Compressed sensing

Or: the equation $Ax = b$, revisited

Terence Tao

University of California, Los Angeles

Mahler Lecture Series

Acquiring signals

- Many types of real-world signals (e.g. sound, images, video) can be viewed as an n-dimensional vector $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbf{R}^n$ of real numbers, where n is large (e.g. $n \sim 10^6$).
- To acquire this signal, we consider a linear measurement model, in which we measure an m-dimensional vector $b = Ax \in \mathbf{R}^m$ for some $m \times n$ measurement matrix A (thus we measure the inner products of x with the rows of A). For instance, if we are measuring a time series in the frequency domain, A would be some sort of Fourier matrix.

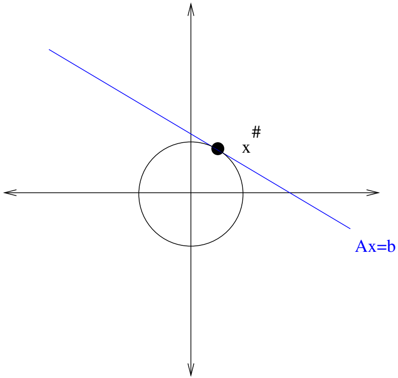
This leads to the following classical question in linear algebra:

- **Question:** How many measurements m do we need to make in order to recover the original signal x exactly from b? What about approximately?
- In other words: when can we solve the equation $Ax = b$?

The classical answer

The classical theory of linear algebra, which we learn as undergraduates, is as follows:

- If there are at least as many measurements as unknowns ($m \geq n$), and A has full rank, then the problem is determined or overdetermined, and one can easily solve $Ax = b$ uniquely (e.g. by gaussian elimination).
- If there are fewer measurements than unknowns ($m < n$), then the problem is underdetermined even when A has full rank. Knowledge of $Ax = b$ restricts x to an (affine) subspace of $\mathbf{R}^n$, but does not determine x completely.
- However, if one has reason to believe that x is "small", one can use the least squares solution $x^\# = \mathrm{argmin}_{x:Ax=b} \|x\|_{\ell^2} = A^*(AA^*)^{-1} b$ as the "best guess" for x.



A low-dimensional example of a least-squares guess.

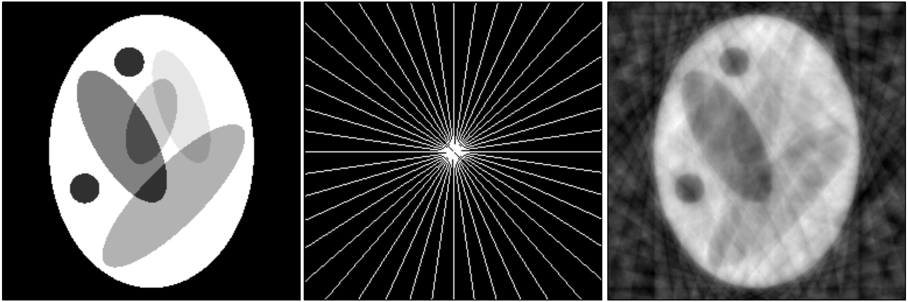
- In many situations the least squares solution is not satisfactory.
- For instance, consider the problem of reconstructing a one-dimensional discrete signal $f : \{1, \ldots, n\} \to \mathbf{C}$ from a partial collection $\hat{f}(\xi_1), \ldots, \hat{f}(\xi_m)$ of Fourier coefficients

$$\hat{f}(\xi) := \frac{1}{n} \sum_{x=1}^{n} f(x) e^{-2\pi i x \xi / n}.$$

The least squares solution $f^\#$ to this problem is easily seen to be the partial Fourier series

$$f^\# := \sum_{j=1}^{m} \hat{f}(\xi_j) e^{-2\pi i \xi_j x/n}$$

which, when m is small, is often very different from the original signal f, especially if f is "spiky" (consider for instance a delta function signal).



The Logan-Shepp phantom and its least squares reconstruction after Fourier-sampling along 22 radial lines (here $m/n \approx 0.05$). This type of measurement is a toy model of that used in MRI.

Sparse recovery

- It is thus of interest to obtain a good estimator for underdetermined problems such as $Ax = b$ in the case in which x is expected to be “spiky” - that is, concentrated in only a few of its coordinates.
- A model case occurs when x is known to be S-sparse for some $1 \leq S \leq n$, which means that at most S of the coefficients of x can be non-zero.

Why sparsity?

