Compressive Filtering

Tom Kealy

October 29, 2015

1 Introduction

The recent work of Candes and Tao [3] and Donoho [8] has established that many real-world signals can be effectively captured via a small number of random projections relative to the dimension of the signal. For example, a 5 megapixel image can be thought of as a vector in $\mathbb{R}^{5,000,000}$. However, it is well known that images have relatively few wavelet coefficients; this is exploited by the JPEG-2000 standard, which can represent the image as a 64-kb file (i.e a point in $\mathbb{R}^{64,000}$.

Classically, for perfect signal reconstruction, we must sample a signal such that the sampling rate must be at least twice the maximum frequency in the bandlimited signal. The continuous time signal can then be recovered using an appropriate reconstruction filter (e.g. a sinc filter). For example, we can represent a sampled continuous signal as a multiplication of the signal with a train of Dirac delta functions at multiples of the sampling period T.

$$x\left(nT\right) = \text{Ш}\left(t - nT\right) x\left(t\right) \tag{1.0.1}$$

where

$$\text{Ш}\left(t - nT\right) = \sum_{k=-\infty}^{\infty} \delta\left(t - kT\right) \tag{1.0.2}$$

Working the frequency domain, this multiplication becomes convolution (which is equivalent to shifting):

$$\hat{X}_s\left(f\right) = \sum_{k=-\infty}^{\infty} x\left(t - kT\right) \tag{1.0.3}$$

Thus if the spectrum of the frequency is supported on the interval $(-B, B)$ then sampling at intervals $\frac{1}{2B}$ will contain enough information to reconstruct the signal $x((t)$. Multiplying the spectrum by a rectangle function (low-pass filtering), to remove any images caused by the periodicity of the function, and the signal $x((t)$ can be reconstructed from its samples:

$$x\left(t\right) = \sum_{n=-\infty}^{\infty} x\left(nT\right) sinc\left(\frac{t_n T}{T}\right) \tag{1.0.4}$$

In contrast Compressive Sampling suggests that by adding randomness into the measurement process, a sparse (or compressible signal) may be accurately sensed with far fewer measurements: $y = Ax + w$

where $A \in \mathbb{R}^{m \times n}$ is a matrix with random entries, $x \in \mathbb{R}^n$ is the signal we capture, $y \in \mathbb{R}^m$ is the result of the measurement process and $w \sim N\left(0, 1\right) \in \mathbb{R}^m$ is additive white Gaussian noise, $m < n$.

Some technical conditions on the matrix A have to satisfied for it: namely the transformation defined by A must behave like an approximate Isometry, and it must be incoherent.

Definition 1.1 (RIP)**.** *We say that a matrix A satisifes the RIP of order δ if $\exists$ a $\delta \in (0,1)$ such that:*

$$(1-\delta)\,||x||_2^2 \leq ||Ax||_2^2 \leq (1=\delta)\,||x||_2^2 \tag{1.0.5}$$

i.e. A approximately preserves the lengths of all s-sparse vectors in $\mathbb{R}^n$.

Definition 1.2 (Coherence)**.** *The mutual coherence of a matrix A is the absolute normalised inner product between different columns from A. Denoting the k-th column in A by a_k, the mutual coherence is given by:*

$$\mu(A) = \max_{1\leq i,j\leq n, i\neq j} \frac{|\langle a_i^T, a_j\rangle|}{||a_i||_2\,||a_j||_2} \tag{1.0.6}$$

This implies that sensing with incoherent systems is good, and efficient mechanisms ought to acquire correlations with random waveforms (e.g. white noise).

Theorem [3] Fix x $\in \mathbb{R}^n$ with a sparse coefficient basis, x_i in ψ. Then a reconstruction from m random measurements is possible with probability $1-\delta$ if:

$$m \geq C\mu^2(A) S \log\left(\frac{n}{\delta}\right) \tag{1.0.7}$$

where $\mu(A)$ is the coherence of the two bases, and S is the number of non-zero entries on the support of the signal.

In this new sensing paradigm, the complexity is shifted to the reconstruction process, where with high probability Donoho proved [7], that the minimiser of the program:

$$\arg\min_x \frac{1}{2}\,||y-Ax||_2^2 + \lambda\,||x||_1 \tag{1.0.8}$$

coincides with the sparsest solution to the under-determined system of linear equations. Thus we are able to sense sparse signals with random waveforms, and reconstruct them via linear programming.

