In this note, we look at Compressed Sensing and more broadly, sparse recovery problems, trading technicality for simplicity.

1 Signals are Sparse and the World is Simple (That is, if you look at them the right way)

We begin our quest with an experiment. Here is a picture of my desk, I put it through a Wavelet Transform and threshold the resulting array retaining only the largest 5% of the coefficients (i.e., setting the rest to 0) and show the reconstructed image in the second picture. The difference between the first and second is noticeable, but subtle at best.

Figure 1: Original Image, Reconstruced Image with Wavelet Thresholding, Pixel-wise Difference Image

For the purpose of the note, the take-home message of the experiment is that data usually have patterns, and are therefore sparse and compressible, if we put them in the right representation. The details of how the Wavelet Transform sparsifies natural images will be left for another time, but for a brief intuition of why "representation" matters and why compression is hopeful, let's take cosine function as an example:

$$cos(2\pi t) = \frac{1}{2}(e^{i2\pi t} + e^{-i2\pi t});$$

if we naively index it with δ-functions as

$$\sum_i c_i \cdot \delta(t - t_i),$$

one would need an infinite sequence of bits for storing the pairs (t_i, c_i); and this is indeed how we view signals when we plot them. However, being a bit precognizant and recognizing the periodicity in the function, if we were to index instead by complex exponential functions

$$\sum_i c_i \cdot e^{i2\pi k_i t},$$

we would just need to know the pairs $(k_i, c_i) = \{(1, 1/2), (-1, 1/2)\}$ in this "transformed" domain to capture all the information, since the rest of the c_i's are zero. This is, in fact, the general philosophy behind image compression schemes such as JPEG – there is a representation of natural images where they can be stored and recovered with a small number of nonzero elements.

The curious may then wonder, if a photograph can (and usually will) be saved in a compressed form for much less storage space without any perceivable difference – why should the camera bother collecting all these extra information in the first place, only to be thrown away later?

2 From Sparse Signals to Sparse Sampling

Compressed sensing is a theory developed to only "measure the important bits". There are two components to the story – the first about what subsampling scheme to use for taking the measurement, and the second about how to recover the image given the collected data. But before we embark on the journey, let's give a motivating example where such a theory turns out to be tremendously useful. In MRI, the time it takes for scanning is proportional to the number of measurements that need to be taken for satisfying reconstruction. Therefore less measurements directly translate to speed-up in the process. The measurements are taken in the so-called K-space, where we get measurements of the type

$$y = Fx$$

for $F \in \mathbb{C}^{d\times d}$ a Fourier Transform matrix and $x \in \mathbb{R}^d$ a sparse signal we wish to recover. The goal is to take a subset of samples $F_u \in \mathbb{C}^{n\times d}$ for $n \ll d$ such that we can still reconstruct the desired x given $y_u \in \mathbb{R}^n = F_u x$. For the non-technical audience, it suffices to think of F as a fixed matrix, and our job is to come up with ways to sample as few number of rows as possible from the matrix, with the goal that the collected data will be enough for an algorithm to recover the sparse (in an appropriate domain) image.

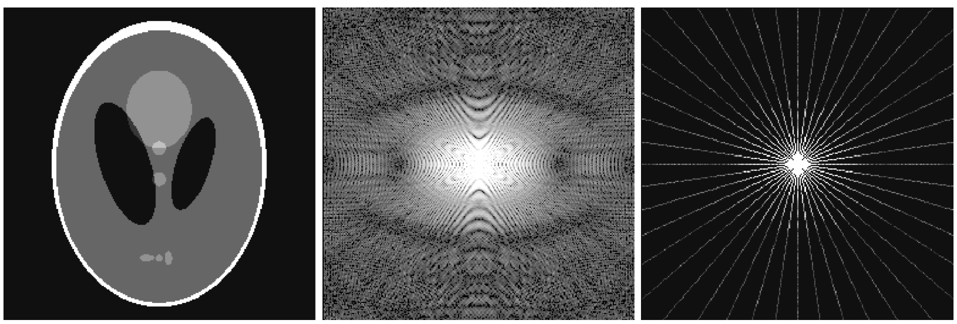

Figure 2: Original Image, Fourier Transform of Image, Subsampled Fourier Transform (dark indicates zero)

Surprising as it may seem in the experiment below, even though some information are lost in the subsampling (i.e., in general we can't hope to recover any arbitrary signal x), with the prior knowledge that the signal is sparse, as natural images tend to be, recovery is possible with much fewer number of measurements!

