Lab 6: Compressed sensing

Let us consider the following toy problem.

PROBLEM 1. We have a polynomial $p(z) = a_0 + a_1 z + \cdots + a_{N-1} z^{N-1}$ of degree $N - 1$ (with N very big) but we know that most of its coefficients are 0. That is, we know that only K coefficients are non zero and K is much smaller than N.

Suppose that we know N and K but we don't know the polynomial. Assume that we can evaluate the polynomial at any point. How many evaluations do we need and what is a possible scheme to recover the coefficients?

A first naive answer is that we need N samples since the polynomials is of degree $N-1$. But this is way too much because we are disregarding the fact that our polynomial is *sparse*. That is it has few non-vanishing coefficients.

On the other hand, if we know exactly which K coefficients are non zero, then with only K queries we could solve the problem just solving a $K \times K$ linear system.

We could try to get a bit more than K points, say M points and for all possible collections of K coefficients and for each of them we solve the $M \times K$ over determined system. It will compatible for one choice of coefficients.

This is too costly. Candès and Tao and Donoho proposed the following strategy:

- Consider the system $Ax = y$ where A is a $N \times M$ matrix. This is the system that we want to solve (it codifies the values of the polynomial at the points that we know). This is heavily undetermined because $M < N$ so there are many possible solutions.
- Choose the solution x that has minimal ℓ^1 norm.

Some remarks are in order: Since we know that there is a solution x which is sparse, ideally we would like to find the solution x such that $Ax = y$ has minimal $\|x\|_0$ norm (minimal number of non-zero coefficients). This, as we mentioned is very hard (it is an NP problem). On the other hand, the minimal ℓ^2 solution is easy to compute by a minimal squares technique, but it will be no sparse at all. A good compromise is the ℓ^1 solution because as you may expect from the geometry of the unit balls, it has many vanishing coefficients. See the figure, where the minimal solution to the system is computed in ℓ^2, ℓ^∞ and ℓ^1 norm.

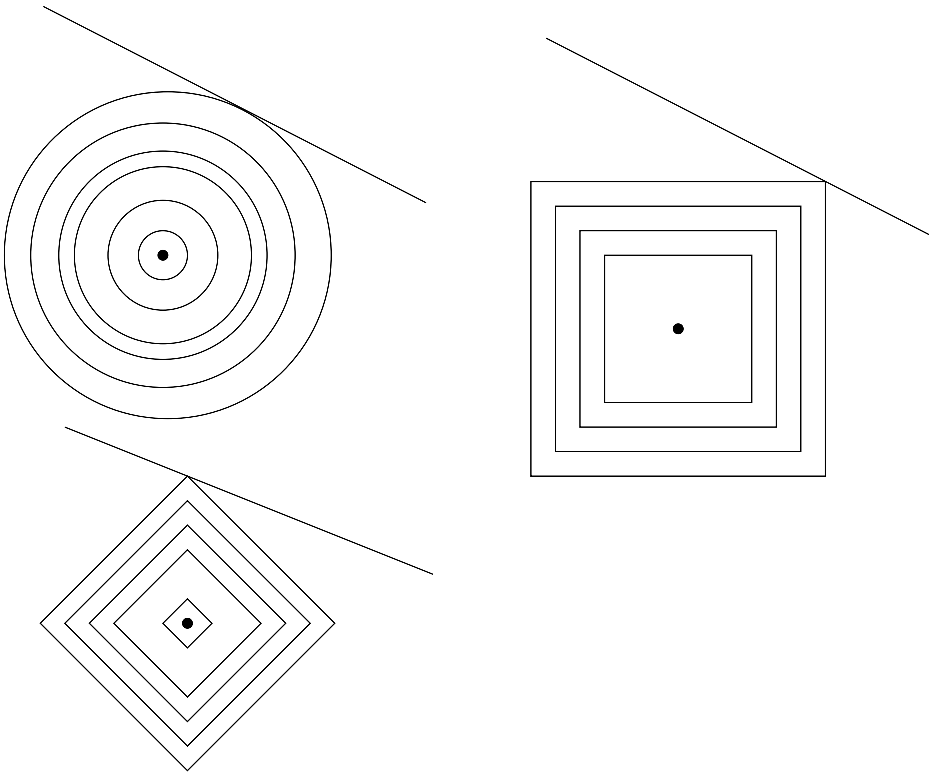

FIGURE 1. The line represents the manifold of solutions to $Ax = y$. In the first picture the tangent point is the solution with minimal ℓ^2 norm. The second picture is the solution with minimal ℓ^∞ norm and the last one is the solution with minimal ℓ^1 norm. This last one has one of the coefficients vanishing. This happens generically for many systems $Ax = y$.

This works most of the time, but it is important that the matrix A which represents the relationship between the basis where the signal is sparse and the measurements that we take, has to have a certain "incoherence". This is achieved for instance if we evaluate the polynomial among the roots of unity (randomly).

