Lecture 22: More On Compressed Sensing

Scribed by Eric Lee, Chengrun Yang, and Sebastian Ament

Nov. 2, 2017

1 Recap and Introduction

Basis pursuit was the method of recovering the sparsest solution to an underdetermined linear system i.e.

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \|x\|_1 \\ \text{subject to} \quad & \|Ax - b\| < \epsilon \end{aligned}$$

Compressed sensing is related to basis pursuit, but stems from a different context, in which there is some underlying signal or function $f \in \mathbb{R}^n$ that we cannot observe. What we can observe is some set of linear observations

$$\left\{ y_i \mid y_i = \psi_i^T f \right\}_{i=1}^m \tag{1}$$

with $\psi_i \in \mathbb{R}^n$. Compressed sensing asks if there is some minimal number set of observations that may be made i.e. gives bounds on the size of m. Let $\Psi = \begin{bmatrix} \psi_1 & \psi_2 & \dots & \psi_m \end{bmatrix} \in \mathbb{R}^{n \times m}$. If $m = n$, we can set up the system of equations

$$\Psi^T f = y \tag{2}$$

with

$$\Psi = \begin{pmatrix} \Psi_1^T \\ \Psi_2^T \\ \vdots \\ \Psi_m^T \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

If Ψ is invertible, the solution is trivial (given by inverting Ψ). However, we are interested in some $m < n$ such that exact recovery is still possible. For general problems, this is not possible; we are interested on problems with a bit more structure to them.

2 Mutual Coherence

To be more precise, we assume that f is data-sparse in some basis Φ:

$$f = \Phi x$$

where $\Phi \in \mathbb{R}^{n \times n}$ is assumed to be orthogonal and $xin\mathbb{R}^n$ is sparse. This basis Φ may be derived from some physical problem or model.

Then expanding f into (2), we observe that this is equivalent to solving

$$\Psi^T \Phi x = y \tag{3}$$

However, if Ψ is not invertible (the case when $m < n$), we are interested in when the basis pursuit-esque formulation where

$$A = \Psi^T \Phi$$

and we would like to solve the optimization problem

$$\begin{aligned} \underset{z}{\text{minimize}} \quad & \|z\|_1 \\ \text{subject to} \quad & Az = y \end{aligned}$$

Note however that we are interested in *exact* solutions and not *approximate* solutions. To simplify things, we make the additional assumption that the columns of Ψ are sampled uniformly at random from an larger orthogonal matrix $\bar{\Psi} \in \mathbb{R}^{n \times n}$

That is,

$$A = (\bar{\Psi} R)^T \Phi$$

where $R \in \mathbb{R}^{n \times m}$ is a sampling matrix that picks out columns of $\bar{\Psi}$

Before we define Mutual Coherence, we first give a toy example to illustrate why Mutual Coherence is important.

2.1 Toy Example

Assume that $\bar{\Psi} = \Phi = F$ where F is the $n \times n$ Discrete Fourier Transform Matrix. Let $x = e_1$, the unit vector associated with the first coordinate. Then regardless of which R we pick, we note that our samples

$$y = \Psi^T \Phi e_1 \tag{4}$$

Will be the zero vector unless Ψ contains the first column of F. This is in some senses, the worst case scenario; we need $m = n$ in order to guarantee that we can extract e_1. Conversely, if $m = n - 1$, we have a very high probability of getting no information from our linear measurements y

2.2 Mutual Coherence

Keeping the toy example in mind, Mutual Coherence is defined as

$$\mu(\bar{\Psi}, \Phi) = \sqrt{n} \max_{i,j} |\psi_i^T \phi_j| \tag{5}$$

Where ϕ_j is the j-th column of the matrix Φ. The way μ is defined here is slightly different from the way it might be defined in basis pursuit. μ here ranges between 1 and $\sqrt{n}$ (with 1 representing the lowest possible score and $\sqrt{n}$ representing the highest possible score). But other than the fact that μ is scaled up by $\sqrt{n}$, there is no difference in its meaning; μ represents how "different" two bases are. Following this logic, low mutual incoherence between two bases indicates guarantees that any vector sparse in one basis cannot be be sparse in the other; this is a critical property to have, as going back to the toy example, we see that having a vector be sparse in both bases leads to sensing problems.

