COMP0114 - Coursework 3

Due: Tuesday March 7th, 2023 at 16:00

Critical analysis:

At the time of writing, the paper has got 8486 citations, which is quite a lot for a research paper, we can already guess that this paper has had a big impact on the area of mathematical signal processing. Its authors are widely-cited too, Emmanuel Candès (147840), Justin Romberg (14420) and Terence Tao (89528), and have collaborated more than once, for example:

- "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information", EJ Candès, J Romberg, T Tao, IEEE Transactions on information theory 52 (2), 489-509 (2006)
- "Near-optimal signal recovery from random projections: Universal encoding strategies?", EJ Candes, T Tao, IEEE transactions on information theory 52 (12), 5406-5425 (2006)

which have respectively 18835 and 8277 citations. The topics of these publications being close to the paper we are studying, we can say that the latter has created almost a new field within signal processing, with novel standard methods for signal estimation.

We can quote one paper citing Candès et al. :

- "Deep-learning-based ghost imaging.", Lyu, Meng, et al, Scientific reports 7.1 (2017): 17865. (2017)

It introduces us to deep learning for ghost imaging. Ghost imaging is a technique that produces an image using two different light sources. We can imagine that we would benefit from taking fewer measurements given that we have already two light sources and that the acquisitions can be costly and long. That is where the idea of Compressed Sensing appears. As we have seen in class, it consists of limiting the number of measurements directly when measuring, and not during the data processing.

And actually, techniques like Compressed Sensing have been made possible thanks to advances in sparse reconstruction that our paper presents.

We also see citations from other areas of machine learning such as few-shots learning where the network generalizes based on a low number of data, and also in other areas of signal processing such as telecommunications, where it is in our interest to reduce the number of useful symbols to increase our bandwidth.

Overall, sparse techniques introduced by Candès et al. have made possible reconstruction and learning from incomplete data, and have made appear new methods in optimization, machine learning, and signal processing.

Let's now answer the following question: ***"How and under what circumstances can a signal f be exactly reconstructed from a discrete set of samples?"***

First of all, what about signals in general? We know we can decompose a signal into a sum of cosines of different frequencies (Fourier), and so reconstruct the signal from this basis of cosines. The problem we have is that usually, we do not have an infinite number of samples to determine the perfect reconstruction. We only have a limited number of observations, this is the discrete set we are talking about here.

Because we have a limited number of observations, we can guess it will be hard to reconstruct exactly a full signal. We would need a sparse signal i.e a signal with more null values than non-zero values. We can intuitively understand that seeing only peak values rather than continuous curves will be easier to reconstruct the signal: when we see a peak, we know a non-zero value is located there, but if we see a Gaussian/parabola/other type of curve, we will have trouble estimating exactly all the values in this interval.

Is it reasonable to assume that such sparse signals can be found? This paper, but also one of its references ([13] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. IEEE Trans. Inform. Theory, 47:2845–2862 (2001)), argues that there are always representations (e.g Fourier transform, Wavelets) in which a signal can be sparse. The latter even proves that this decomposition can be unique. Theorem 2 from our paper also lays out a condition for which approximately sparse signals can also be reconstructed stably. Overall, we can often find a way to obtain a sparse representation.

Once we obtain this sparse representation, we have to find a way to observe it. If we only pay attention to one part of the signal, it will not help us guess what is happening in the other parts. If we already knew the position of the non-zero values, it would be too easy (this is the oracle procedure shown in the paper in 1.3). So, random matrices or random Fourier ensembles can work, as long as they obey a "uniform uncertainty principle" (1.1).
Let's add that the way we build the observation matrix determines the accepted sparsity level of our signal.

But even with these conditions, an estimator like the least-squares would only average the result and give us a completely wrong reconstruction. We need to enforce sparsity, by adding the condition $\min \|x\|$. The l_0 norm counts the number of non-zero elements. The l_1 norm is used in convex programs where we can use the classical methods of linear programming.

Now that we know all of this, we have to select an observation matrix that is able to both give us a sparse representation of the signal, and give us a nice framework for observation (e.g an orthonormal basis), such as example 3 in 1.3. This matrix should give us enough room to estimate sparse signals. Once this preliminary work is done, we can use optimisation libraries to solve the convex problem (P_2).