- Sparsity is a simple but effective model for many real-life signals. For instance, an image may be many megapixels in size, but when viewed in the right basis (e.g. a wavelet basis), many of the coefficients may be negligible, and so the image may be compressible into a file of much smaller size without seriously affecting the image quality. (This is the basis behind algorithms such as the JPEG2000 protocol.) In other words, many images are effectively sparse in the wavelet basis.
- More complicated models than sparse signals can also be studied, but for simplicity we will restrict attention to the sparse case here.

Sparsity helps!

- Intuitively, if a signal $x \in \mathbf{R}^n$ is S-sparse, then it should only have S degrees of freedom rather than n. In principle, one should now only need S measurements or so to reconstruct x, rather than n. This is the underlying philosophy of compressive sensing: one only needs a number of measurements proportional to the compressed size of the signal, rather than the uncompressed size.
- An analogy would be with the classic twelve coins puzzle: given twelve coins, one of them counterfeit (and thus heavier or lighter than the others), one can determine the counterfeit coin in just three weighings, by weighing the coins in suitably chosen batches. The key point is that the counterfeit data is sparse.

Compressed sensing is advantageous whenever

- signals are sparse in a known basis;
- measurements (or computation at the sensor end) are expensive; but
- computations at the receiver end are cheap.

Such situations can arise in

- Imaging (e.g. the “single-pixel camera”)
- Sensor networks
- MRI
- Astronomy
- ...

But can compressed sensing work?

- **Proposition:** Suppose that any $2S$ columns of the $m \times n$ matrix A are linearly independent. (This is a reasonable assumption once $m \geq 2S$.) Then, any S-sparse signal $x \in \mathbf{R}^n$ can be reconstructed uniquely from Ax.
- **Proof:** Suppose not; then there are two S-sparse signals $x, x' \in \mathbf{R}^n$ with $Ax = Ax'$, which implies $A(x - x') = 0$. But $x - x'$ is $2S$-sparse, so there is a linear dependence between $2S$ columns of A, contradiction. $\square$

- In fact, the above proof also shows how to reconstruct an S-sparse signal $x \in \mathbf{R}^n$ from the measurements $b = Ax$: x is the unique sparsest solution to $Ax = b$. In other words,

$$x = \operatorname{argmin}_{x:Ax=b} \|x\|_{\ell^0}$$

where

$$\|x\|_{\ell^0} := \sum_{i=1}^{n} |x_i|^0 = \#\{1 \leq i \leq n : x_i \neq 0\}$$

is the sparsity of x.

- Unfortunately, in contrast to the ℓ^2 minimisation problem (least-squares), ℓ^0 minimisation is computationally intractable (in fact, it is an NP-hard problem in general). In part, this is because ℓ^0 minimisation is not a convex optimisation problem.

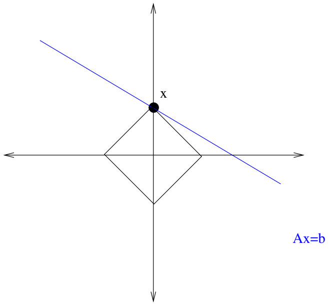
- To summarise so far: when solving an underdetermined problem $Ax = b$, ℓ^2 minimisation is easy to compute, but often wrong.
- When x is sparse, ℓ^0 minimisation is often correct, but very difficult to compute.
- Is there a way to split the difference?

Basis pursuit

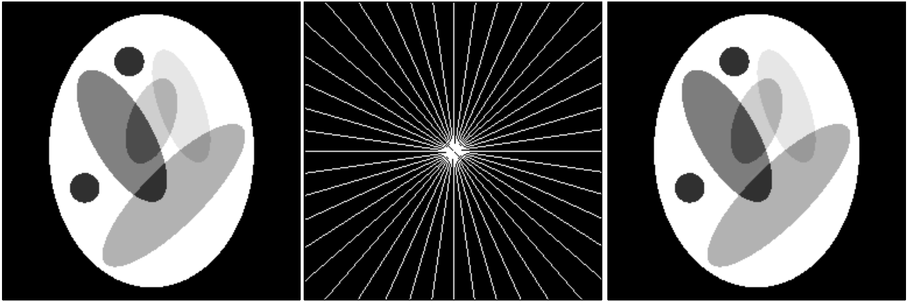
- A simple, yet surprisingly effective, way to do so is ℓ^1 minimisation or basis pursuit; thus, our guess $x^\sharp$ for the problem $Ax = b$ is given by the formula

$$x^\sharp = \operatorname{argmin}_{x: Ax=b} \|x\|_{\ell^1}.$$

- This is a convex optimisation problem and can be solved fairly quickly by linear programming methods (several specialised software packages are now also available). And... it works!



In this example, the signal x is reconstructed exactly from $b = Ax$ by ℓ^1 minimisation.