Theorem 2.1 *Given $f \in \mathbb{R}^n$, $x \in \mathbb{R}^n$, $\Phi \in \mathbb{R}^{n\times n}$, and $\bar{\Psi} \in \mathbb{R}^{n\times n}$ such that*

$$f = \Phi x$$

with $\bar{\Psi}$ and Φ orthogonal, and x k-sparse (having only k nonzero entries), select m columns of $\bar{\Psi}$ uniformly at random and put in Ψ. Then if

$$m \geq c\ \mu(\bar{\Psi}, \Phi)^2\ k\ log(n)$$

for some constant c, the solution $\hat{x}$ to the optimization problem

$$\begin{aligned} \underset{z}{minimize} \quad & \|z\|_1 \\ subject\ to \quad & \Psi^T \Phi z = y \end{aligned}$$

is exact with high probability.

Now, in both exact and approximate sparse recovery cases, we want to find some structures of A to give provable guarantees and bounds for the effect of our sparse recovery. Thus we define Restricted Isometry Property (RIP) as follows:

Definition. Restricted Isometry Property (RIP) is defined as: For $k = 1, 2, ...,$ define δ_k (a constant) as the smallest number such that

$$(1 - \delta_k)||x||_2^2 \leq ||Ax||_2^2 \leq (1 + \delta_k)||x||_2^2$$

for all k-sparse x. Then we say the matrix A satisfies k-Restricted Isometry Property with restricted isometry constant δ_k.

For a matrix A that satisfies the RIP, any subset of k columns of A are "well-behaved", i.e. A is guaranteed to project any vector x with corresponding nonzero pattern to another vector that has 2-norm close to x. A has the RIP if $\delta_1, \delta_2, \cdots, \delta_{2k}$ are "small", which leads to the alternative form of RIP:

$$\delta_k = \max_{\text{card}(S)\leq s} ||A_S^* A_S - I||_2$$

should be small, in which the 2-norm is the spectral norm.

Let's say we want to recover a k-sparse signal and δ_{2k} is sufficiently < 1. We have an equivalent form of RIP being $(1-\delta_{2k})||x_1 - x_2||_2^2 \leq ||Ax_1 - Ax_2||^2 \leq (1+\delta_{2k})||x_1 - x_2||_2^2$. Here $2k$ instead of k exists because we can only guarantee $x_1 - x_2$ to be $2k$-sparse.

Theorem 2.2 *Assume $\delta_{2k} < \sqrt{2} - 1$, then $\hat{x}$, the solution to*

$$\begin{aligned} &\underset{z}{\textit{minimize}} && \|z\|_1 \\ &\textit{subject to} && Az = y \end{aligned}$$

obeys $||\hat{x} - x||_2 \leq c||x - x_k||_1/\sqrt{k}$ and $||\hat{x} - x||_1 \leq c||x - x_k||_1$, where $y = Ax$ and x_k is the best k-sparse approximation of x.

Note: x is not necessarily sparse here.

Now, suppose $y = Ax + g$, in which g is noise and $||g||_2 \leq \epsilon$, and let $\hat{x}$ solve

$$\begin{aligned} &\underset{z}{\text{minimize}} && \|z\|_1 \\ &\text{subject to} && ||Az - y||_2 \leq \epsilon \end{aligned}$$

which is noise-tolerant instead of the exact case in the above theorem.

Theorem 2.3 *Assume $\delta_{2k} \leq \sqrt{2} - 1$, then $\hat{x}$ satisfies*

$$||\hat{x} - x||_2 \leq c_0||x - x_k||_1/\sqrt{k} + c_1\epsilon$$

for constants c_0 and c_1.

For example, if $\delta_{2k} < \frac{1}{4}$, then $c_0 \leq 5.5$, $c_1 \leq 6$.

Some more conclusions on the structure of A:

What general classes of matrices A satisfy the RIP? As it turns out, if random matrices do the job. To be more precise, we want an A with m close to k satisfying the RIP. In the cases where either $A \in \mathbb{R}^{m\times n}$ with i.i.d. normal mean and variance $\frac{1}{m}$ entries, or $A \in \mathbb{R}^{m\times n}$ has each column sampled uniformly at random on the unit sphere in $\mathbb{R}^m$, A obeys the RIP for δ_{2k} ($< \sqrt{2} - 1$) with high probability if $m \geq ck\log(\frac{n}{k})$.

In the case $A = R\bar{\Psi}^T\Phi$, in which R is a random subsampling matrix, it is sufficient to have $m \geq k(\log n)^4$ and get the RIP with high probability for δ_{2k}.