To conclude, here are the main points to ensure the exact reconstruction of a signal from a discrete set of samples:

1. **Circumstances:** we need to know a basis in which the decomposition of the signal can be sparse, and find another matrix constructed in a way that it can observe enough samples of the signal
2. **How:** then, we construct a convex problem built to minimize the noise in the observations and to minimize the norm of our recovered signal. It can be solved using linear programming methods.

Numerical experiments:

Firstly, as recommended in the coursework sheet, we will create the A matrix, starting from two orthogonal bases (with MATLAB's `orth`). We're just left with designing the W matrix - the spectrum of A.

Then, following what is suggested in the paper (see 3rd bullet point in 1.3), we take W as the identity matrix (so that we sample the rows of U, which is an orthonormal matrix), the spectrum of A is, therefore, comprised of the same value (diagonal of ones):

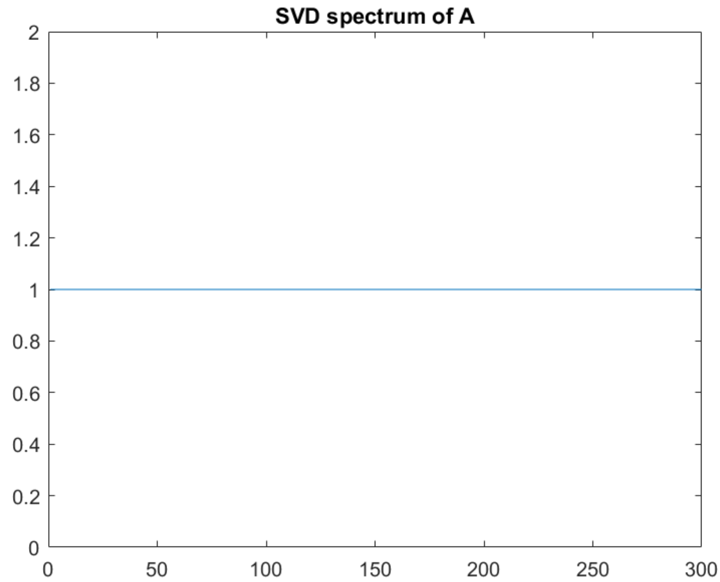


Figure 1: Spectrum of matrix A (well-conditioned)

We know from the previous courseworks that A is well-conditioned and that it can be inverted without too many numerical errors.

Let's now compute the Gaussian measurements:

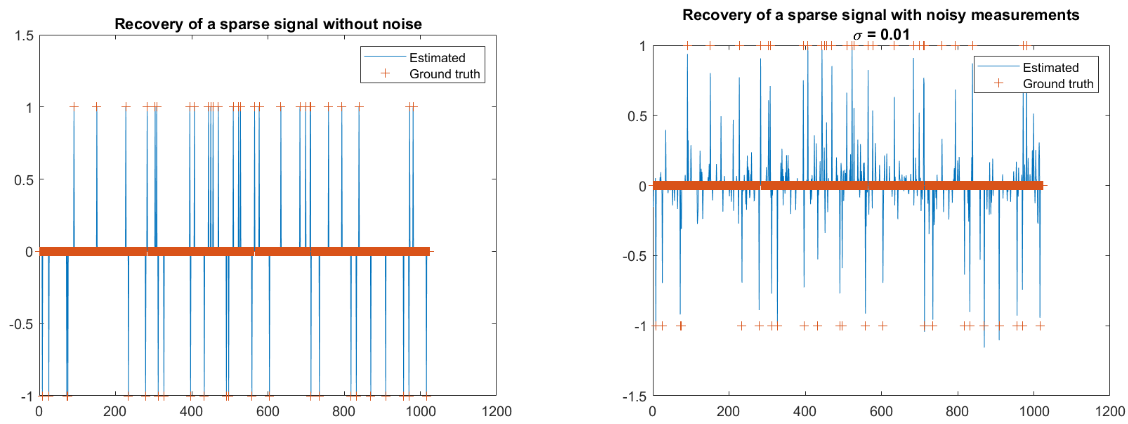
$$b = A * s + e$$

where s is our signal of length 1024 with 50 of its samples equal to either 1 or -1 (the others being null), and e is a Gaussian vector of mean 0 and variance σ of length 300 (number of measurements).