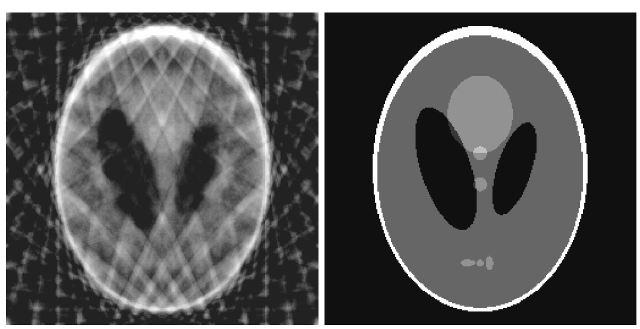

Figure 3: Classical Reconstruction, Compressive Sensing Reconstruction

Now we revisit the first question above and try to unravel the mystery – Is sparsity in signal alone enough for reconstruction or is there something special about the sampling matrix F_u that's required for its success? Let's use a simple example for illustration, which shows that sampling vectors cannot be sparse otherwise most of the measurements will be 0, i.e., contain no information.

$$\underbrace{\begin{bmatrix}0\\0\\0\\0\end{bmatrix}}_{\text{acquired measurement}} = \underbrace{\begin{bmatrix}1&&&&&&&&&\\&&&&&1&&&&\\&&&&&&1&&&\\&&&&&&&&&1\end{bmatrix}}_{\text{sensing matrix}} \underbrace{\begin{bmatrix}0\\0\\0\\*\\0\\0\\0\\*\\0\\0\end{bmatrix}}_{\text{sparse signal}}$$

Similarly, to capture information in different "parts" of the signal, so we don't end up measuring the same coordinate over and over again, we would like the row vectors to be diverse and spread-out as well. The theory of Compressed Sensing suggests that, in fact, random subsampling works great for these purposes. Effectively this turns an ill-posed problem, where the solution to $F_u x = y_u$ is non-unique (we're solving an underdetermined system of linear equations after all), into a sparse signal denoising problem, where the "noise" is induced by the random subsampling operation[1]. Therefore one may hope that if we manage to impose our prior knowledge about the sparsity of the signal, out of the many solutions consistent with the data there might exists *only one sparse solution*, corresponding to the true signal that we aim to recover.

3 "Out of Nothing I Have Created a Wonderful New Universe"

Now that we have the under-sampled measurements y_u, the question remains on how to retrieve the desired x. For this, the proposal is to solve the following optimization problem:

$$\min_x \; \|F_u x - y_u\|_2^2 + \lambda \|\Psi x\|_1 \tag{1}$$

for F_u the (randomly) subsampled measurement matrix, x the image we are trying to recover and Ψ a sparsifying transform e.g., Wavelet Transform. This objective function is finding among all solutions that satisfy the data constraints $F_u x = y_u$, the one with the minimum ℓ_1 norm after Wavelet Transform. The parameter $\lambda > 0$ aims to balance the two parts of the objective. While not being entirely obvious, ℓ_1 norm is known to encourage sparse solution, which makes sense for what we are after as we expect our signal to have relatively few number of non-zeros in the transformed domain. A pictorial illustration of how ℓ_1 norm promotes sparse optimal solution in 2D is given below, thanks in part to its diamond-shape level set.