If $A = G\Phi$ with G being an $m \times n$ random matrix as before, and Φ being a set of fixed orthogonal bases, A obeys the RIP with high probability if $m \geq ck\log(\frac{n}{k})$.

3 Supplementary Information: Application of Compressed Sensing to MRI Imaging

Compressed sensing has found innumerable applications in imaging, in particular medical imaging, and seismic imaging, where the cost of measurement is high, but the data can usually be represented in a sparse format. Further, it has found applications in biological sensing, radar systems, communication networks, and many more [1].

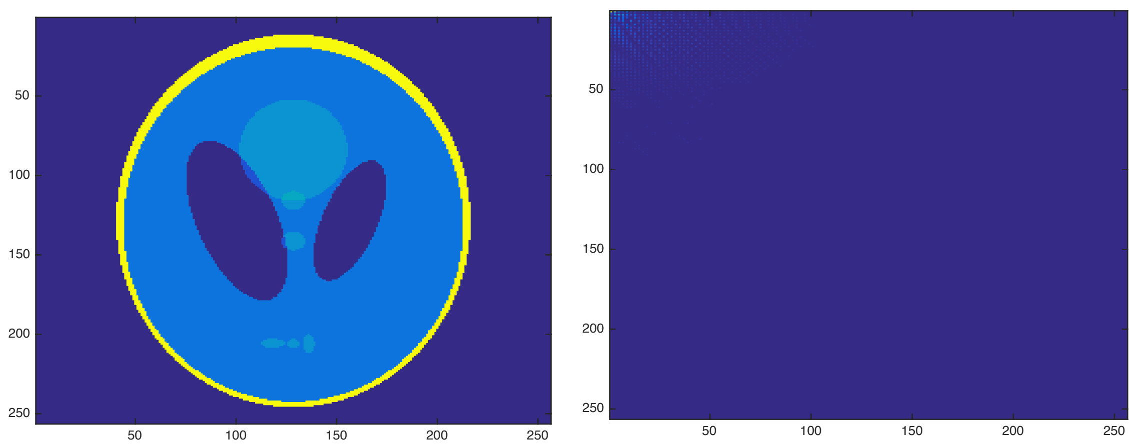

Figure 1: Shepp-Logan phantom (left), and the magnitude of its 2D discrete-cosine transform coefficients (right)

[2] used compressed sensing to accelerate magnetic resonance imaging (MRI) by massively under-sampling the Fourier domain. The paper shows that MRI images are sparse in many domains, including DCT, wavelet, and spatial finite-difference domains. Figure 1 demonstrates this: The left side of figure 1 the Shepp-Logan phantom, which is a schematic of the human brain used to compare medical imaging algorithms. The right image shows the magnitudes of the 2D discrete cosine transform coefficients. Evidently, the image is extremely sparse in this domain, whereas there are large regions with significant intensities in the image domain. Further, several decisions have to be made to apply compressed sensing in practice.

For example, the paper discusses how to incorporate prior knowledge of the sparsity pattern into a compressed sensing algorithm. In practice, many signals have large Fourier coefficients for small frequencies, while also a few critically important high frequency components. *Variable density sampling* incorporates this knowledge, by under-sampling the frequency domain less for small frequencies, and more for high frequencies. The paper shows that this method achieves superior performance in practice, when the frequencies are sampled according to a power law.

Further, there are multiple ways of taking a three dimensional MRI image: by subsequently scanning 2D slices, or by a pure 3D scan. The authors point out that randomly sub-sampling 3D space, instead of subsequent 2D images, reveals more of the image redundancy, capturing most of the advantage of compressed sensing.

Lastly, it is intriguing that the authors propose fixing a sampling pattern after finding one with good image reconstruction properties, as measured on a few sample images. This makes the algorithm deterministic and sacrifices the theoretical, probabilistic guarantees of the under-sampling scheme. The practical justification of this is that one expects the images of a particular body region to be sufficiently similar, in order for the good reconstruction behavior to apply to MRI scans of different people.

References

[1] S. Qaisar, R. M. Bilal, W. Iqbal, M. Naureen, and S. Lee. Compressive sensing: From theory to applications, a survey. *Journal of Communications and Networks*, 15(5):443–456, 10 2013.

[2] Michael Lustig, David Donoho, and John M. Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 12 2007.