We use the l1-magic library to perform the optimization, in particular the function `l1eq_pd`. You can find the results of our reconstruction in Figure 2 below (without and with noise respectively).

We can see that the optimization routine manages to completely recover the sparse signal, which validates the theory from the paper. For the noisy case, we can clearly see estimated peaks at the location of the ground truth peaks, but the magnitude of the signal is not fully retrieved: the magnitude is rather diffused between the peaks and the noise between the peaks. Therefore, we can have a rough estimation of the sparse signal, but we need to have another process to determine with more precision the magnitude at each time step.

We can also conjecture that this reconstruction depends on the condition number of matrix A. Let's test this hypothesis by changing the spectrum of A by $\exp(-k/100)$ where k is the index of the diagonal. We can plot this spectrum:



(a) without noise

(b) with noisy measurements, for σ=0.01

Figure 2: Reconstruction of the sparse signal for a well-conditioned matrix

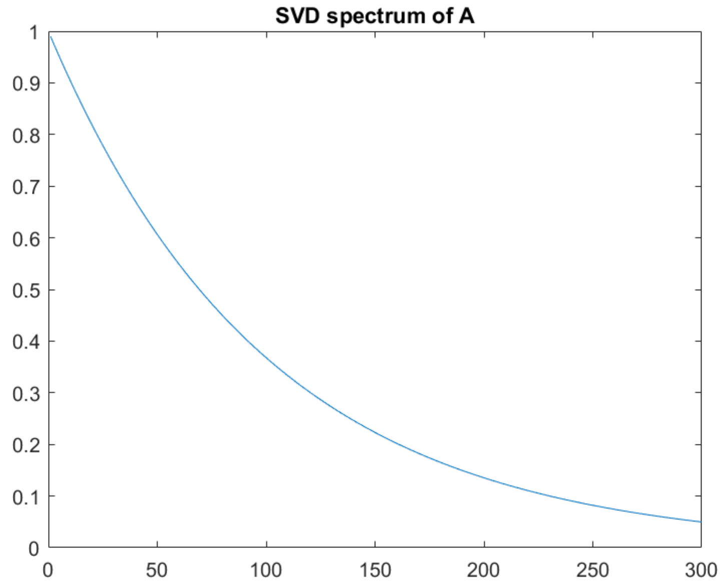


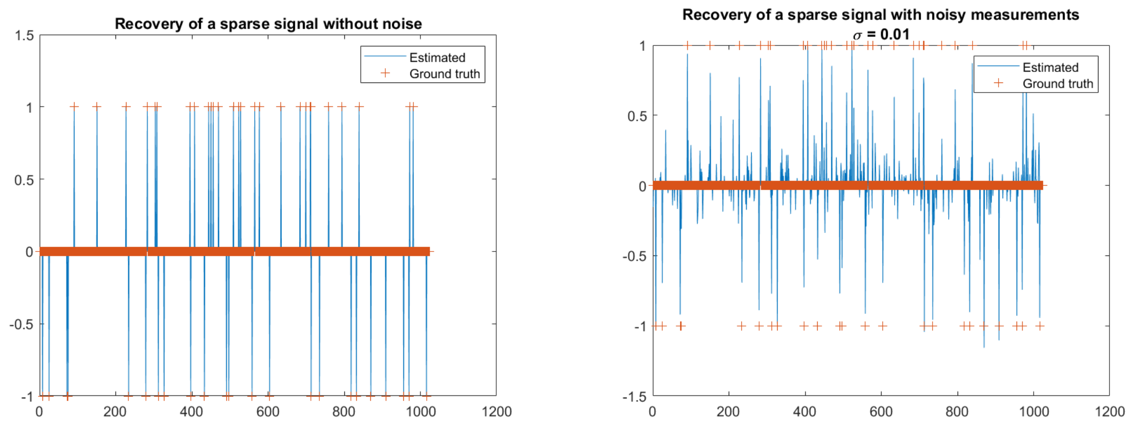
Figure 3: Spectrum of matrix A (ill-conditioned)

We repeat the previous routine on noiseless and noisy measurements, the corresponding graphs are shown in Figure 4 below.

