A Quick Introduction on Compressive Sensing

Yubei Chen
Pressure is on DSP

- Shannon/Nyquist sampling theorem
  - For general band-limited signal, no information loss if we sample at 2x signal bandwidth

- DSP revolution:
  - sample first and ask questions later

- Increasing pressure on DSP hardware, algorithms
  - ever faster sampling and processing rates
  - ever larger dynamic range
  - ever larger, higher-dimensional data
  - ever lower energy consumption

...
But wait ... Do we really need general signal?

- What does general signal mean (say a general image)?
  - With high probability it’s something like this ...
But wait ... Do we really need general signal?

- What does general signal mean (say a general image)?
  - But usually we are just interested in the structured signal like this ...
From general signal to structured signal

- We want to make more assumption (rational?) about our signal in order to lower the limit of sampling rate, let compare more between our previous assumption and the reality:

- A lot of high/low level structures, e.g. edges, continuation.
Sparse Representation

- Inspired by the statistics of some transforms (DCT, Wavelet ...), we assume our signal is sparse in some domain.

\[ x = \sum_{i=1}^{N} \alpha_i \psi_i \quad \alpha_i = \langle x, \psi_i \rangle \]

- \( K \)-sparse: \( K \) large coefficients, where \( K \ll N \)

\[ x \approx \sum_{K \ll N \text{ largest terms}} \alpha_i \psi_i \]
Transform Coding and Its Inefficiency

• The previous sample-then-compress framework:
  – K largest coefficients are located and the rest N-K smallest coefficients are discarded.
  – The K values and locations of the largest coefficients are encoded.

• Major three inefficiencies:
  – N maybe large even K is small
  – N coefficients need to be computed even N-K of them will be discarded
  – Encoding of locations introduce an overhead
The Compressive Sensing Problem

• A general M-dimensional linear measurement:

\[ y = \Phi x = \Phi \Psi s = \Theta s \]

• \( \Phi \) is not adaptive and it’s stable – the salient information in any K-sparse or compressive signal is not damaged by this dimension reduction process.

• Then we recover the signal based on \( y, \Phi, \Psi \)
The Solution to CS

- If the signal is not compressive, the problem is ill-conditioned.
- If the K locations are known, a necessary and sufficient condition for well-conditioning is:
  - The matrix $\Theta$ preserves the lengths of these K-sparse vectors.

$$1 - \epsilon \leq \frac{\|\Theta v\|_2}{\|v\|_2} \leq 1 + \epsilon$$

- A sufficient condition for a stable solution for both K-sparse and compressible signals is $\Theta$ satisfies the above condition for an arbitrary 3K-sparse vector $v$. This is referred to as restricted isometry property (RIP).
- Incoherence: rows of $\Phi$ do not sparsely represent $\Psi$. 
The Solution to CS: Sampling

- However fortunately such a $\Phi$ is easy to construct as a random matrix and the previous condition is satisfied with high probability.
- For instance: let each elements of $\Phi$ be i.i.d Gaussian random variables $\sim \mathcal{N}(0, 1/M)$.
  - Quick verification by expectation and Chebyshev inequality ...
The Solution to CS: Reconstruction

• E.g. Lasso Regression

\[ \min_{s} \| y - \Phi \Psi s \|^2_2 + \lambda \| s \|_1 \]

• The basic idea is to preserve the information and also seek sparse representation.

• Other similar convex optimization based on this idea can also be formulated, say, make the L2 norm a convex constraint.
An Image Filling-in Case: Single Pixel Camera
Further Relaxation

- Non-orthogonal basis map
- Redundant basis map
- Dictionary learning: Sparse coding, Independent component analysis (infomax etc.)
Conclusion of Compressive Sensing Approach

• Transform Sparsity

• Non-coherency

• Non-linear Construction (Optimization)
Thanks!