EECS 275 Matrix Computation

Ming-Hsuan Yang

Electrical Engineering and Computer Science
University of California at Merced
Merced, CA 95344
http://faculty.ucmerced.edu/mhyang

Lecture 26

Overview

- Compressive sensing
- Applications

Compressive sensing: Background

- Nyquist/Shannon sampling theory:
  - the number of samples needed to reconstruct a signal without error is dictated by its bandwidth, i.e., at least 2 times the bandwidth
  - the length of the shortest interval which contains the support of the spectrum of the signal
- Has significant interactions and bearings on some fields in the applied science and engineering such as statistics, information theory, coding theory, theoretical computer science
- Sparsity: has bearing on the data acquisition process itself, and leads to efficient data acquisition protocols
- Sparse coding: well known in numerous fields such as visual cortex, neuroscience, computer vision, image processing, and machine learning

Compressible signals

- Consider a real-valued finite-length one-dimensional, discrete-time signal $\mathbf{x} \in \mathbb{R}^N$
- Any signal in $\mathbb{R}^N$ can be represented in terms of basis $N \times 1$ vectors $\{\boldsymbol{\psi}_i\}_{i=1}^N$
- For simplicity, assume orthonormal basis, and $\mathbf{x}$ is represented as

$$\mathbf{x} = \sum_{i=1}^{N} s_i \boldsymbol{\psi}_i, \quad \text{or} \quad \mathbf{x} = \boldsymbol{\Psi}\mathbf{s} \tag{1}$$

  where $\mathbf{s}$ is the $N \times 1$ vector of weighted coefficients, $s_i = \langle \mathbf{x}, \boldsymbol{\psi}_i \rangle = \boldsymbol{\psi}_i^\top \mathbf{x}$
- The signal $\mathbf{x}$ is K-sparse if it is a linear combination of only K basis vectors, i.e., , only K of the s_i coefficients are nonzero and $N - K$ are zero (of great interest when $K \ll N$)
- The signal is compressible if (1) has only a few large coefficients and many small coefficients

Transform coding

- The fact that compressible signals are well approximated by K-sparse representations forms the foundation of transform coding
- In data acquisition (e.g., digital cameras) transform coding plays a central role:
    1. full N-sample signal $\mathbf{x}$ is acquired
    2. complete set of transform coefficients $\{s_i\}$ is computed via $\mathbf{s} = \Psi^\top \mathbf{x}$
    3. K largest coefficients are located and the $N - K$ smallest coefficients are discarded
    4. K values and locations of the largest coefficients re encoded
- JPEG: exploits sparse representation based on discrete cosine transform (DCT)
- JPEG 2000: exploits sparse representation based on discrete wavelet transform (DWT)

Compressive sensing: Main idea

- Standard compression schemes
  - acquire the full signal
  - compute the complete set of transform coefficients
  - encode the largest coefficients
  - discard all the others
  - operate at Nyquist rate
- Compressive sensing/sampling (aka compressed sensing):
  - directly acquire the data in already compressed form
  - does not need to throw away anything
  - sample at rate lower than Nyquist rate

Compressive sensing: Two principles

- Sparsity:
  - information rate of a continuous time signal may be much smaller than suggested by its bandwidth
  - discrete time signal depends on a number of degree of freedom which is much smaller than its length
  - many natural signals are sparse or compressible
- Incoherence:
  - extends the duality between time and frequency domains
  - objects having a sparse representation in Ψ must be spread out in the domain in which they are sampled, just as a Dirac or spike in the time domain is spread out in the frequency domain
  - unlike the signal of interest, the sampling/sensing waveforms have a extremely dense representation in Ψ

Compressive sensing

- Consider a general linear measurement process that computes $M < N$ inner products between $\mathbf{x}$ and a collection of vectors $\{\phi_j\}_{j=1}^{M}$ as in $y_j = \langle \mathbf{x}, \phi_j \rangle$
- Arrange the measurements y_j in an $M \times 1$ vector $\mathbf{y}$ and the measurement vectors $\phi_j^\top$ as rows in an $M \times N$ matrix Φ, we have

$$\mathbf{y} = \Phi \mathbf{x} = \Phi \Psi \mathbf{s} = \Theta \mathbf{s}$$

  where $\Theta = \Phi\Psi$ is an $M \times N$ matrix
- Nonadaptive measurement process, i.e., Φ is fixed and does not depend on $\mathbf{x}$
  - need to find a stable measurement matrix Φ such that the salient information in any K-sparse or compressible signals is not damaged by dimensionality reduction from $\mathbf{x} \in \mathbb{R}^N$ to $\mathbf{y} \in \mathbb{R}^M$ $(M < N)$
  - a reconstruction algorithm to recover $\mathbf{x}$ from only $M \approx K$ measurements $\mathbf{y}$

Designing a stable measurement matrix

- The measurement matrix Φ must allow reconstruction of the length N signal $\mathbf{x}$ from $M < N$ measurements (i.e., $\mathbf{y}$)
- Appear to be ill-posed at first glance
- If, however, $\mathbf{x}$ is K-sparse and the K locations of nonzero coefficients in $\mathbf{s}$ are known, then the problem can be solved provided $M \geq K$
- A necessary and sufficient condition is, for any vector $\mathbf{v}$ sharing the same K nonzero entries as $\mathbf{s}$ and for some $\varepsilon > 0$

$$1 - \varepsilon \leq \frac{\|\Theta \mathbf{v}\|_2}{\|\mathbf{v}\|_2} \leq 1 + \varepsilon \tag{2}$$