The reconstruction from noiseless data is still precise, but the performance with noisy measurements has severely decayed. We have approximately the same reconstruction for the ill-conditioned at $\sigma = 0.01$ as the well-conditioned at $\sigma = 0.05$. Smaller changes in input beget bigger changes in output, the reconstruction is definitely less stable for the ill-conditioned matrix.

We have also run reconstructions for signals with more than 50 non-zero values, to check the assumption of the paper, found in Figures 5 to 7. We can clearly see that even doubling the initial number of non-zero values makes the recovery almost impossible.

Overall, our numerical experiments confirm the results from the paper and its assumptions about the observation matrix and the number of non-null samples.



(a) without noise (b) with noisy measurements, for σ=0.01

Figure 4: Reconstruction of the sparse signal for an ill-conditioned matrix

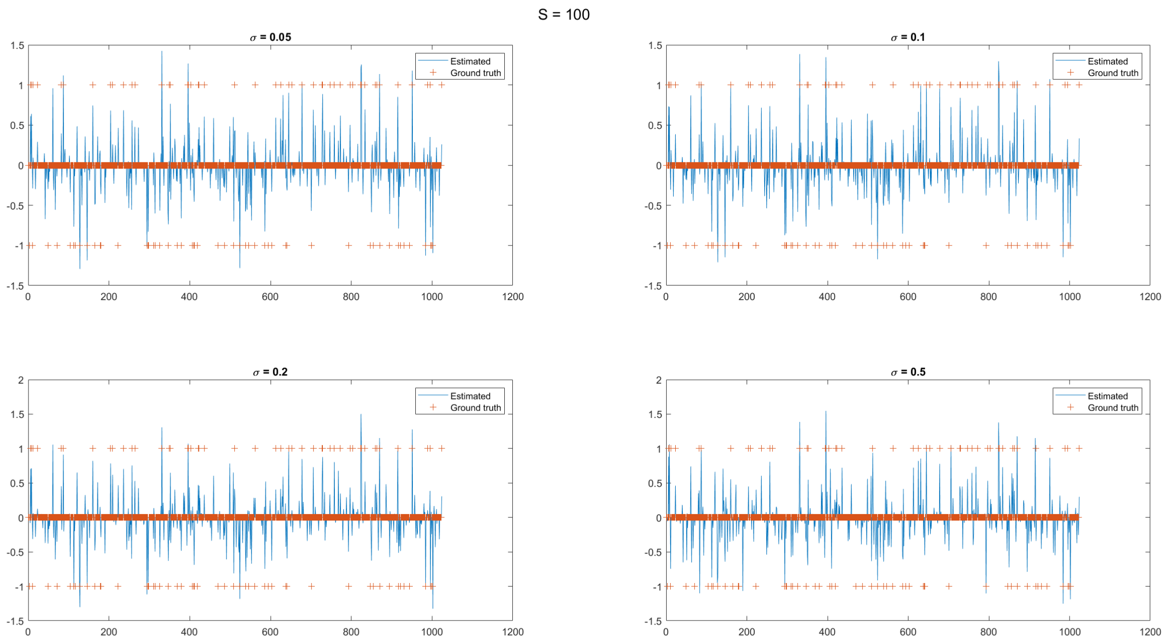


Figure 5: Reconstruction of the signal for 100 non-zero values, for a well-conditioned matrix, for different noise values