The second important observation is that the problem of minimizing $\|x\|_1$ subject to the restriction $Ax = y$ is easily reduced to a linear-programming problem, and this is a well studied problem with many good solutions.

We define $x = u - v$ where $u_j = x_j^+$ and $v_j = x_j^-$. So we want to minimize

$$\sum_{i=1}^{N} u_i + v_i,$$

subject to $A(u-v)=y$ and $u_j \geq 0, v_j \geq 0$. This can be solved with a standard linear programming solver as provided for instance with Matlab `linprog` or Octave `glpk`.

This is formalized in a result of Tao, Candès and Romberg that says the following:

THEOREM 1 (Candès, Romberg, Tao). *Suppose that we have a discrete finite signal of length N and a randomly chosen set of frequencies $\Omega \subset \{2\pi i/N\}_{i=0}^{N-1}$. Suppose that f is a superposition of $|T|$ shifted deltas, i.e.*

$$f[n] = \sum_{t \in T} f[t]\delta(n-t),$$

and

$$|T| \leq \frac{C_M}{\log N}|\Omega|,$$

for some constant $C_M > 0$. We do not know the locations of the delta, nor its amplitudes. Then, with probability $1 - O(N^{-M})$, f can be reconstructed exactly as the solution of the ℓ_1 minimization problem:

$$\min_g \sum_{k=0}^{N-1} |g[k]|, \qquad \text{such that } G(w) = F(w), \quad \forall \omega \in \Omega.$$

EXERCISE 1. *Make a program that solves problem 1 in the following sense: you hard code a sparse trigonometric polynomial (say $p(x) = 2\sin(x) + \sin(2x) + 21\sin(300x)$) and then try to guess it by evaluating it on real points only 30 or 40 times taken at random in $[0, 400]$. You only know that it is a trigonometric polynomial of degree smaller than 400 and with few components.*

This is a very flexible tool that allows to be used in a lot of different contexts. One particular interesting cases is the Radon transform. Here till the end I quote directly from Tao, Romberg and Candès:

Compressed sensing and the Radon transform

This idea is best motivated by an experiment with surprisingly positive results. Consider a simplified version of the classical 'tomography' problem in medical imaging: we wish to reconstruct a 2D image $f(t_1, t_2)$ from samples $\hat{f}|_\Omega$ of its discrete Fourier transform on a star-shaped domain Ω. Our choice of domain is not contrived; many real imaging devices can collect high-resolution samples along radial lines at relatively few angles. Figure 2(b) illustrates a typical case where one gathers 512 samples along each of 22 radial lines.

Frequently discussed approaches in the literature of medical imaging for reconstructing an object from 'polar' frequency samples are the so-called filtered backprojection algorithms. In a nutshell, one assumes that the Fourier coefficients at all of the unobserved frequencies are

