random linear projections of temperature at M nodes over a MAC channel [30].

CS for Compressible Signals

If A satisfies $\delta_{3S} + \delta_{4S} < 1$, then the reconstruction error satisfies

$$||x - \hat{x}||_{\ell_1}^2 \leq C||x - x_S||_{\ell_1}$$

i.e. with $K = O(S \log N)$ measurements the reconstruction error is of the same order as that with keeping only the S largest coefficients of x.

- Compressible signals are usually modeled as

$$|x_{(k)}| \leq Rk^{-1/p},\ 1 < p < 2$$

- This can be used to show that

$$||x - x_K||_{\ell_1} \leq C_1 R K^{-(1/p - 1/2)}$$

But if I did CS using K measurements (i.e. $S = \alpha K / \log N$), the error is

$$||x - \hat{x}||_{\ell_1}^2 \leq C_2 R(\alpha K / \log N)^{-(1/p-1/2)}$$

In other words, the error has increased about $(\alpha \log N)^{(1/p-1/2)}$ times that with knowing the signal and keeping top K coefficients.

CS for Noisy Signals

- Sparse Signals: Let x is S-sparse and A satisfies $\delta_{2S} + \delta_{3S} < 1$. Assume that the observation is

$$y = Ax + z, \;\; z \sim \mathcal{N}(0, \sigma^2 I_K)$$

Let $\lambda := \sqrt{2 \log N (1+\beta)}$. Let $\hat{x}$ is the solution to

$$(P2) \quad \min ||\tilde{x}||_1 \;\; s.t. \;\; ||A^T(y - A\tilde{x})||_\infty \leq \lambda\sigma$$

Then w.p. $\geq 1 - (\sqrt{\pi \log N} p^\beta)^{-1}$, $\hat{x}$ satisfies

$$||x - \hat{x}||_2^2 \leq C_1^2 \lambda^2 S \sigma^2$$

with $C_1 := 4/(1 - \delta_{2S} - \delta_{3S})$.

- Compressible signals: Let x is compressible and A satisfies

$\delta_{2S} + \delta_{3S} < 1$. Then w.h.p. $\hat{x}$ satisfies

$$||x - \hat{x}||_2^2 \leq C_3^2 \lambda^2 \min_{1 \leq s \leq S} \left(s\sigma^2 + R^2 s^{-(2/p-1)} \right)$$

Papers to Read

- Decoding by Linear Programming (CS without noise, sparse signals)
- Dantzig Selector (CS with noise)
- Near Optimal Signal Recovery (CS for compressible signals)
- Applications