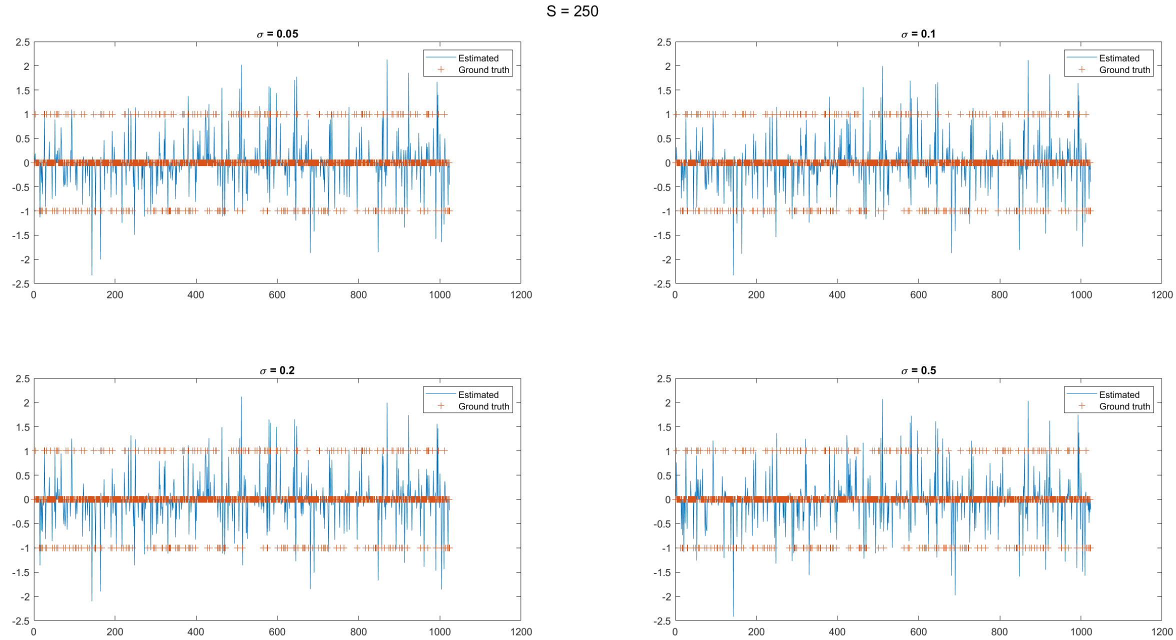


Figure 6: Reconstruction of the signal for 250 non-zero values, for a well-conditioned matrix, for different noise values

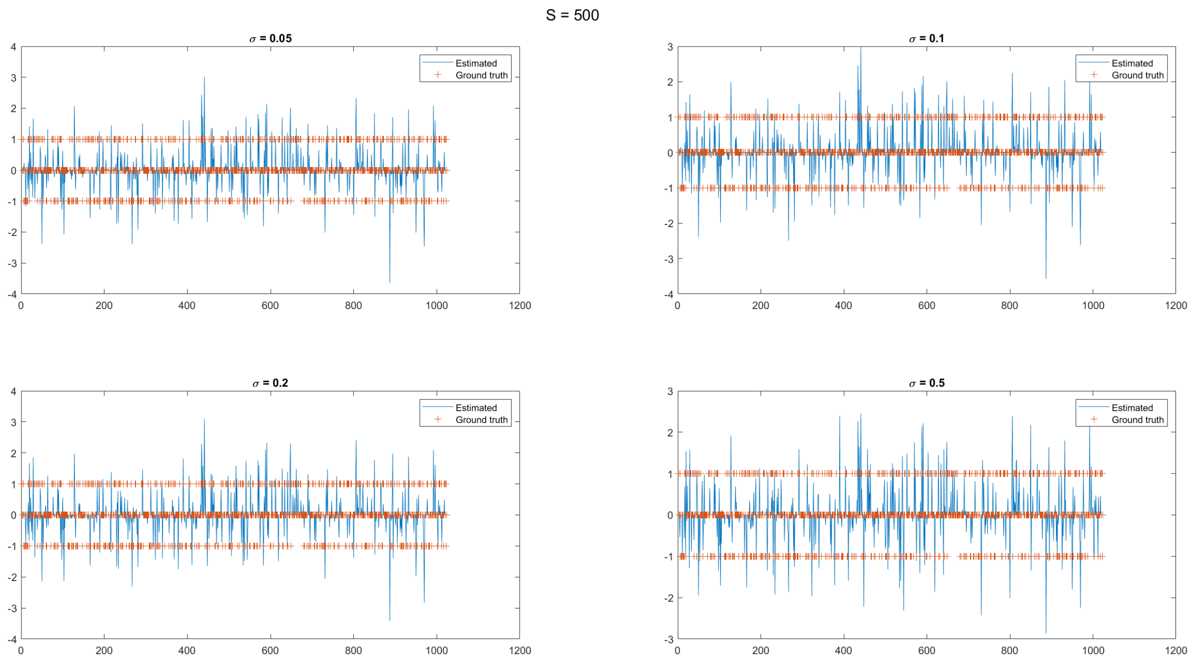


Figure 7: Reconstruction of the signal for 500 non-zero values, for a well-conditioned matrix, for different noise values