However, signal reconstruction is not the only interesting signal processing task. Filtering, classification, detection, and estimation are also required in real world systems. For these tasks it was thought that signal reconstruction must be performed first, and then classical signal processing techniques could be brought to bear on the reconstruction.

There is some tension in this idea however: since the measurement matrix is an approximate isometry, some (as yet unspecified) operations on the measurements y should correspond to inference tasks (such as filtering and estimation) on x. This means that performing inference needn't require the reconstruction of the signal.

The papers [4] and [5] provide an introductory answer for the cases of filtering, detection, classification and estimation.

The structure of this document is as follows: sections (2), is a literature review of relevant material from compressed sensing, Wishart matrices, and maximum likelihood estimation of uncompressed signals in noise. Section (3) gives an overview of the problem of estimating a signal from a known set of basis functions.

2 Preliminaries

2.1 RIP and Stable Embeddings

Given a signal $x \in \mathbb{R}^n$, a matrix $A \in \mathbb{R}^{m\times n}$ we can acquire the signal via the set of linear measurements:

$$y = Ax \tag{2.1.9}$$

where in this case A represents the sampling system. In contrast to classical sensing, which requires that $m = n$ for there to be no loss of information, it is possible to reconstruct x from an under-determined set of measurements as long as x is sparse in some basis.

To make this precise, we define Σ_s as the set of s-sparse signals in $\mathbb{R}^n$:

Definition 2.1.

$$\Sigma_s = \{x \in \mathbb{R}^n : \text{supp}\,(x) \leq s\} \tag{2.1.10}$$

where $\text{supp}\,(x)$ *is the set of indices on which* x *is non-zero.*

Definition 2.2 (RIP)**.** *We say that a matrix* A *satisifes the RIP of order* s *if there exists a* $\delta \in (0,1)$ *such that for all* $x \in \Sigma_s$*:*

$$(1-\delta)\,||x||_2^2 \leq ||Ax||_2^2 \leq (1+\delta)\,||x||_2^2 \tag{2.1.11}$$

i.e. A approximately preserves the lengths of all s*-sparse vectors in* $\mathbb{R}^n$*.*

Remark 2.3 (Information Preservation)**.** *A necessary condition to recover all* s*-sparse vectors from the measurements* Ax *is that* $Ax_1 \neq Ax_2$ *for any pair* $x_{/} = x_2$*,* $x_1, x_2 \in \Sigma_s$*, which is equivalent to* $||A\,(x_1 - x_2)||_2^2 > 0$*.*

This is guaranteed as long as A *satisfies the RIP of order* $2s$ *with constant* δ *- as the vector* $x_1 - x_2$ *will hve at most* $2s$ *non-zero entries, and so will be distinguishable after multiplication with* A*. To complete the argument take* $x = x_1 - x_2$ *in definition* (2.2)*, guaranteeing* $||A\,(x_1 - x_2)||_2^2 > 0$*, and requiring the RIP order of* A *to be* $2s$*.*

Remark 2.4 (Stability)**.** *[?] We also require that the dimensionality reduction of compressed sensing is the preservation of relative distances: that is if* x_1 *and* x_2 *are far apart in* $\mathbb{R}^n$ *then their projections* Ax_1 *and* Ax_2 *are far apart in* $\mathbb{R}^m$*. This will guarantee that the dimensionality reduction is robust to noise.*

A requirement on the matrix A that satisfies both of these conditions is the following:

Definition 2.5 (δ-stable embedding)**.** *We say that a mapping is a* δ*-stable embedding of* $U, V \subset \mathbb{R}^n$ *if*

$$(1-\delta)\,||u-v||_2^2 \leq ||Au - Av||_2^2 \leq (1+\delta)\,||u-v||_2^2 \tag{2.1.12}$$

for all $u \in U$ *and* $v \in V$*.*

Remark 2.6. *Note that a matrix* A*, satisfying the RIP of order* $2s$ *is a* δ*-stable embedding of* Σ_s, Σ_s*.*

Remark 2.7. *Definition 2.5 has a simple interpretation: the matrix* A *must approximately preserve Euclidean distances between all points in the signal model* Σ_s*.*

2.2 Random Matrix Constructions

To construct matrices satisfying definition 2.5, given m, n we generate A by A_{ij} being i.i.d random variables from distributions with the following conditions [4]