Exact reconstruction of the Logan-Shepp phantom from partial Fourier data by ℓ^1 minimisation (or more precisely, total variation minimisation, i.e. the ℓ^1 norm of the gradient).

- Basis pursuit was introduced empirically in the sciences (e.g. in seismology by Claerbout-Muir and others) in the 1970s, and then studied mathematically in the 1990s by by Chen, Donoho, Huo, Logan, Saunders, and others.
- Near-optimal performance guarantees emerged in the 2000s by Candés-Romberg-Tao, Donoho, and others.
- There are also several other compressed sensing algorithms known (e.g. matching pursuit and its refinements), but we will focus on basis pursuit here as it is relatively simple to state.

Theoretical results

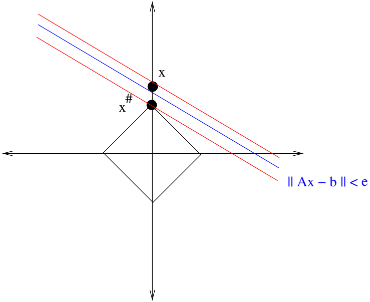
- There are now several theoretical results ensuring that basis pursuit works whenever the measurement matrix A is sufficiently “incoherent”, which roughly means that its matrix entries are uniform in magnitude. (It’s somewhat analogous to how the secret to solving the twelve coins problem is to weigh several of the coins at once.) Here is one typical result:
- **Theorem:** (Candés-Romberg-T. 2004). Let $\xi_1, \ldots, \xi_m \in \{1, \ldots, n\}$ be chosen randomly. Then with high probability, every S-sparse signal $f : \{1, \ldots, n\} \to \mathbf{C}$ can be recovered from $\hat{f}(\xi_1), \ldots, \hat{f}(\xi_m)$, so long as $m > CS \log n$ for some absolute constant C.

- Numerical experiments suggest that in practice, most S-sparse signals are in fact recovered exactly once $m \geq 4S$ or so.
- It turns out that basis pursuit is effective not only for Fourier measurements, but for a much wider class of measurement matrices. The necessary condition that every $2S$ columns of A have to be linearly independent, has to be strengthened somewhat (for instance, to the assertion that every $4S$ columns of A are approximately orthonormal). The precise condition used in the literature is called the Restricted Isometry Property (RIP), and is obeyed by many types of matrices (e.g. gaussian random matrices obey the RIP with high probability).

Variants and extensions

There are many variants and extensions of compressed sensing in the literature (200+ papers in the last 3 years!). Here is a quick sample:

- Compressed sensing algorithms such as basis pursuit can not only recover sparse data x exactly from $b = Ax$, but can also recover compressible data (data which is approximately sparse) approximately, by a slight modification of the algorithm.
- In a similar spirit, these algorithms are also robust with regards to noise: if one only has some approximate measurements $b = Ax + z$ of the signal x, where z is a noise vector (e.g. gaussian white noise), then basis pursuit will still recover a good approximation $x^\sharp$ to x.



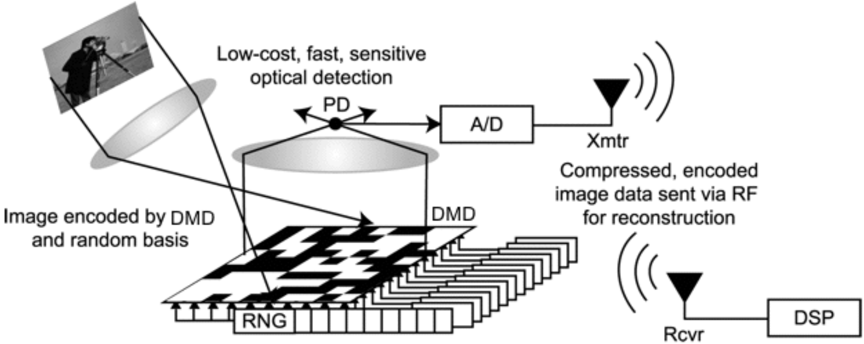
Reconstructing a sparse signal x approximately from noisy data $b = Ax + z$, assuming that z has norm less than some error tolerance e.

- There is a “dual” to compressed sensing, namely linear coding in which a signal $x \in \mathbf{R}^n$ is expanded into a larger signal $Ax \in \mathbf{R}^m$ (where now $m > n$ instead of $m < n$) to be transmitted over a noisy network. Even if parts of the transmitted signal are corrupted, so that the data received is of the form $b = Ax + e$ for some sparse e (representing packet loss or corruption), one can recover x exactly in many cases. (The trick is to view e, rather than x, as the signal, in which case one can convert things back to a compressed sensing problem.)