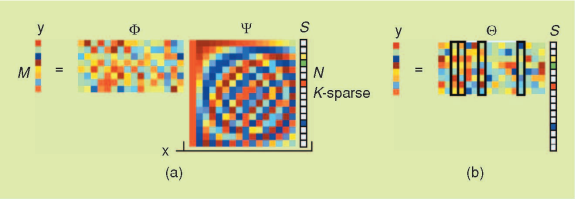
- That is, the matrix Θ must preserve the lengths of these particular K-sparse vectors
- In general, the locations of the K nonzero entries in $\mathbf{s}$ are not known

Designing a stable measurement matrix (cont'd)

- Restricted isometry property (RIP): A sufficient condition for a stable solution for both K-sparse and compressible signals is that Θ satisfies (2) for an arbitrary $3K$-sparse vector $\mathbf{v}$
- Incoherence: A related condition that requires the row $\{\phi_i\}$ of Φ cannot sparsely represent the columns $\{\phi_i\}$ of Ψ, and vice versa
- Direct construction of Φ such that $\Theta = \Phi\Psi$ has the RIP requires verifying (2) for each of the $\binom{N}{K}$ possible combinations
- However, both the RIP and incoherence can be achieved with high probability simply by selecting Φ as a random matrix
- For example, let the matrix elements $\phi_{j,i}$ be independently and identically distributed (iid) random variables form a Gaussian probability density function with zero mean and variance $1/N$
- Then the measurements $\mathbf{y}$ are merely M different randomly weighted linear combinations of the elements of $\mathbf{x}$

Gaussian random measurements

- The measurement matrix Φ is incoherent with the basis $\Psi = I$ of data spikes with high probability
- More specifically, an $M \times N$ iid Gaussian matrix $\Theta = \Phi I = \Phi$ can be shown to have the RIP with high probability if $M \geq c\ K \log(N/K)$ with c a small constant
- Thus, K-sparse and compressible signals of length N can be recovered from only $M \geq c\ K \log(N/K) \ll N$ random Gaussian measurements
- The matrix Φ is universal in the sense that $\Theta = \Phi\Psi$ will be iid Gaussian and thus have the RIP with high probability regardless of choice of orthonormal basis Ψ



Signal reconstruction

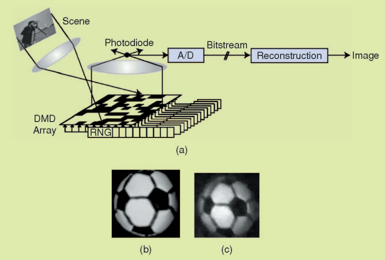
- Take M measurements in the vector $\mathbf{y}$, the random measurement matrix Φ, and the basis Ψ to reconstruct $\mathbf{x}$ (or equivalently its sparse coefficient $\mathbf{s}$)
- For K-sparse signals, there are infinitely man $\mathbf{s}'$ that satisfy $\Theta\mathbf{s}' = \mathbf{y}'$ since $M < N$
- An underconstrained problem and for $\Theta\mathbf{s} = \mathbf{y}$, there exists $\Theta(\mathbf{s} + \mathbf{r}) = \mathbf{y}$ for any vector $\mathbf{r}$ in the null space $\mathcal{N}(\Theta)$ of Θ
- The signal reconstruction algorithm aims to find the signal's sparse coefficient vector in the $(N - M)$-dimensional translated null space $\mathcal{H} = \mathcal{N}(\Theta) + \mathbf{s}$

Optimization problems

- Minimum ℓ_2-norm reconstruction
$$\hat{\mathbf{s}} = \arg\min_{\|\mathbf{s}'\|_2} \Theta\mathbf{s}' = \mathbf{y}$$
which can be solved with closed form solution, $\hat{\mathbf{s}} = \Theta^\top(\Theta\Theta^\top)^{-1}\mathbf{y}$, but almost never find a K-sparse solution
- Minimum ℓ_0-norm reconstruction
$$\hat{\mathbf{s}} = \arg\min_{\|\mathbf{s}'\|_0} \Theta\mathbf{s}' = \mathbf{y}$$
which can recover a K-sparse signals with only $M = K + 1$ iid Gaussian measurements, but it is both numerically unstable and NP-complete
- Minimum ℓ_1-norm reconstruction
$$\hat{\mathbf{s}} = \arg\min_{\|\mathbf{s}'\|_1} \Theta\mathbf{s}' = \mathbf{y}$$
which can recover K-sparse signals and closely approximate compressible signals with high probability using only $M \leq c\ K \log(N/K)$ iid Gaussian measurements via convex optimization

Applications

- Single pixel, compressive digital camera that directly acquires M random linear measurements without first collecting the N pixel values
- Use digital micromirror device consisting of an array of N tiny mirrors where each one can be independently oriented
- To collect measurements, a random number generator sets the mirror orientations in a pseudorandom pattern to create the measurement ϕ_j and the voltage at the photodiode equals y_j, the inner product between ϕ_j and $\mathbf{x}$
- The process repeats M times to obtain $\mathbf{y}$



[Baraniuk 07]