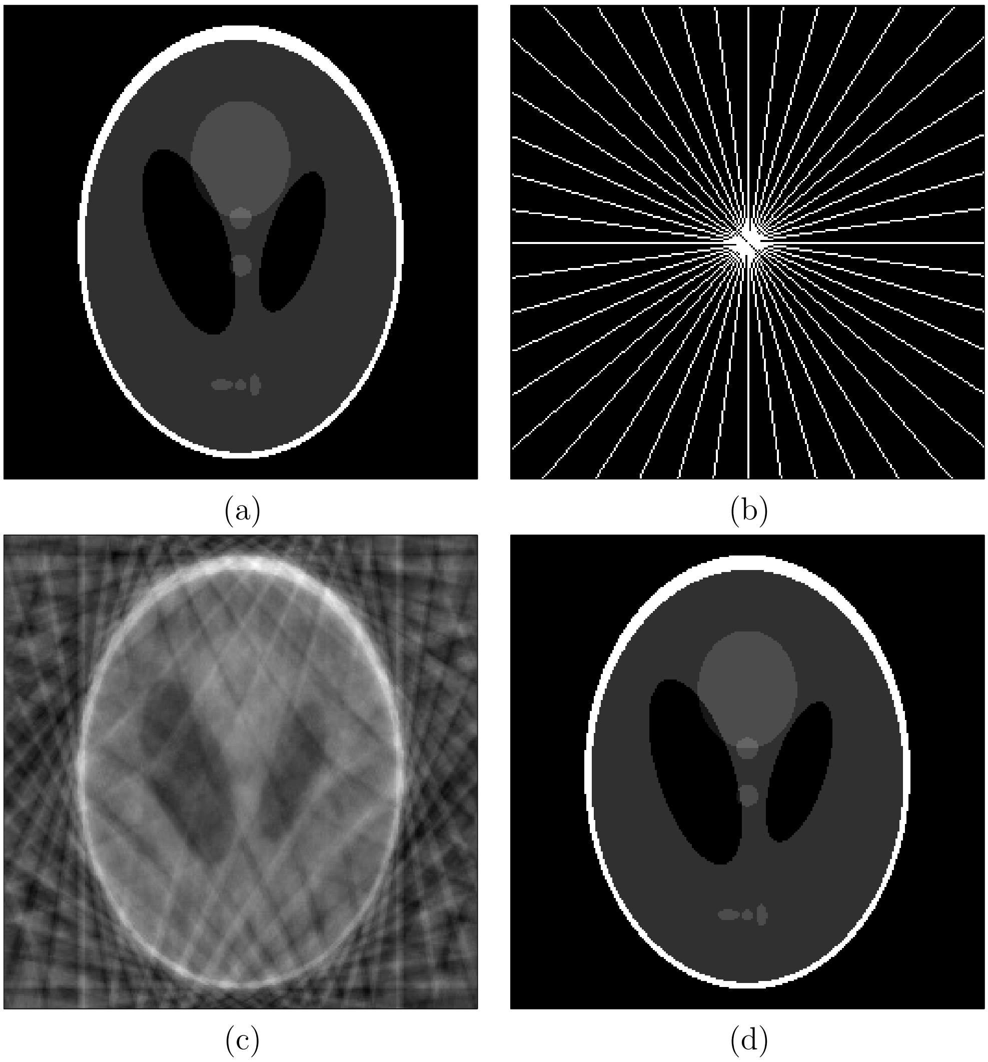

Figure 2. Example of a simple recovery problem. (a) The Logan-Shepp phantom test image. (b) Sampling 'domain' in the frequency plane; Fourier coefficients are sampled along 22 approximately radial lines. (c) Minimum energy reconstruction obtained by setting unobserved Fourier coefficients to zero. (d) Reconstruction obtained by minimizing the total-variation, as in (1). The reconstruction is an exact replica of the image in (a).

zero (thus reconstructing the image of "minimal energy" under the observation constraints). This strategy does not perform very well, and could hardly be used for medical diagnostic. The reconstructed image, shown in Figure 2(c), has severe nonlocal artifacts caused by the angular undersampling. A good reconstruction algorithm, it seems, would have to guess the values of the missing Fourier coefficients. In other words, one would need to interpolate $\hat{f}(\omega_1, \omega_2)$. This is highly problematic, however; predictions of Fourier coefficients from their neighbors are very delicate, due to the global and highly oscillatory nature of the Fourier transform. Going back to our example, we can see the problem immediately. To recover frequency information near (ω_1, ω_2), where ω_1 is near $\pm\pi$, we would need to interpolate $\hat{f}$ at the Nyquist rate $2\pi/N$. However, we only have samples at rate about $\pi/22$; the sampling rate is almost 50 times smaller than the Nyquist rate!

We propose instead a strategy based on convex optimization. Let $\|g\|_{BV}$ be the total-variation norm of a two-dimensional object g which for discrete data $g(t_1, t_2)$, $0 \leq t_1, t_2 \leq N-1$, takes the form

$$\|g\|_{BV} = \sum_{t_1, t_2} \sqrt{|D_1 g(t_1, t_2)|^2 + |D_2 g(t_1, t_2)|^2},$$

where D_1 is the finite difference $D_1 g = g(t_1, t_2) - g(t_1 - 1, t_2)$ and $D_2 g = g(t_1, t_2) - g(t_1, t_2 - 1)$. To recover f from partial Fourier samples, we find a solution $f^\sharp$ to the optimization problem

$$\min \|g\|_{BV} \qquad \text{subject to} \qquad \hat{g}(\omega) = \hat{f}(\omega) \text{ for all } \omega \in \Omega. \tag{1}$$

In a nutshell, given partial observation $\hat{f}_{|\Omega}$, we seek a solution $f^\sharp$ with minimum complexity—here Total Variation (TV)—and whose 'visible' coefficients match those of the unknown object f. Our hope here is to partially erase some of the artifacts classical reconstruction methods exhibit (which tend to have large TV norm) while maintaining fidelity to the observed data via the constraints on the Fourier coefficients of the reconstruction.

When we use (1) for the recovery problem illustrated in Figure 2 (with the popular Logan-Shepp phantom as a test image), the results are surprising. The reconstruction is *exact*; that is, $f^\sharp = f$! Now this numerical result is not special to this phantom. In fact, we performed a series of experiments of this type and obtained perfect reconstruction on many similar test phantoms.

EXERCISE 2. *For the very brave: Try to reproduce the numerical experiment suggested. There is a program to find the minimum the total variation subject to some linear constrains in the web page of E. Candès*

Bibliography

[1] J. Bruna Aproximació no lineal i mostratge comprimit *Butlletí de la SCM*, 30(2): 109–123, 2015.

[2] E. J. Candès, J. Romberg, T. Tao Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2): 489–509, 2006.

[3] D. Mackenzie Compressed Sensing Makes Every Pixel Count What is happening in the Mathematical Sciences., 7: 114–127, 2009 http://www.ams.org/publicoutreach/math-history/hap7-pixel.pdf