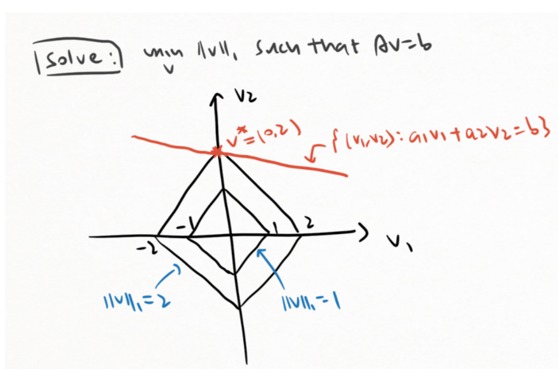

Figure 4: Sparse optimal solution at intersection of linear constraint and ℓ_1 norm level set

[1]The emphasis on *random* subsampling, as opposed to equispaced subsampling, is to make the noise random and white-like. Readers with background in signal processing will recognize that equispaced under-sampling scheme introduces structured noise where a portion of the information is constantly missing. An experiment is included in Section 4 for illustration.

This turns out to be a mathematical truth! We state below the theorem that combines all the necessary ingredients introduced so far.

Theorem (Informal)**.** If

- x has at most k non-zero terms;
- Fourier coefficients are selected *at random*,

solving (1) *exactly* recovers the true signal if number of measurements taken $>$ small multiple of k (i.e., the information content). Moreover, essentially one cannot do better with other measurements and/or algorithm choices.

What makes (1) all the more attractive for applications is the fact that there are very efficient algorithms to solve them to high-accuracy. Moreover, the reason the theory proves to be really useful in practice is that one could still recover for *approximately* sparse signal with noisy measurements, and the accuracy degrades "gracefully" as noise power increases. This general paradigm of recovering structured low-complexity signal from incomplete data turns out to find applications in many other problems across science and engineering domains and has inspired a whole new line of research. "There is nothing more practical than a good theory", once said Kurt Lewin, this story serves as a reminder of that.

4 Randomness is Too Important to be Left to Chance Alone

I hope to convince through the experiment below that sampling pattern plays an important role in practice and it's not just a theoretical construct. Below is a brain MRI scan, where a simple recovery scheme (i.e., inverse Fourier Transform) is performed on the undersampled data – one with equispaced sampling and the other with random sampling. The middle picture comes at no surprise to engineers, who are ingrained to take as an unfortunate fact of life that equispaced subsampling introduces noise that makes recovery inherently impossible. In contrast, the surprising finding is that random sampling (third picture) gives much better reconstruction, even with simple linear recovery operation. The magic of Compressed Sensing is that one can really get away with less measurements using (1) non-uniform sampling; (2) nonlinear reconstruction scheme as introduced in Section 3, under the (mostly harmless) assumption that the signal is sparse in an appropriate sense.

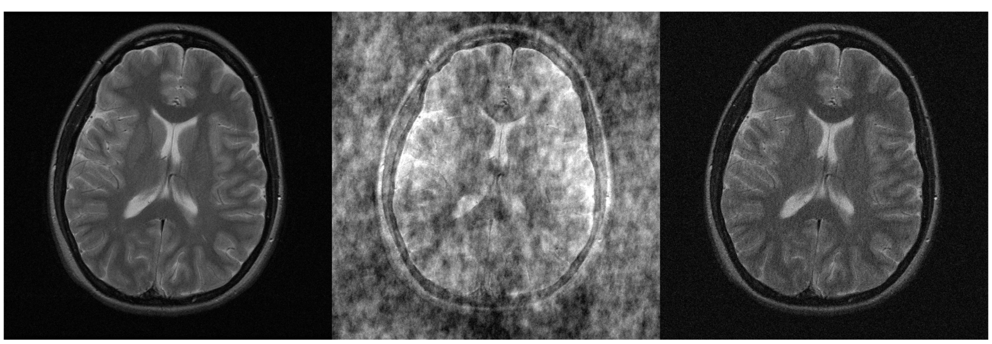

Figure 5: Original, naïve linear reconstruction from x3 equispaced (M) and randomly (R) subsampled data

I hope you enjoyed reading the note as much as I enjoyed writing it. – Qijia