Condition 1 (Norm preservation)**.** $\mathbb{E}A_{ij}^2 = \frac{1}{m}$

Condition 2 (sub-Gaussian)**.** $\mathbb{E}\left(e^{A_{ij}t}\right) \leq e^{C^2t^2/2}$

Random variables A_{ij} satisfying conditions (1) and (2) satisfy the following concentration inequality [1], [6][Lemma 6.1]:

Condition 3 (sub-Gaussian)**.**

$$\mathbb{P}\left(|\;||Ax||_2^2 - ||x||_2^2\;| \geq \varepsilon\,||x||_2^2\right) \leq 2e^{-cM\varepsilon^2} \tag{2.2.13}$$

Then in [1] the following theorem is proved:

Theorem 2.8. *Suppose that m, n and $0 < \delta < 1$ are given. If the probability distribution generating A satisfies condition* (2.2.13)*, then there exist constants c_1, c_2 depending only on δ such that the RIP* (2.2) *holds for A with the prescribed δ and any $s \leq \frac{c_1 n}{\log n/s}$ with probability $\geq 1 - 2e^{-c_2 n}$*

For example, if we take $A_{ij} \sim \mathcal{N}(0, 1/m)$, these conditions are satisfied.

2.3 Wishart Matrices

Let $\{X_i\}_{i=1}^r$ be a set of i.i.d $1 \times p$ random vectors drawn from the multivariate normal distribution with mean 0 and covariance matrix H.

$$X_i = (x_1, \ldots, x_p) \sim N(0, H) \tag{2.3.14}$$

We form the matrix X by concatenating the r random vectors into a $r \times p$ matrix.

Definition 2.9 (Wishart Matrix)**.** *Let*

$$W = \sum_{j=1}^{r} X_j X_j^T = XX^T \tag{2.3.15}$$

Then W has the Wishart distribution with parameters

$$W_r(H, p) \tag{2.3.16}$$

where p is the number of degrees of freedom.

Remark 2.10. *This distribution is a generalisation of the Chi-squared distribution: let $m = H = 1$.*

Theorem 2.11 (Expected Value)**.**

$$\mathbb{E}(W) = rH \tag{2.3.17}$$

Proof.

$$\begin{aligned}
\mathbb{E}(W) &= \mathbb{E}\left(\sum_{j=1}^{r} X_j X_j^T\right) \\
&= \sum_{j=1}^{r} \mathbb{E}\left(X_j X_j^T\right) \\
&= \sum_{j=1}^{r} \left(\mathrm{Var}(X_j) + \mathbb{E}(X_j)\mathbb{E}(X_j^T)\right) \\
&= rH
\end{aligned}$$

Where the last line follows as X_j is drawn from a distribution with zero mean. □

Remark 2.12. *The matrix $M = A^T A$, where A is constructed by the methods from section 2.2, will have a Wishart distribution. In particular, it will have $\mathbb{E}M = \frac{1}{m} I_n$*

2.4 Maximum Likelihood estimation: non-compressive case

Consider a received signal y, composed of a deterministic signal s corrupted by noise n (assumed to have zero mean and unit variance), i.e.

$$y = s + n \tag{2.4.18}$$

We assume that s can be expanded in some orthogonal basis, and that we have access to the basis functions $\{\phi_i\}_{i=1}^n$:

$$s = \sum_{i=1}^{n} \alpha_i \phi_i \tag{2.4.19}$$

We can write the likelihood for y down as s is deterministic: y is a Gaussian random variable with mean s:

$$f\left(y \mid s\right) = \left(\frac{1}{\sqrt{2 * \pi}}\right)^n \exp\left(\frac{\left(y-s\right)^T \left(y-s\right)}{2}\right) \tag{2.4.20}$$

Maximising this is equivalent to maximising:

$$lnf = ||y||_2^2 + 2\langle y, s\rangle + ||s||_2^2 \tag{2.4.21}$$

i.e by maximising the inner product $\langle y, s\rangle$.
Given y we can also estimate s, by calculating

$$\begin{aligned} \langle y, \phi_i\rangle \ &= \sum_{j=1}^{n} \alpha_j \phi_j^T \phi_i + n^T \phi_i && (2.4.22)\\ &= \alpha_j + \varepsilon && (2.4.23) \end{aligned}$$