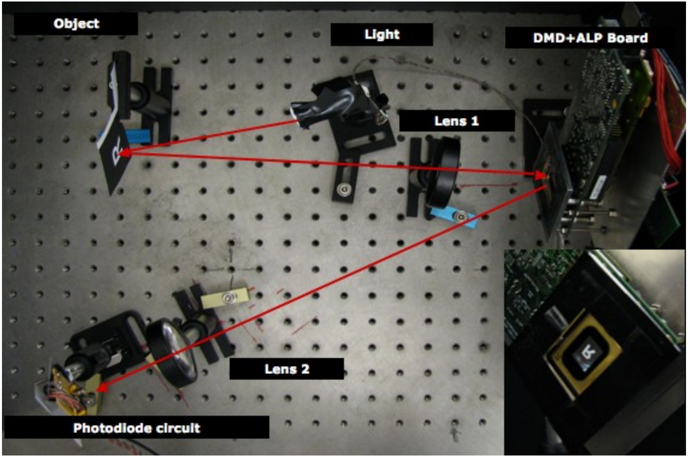
- There are also matrix versions of compressed sensing: instead of trying to reconstruct a sparse vector x from some measurements Ax, one instead tries to reconstruct a low-rank matrix M from some coefficients of that matrix. There is an analogue of basis pursuit (using the nuclear norm $\|M\|_1$ rather than the ℓ^1 norm) which is effective in many cases. This type of problem is relevant to real-life matrix completion problems (e.g. the Netflix prize).

Practical applications of compressed sensing

- Compressed sensing is a fairly new paradigm, but is already being used in practical settings, for instance to speed up MRI scans by requiring fewer measurements to achieve a given amount of resolution.
- One of the first prototype demonstrations of compressed sensing is the single pixel camera, developed by Rice University.

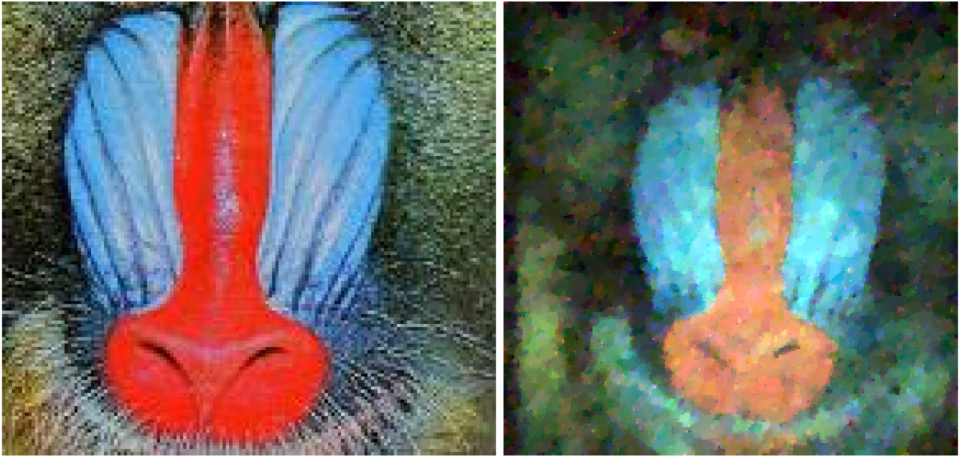


A schematic of the Rice single pixel camera
(`http://dsp.rice.edu/cscamera`)

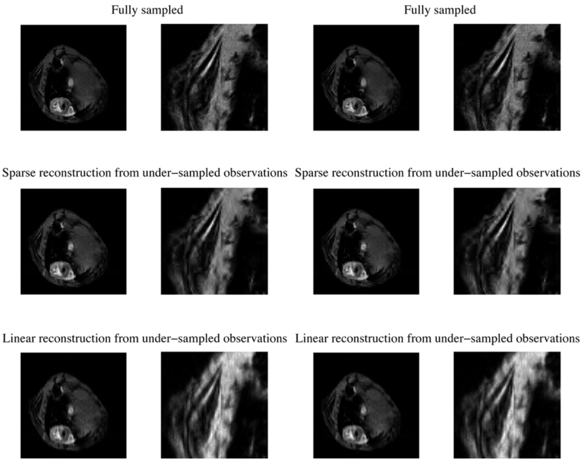


Physical implementation
(`http://dsp.rice.edu/cscamera`)

Test image (16384 pixels) and CS reconstruction using 1600 and 3300 measurements
(`http://dsp.rice.edu/cscamera`)



Test image (65536 pixels) and CS reconstruction using 6600 measurements (`http://dsp.rice.edu/cscamera`)



MRI image of a mouse heart, and CS reconstruction using 20% of available measurements (Blumensath-Davies)