That is, the maximum likelihood estimate of s is

$$\langle y, \phi_i\rangle \tag{2.4.24}$$

3 Compressive Estimation

When we receive the signal $y = Ax$, where $A \in \mathbb{R}^{m\times n}$, $x = \sum_{i=1}^n \alpha_i \phi_i$ and $A_{ij} \sim \mathcal{N}\left(0, 1/m\right)$, we can instead calculate ([5], [2]):

$$\hat{\Theta} = \arg\max_i y^T (A\phi_i) \tag{3.0.25}$$

Note that this

$$\begin{aligned} \mathbb{E}\big(y^T A\phi_i\big) &= y^T \mathbb{E}\big(A^t A\big)\phi_i && (3.0.26)\\ &= \frac{1}{m}\langle y, \phi_i\rangle && (3.0.27) \end{aligned}$$

Where we have used $y = A(s+n)$, in the first line along with linearity of expectanal, and in the third line we have used $\mathbb{E}\big(A^T A\big) = cI_n$, a standard assumption from Compressive sensing theory.

3.1 Example: Single Spike

A signal $s \in \mathbb{R}^{300}$ composed of a single (random) delta function, with coefficients drawn from a Normal distribution (with mean 100, and variance 1).

i.e

$$s = \alpha_i \delta_i \tag{3.1.28}$$

with

$$a_i \sim \mathcal{N}\left(100, \sigma^2\right) \tag{3.1.29}$$

and the index i chosen uniformly at random from $[1, n]$.

The signal was measured via a random Gaussian matrix $A \in \mathbb{R}^{100 \times 300}$, with variance $\sigma^2 = 1/100$ and the inner product between $y = As$ and all 300 delta functions projected onto $\mathbb{R}^{100}$ (i.e. $A\delta_j \forall j = 1 \dots n$) was calculated.

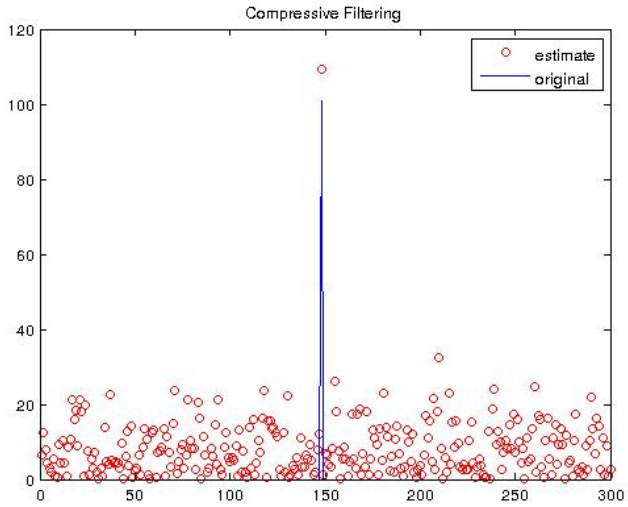

Figure 3.1:

References

[1] Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, 2008.

[2] Martin Braun, Jens P Elsner, and Friedrich K Jondral. Signal detection for cognitive radios with smashed filtering. In *Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th*, pages 1–5. IEEE, 2009.

[3] Emmanuel J Candes, Justin Romberg, and Terence Tao. Robust Uncertainty Principles : Exact Signal Frequency Information. 52(2):489–509, 2006.

[4] Mark Davenport, Petros T Boufounos, Michael B Wakin, Richard G Baraniuk, et al. Signal processing with compressive measurements. *Selected Topics in Signal Processing, IEEE Journal of*, 4(2):445–460, 2010.

[5] Mark A Davenport, Marco F Duarte, Michael B Wakin, Jason N Laska, Dharmpal Takhar, Kevin F Kelly, and Richard G Baraniuk. The smashed filter for compressive classification and target recognition. In *Electronic Imaging 2007*, pages 64980H–64980H. International Society for Optics and Photonics, 2007.

[6] Ronald Devore, Guergana Petrova, and Przemyslaw Wojtaszczyk. Instance-optimality in probability with an 1-minimization decoder. *Appl. Comput. Harmon. Anal*, pages 275–288, 2009.

[7] David L Donoho. Neighborly polytopes and sparse solution of underdetermined linear equations. 2004.

[8] David L Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006.