

Compressive Filtering

Tom Kealy

May 30, 2016

1 Classical Filtering and Compressive Sensing

Compressive Sensing has established that non-uniform sampling with random waveforms can be used to capture low-complexity signals, at rates far below those specified by Nyquist. The work of Candes and Tao [3] and Donoho [8] has established that this sensing framework is universal: the sensing mechanism and reconstruction algorithms are insensitive as to whether the signal is an image, a song, or frequency spectra as long as it is sparse. The computational effort in these systems is backloaded to the reconstruction algorithm, which can have considerable computational complexity. This is in contrast to traditional sampling [11], where the processing required to obtain the digital samples is the difficult part. Classical reconstruction is simply interpolation through an appropriate filter.

There are some technical specifications that on the compressive sensing mechanism which means that it can achieve this dimensionality reduction. Firstly the sensing waveforms must be incoherent with the signals being sensed [2]. This is a statement similar to the T_s -spaced delta function requirement from classical sensing: we need to sample the signal in a domain where its information is spread out. As a signal cannot be sparse in two bases simultaneously, we can exploit the uncertainty relation between the bases to sample the signal efficiently. The other requirement on the sensing waveforms is that they are orthogonal, and change the norm of the signal by very little. In particular, they must not send the signal to 0, or recovery would be impossible.

The structure of this document is as follows: sections (1.3), is a literature review of relevant material from compressed sensing, Wishart matrices, and maximum likelihood estimation of un-compressed signals in noise. Section (2) gives an overview of the problem of estimating a signal from a known set of basis functions.

1.1 Classical Sampling

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 [11]. 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 .

$$x(nT) = \text{III}(t - nT) x(t) \quad (1.1.1)$$

where

$$\text{III}(t - nT) = \sum_{k=-\infty}^{\infty} \delta(t - kT) \quad (1.1.2)$$

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

$$\hat{X}_s(f) = \sum_{k=-\infty}^{\infty} x(t - kT) \quad (1.1.3)$$

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:

$$x(t) = \sum_{n=-\infty}^{\infty} x(nT) \operatorname{sinc}\left(\frac{t_n T}{T}\right) \quad (1.1.4)$$

1.2 Maximum Likelihood estimation: non-compressive case

Consider a received signal $y \in \mathbb{R}^n$, composed of a deterministic signal $\bar{s} \in \mathbb{R}^n$ corrupted by noise $n \in \mathbb{R}^n$ (assumed to have zero mean and unit variance), i.e.

$$y = s + n \quad (1.2.5)$$

We assume $s(\Theta)$ comes from a fixed class of signals, with parameters indexed by a set Θ . For example

- The signal s is composed of a single frequency with unit amplitude $s = e^{i\omega_0 k}$. In this case $\Theta = \{\omega_j\}_{j=1}^n$ is the set of possible frequencies the signal may take on.
- The signal s is a scaled and shifted version of a model signal f : $s(t) = Cf(t - \tau)$. In this case $\Theta = (C, \tau)$.
- The signal s can be expanded in some orthonormal basis, and that we have access to the basis functions $\{\phi_i\}_{i=1}^n$:

$$s = \sum_{i=1}^n \alpha_i \phi_i \quad (1.2.6)$$

In this case $\Theta = \{\alpha_i, \phi_i\}_{i=1}^n$

We can write the likelihood for y down as, for a given Θ , s is deterministic. Therefore y is a Gaussian random variable with mean s :

$$f(y | s) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{(y - s(\Theta))^T (y - s(\Theta))}{2}\right) \quad (1.2.7)$$

Maximising this is equivalent to maximising:

$$\ln f(\Theta) = -\|y\|_2^2 + 2\langle y, s(\Theta) \rangle - \|s\|_2^2 \quad (1.2.8)$$

Since the terms $\|y\|_2^2$ and $\|s\|_2^2$ do not change, we can write the ML estimate of Θ as:

$$\hat{\Theta} = \arg \max_{\Theta} \langle y, s(\Theta) \rangle \quad (1.2.9)$$

So, for example:

- If we receive a single tone corrupted by noise $y = e^{i\omega_0 k} + n$ then we can estimate ω_0 by calculating $\arg \max_{\omega_i} \langle y, e^{i\omega_j k} \rangle = e^{i\omega_0 k}$ as the functions $e^{i\omega_0 j k}$ are orthonormal: $\langle e^{i\omega_l k}, e^{i\omega_r k} \rangle = \delta_{lr}$.
- If we receive a signal composed of a sum of orthonormal basis functions we can estimate the coefficients α_i as:

$$\langle y, \phi_i \rangle = \sum_{j=1}^n \alpha_j \phi_j^T \phi_i + n^T \phi_i \quad (1.2.10)$$

$$= \alpha_i + \varepsilon_i \quad (1.2.11)$$

where $\varepsilon_i = \langle n^T, \phi_i \rangle$ is some small error. Thus the maximum likelihood estimate of s is:

$$\hat{s} = \sum_{i=1}^n \hat{\alpha}_i \phi_i \quad (1.2.12)$$

where

$$\hat{\alpha} = \langle y, \phi_i \rangle \quad (1.2.13)$$

1.3 Compressive Sensing Preliminaries

In contrast to Classical Sensing, Compressive Sampling suggests that by adding randomness into the measurement process, a sparse (or compressible signal) may be accurately sensed with far fewer measurements:

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:

$$y = Ax \quad (1.3.14)$$

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 1.1.

$$\Sigma_s = \{x \in \mathbb{R}^n : \text{supp}(x) \leq s\} \quad (1.3.15)$$

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

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

Definition 1.2 (RIP). We say that a matrix A satisfies the RIP of order s if there exists a $\delta \in (0, 1)$ such that for all $x \in \Sigma_s$:

$$(1 - \delta) \|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta) \|x\|_2^2 \quad (1.3.16)$$

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

Definition 1.3 (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:

$$\mu(A) = \max_{1 \leq i, j \leq n, i \neq j} \frac{|\langle a_i^T, a_j \rangle|}{\|a_i\|_2 \|a_j\|_2} \quad (1.3.17)$$

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

Remark 1.4 (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 $x_1 \neq x_2$, $x_1, x_2 \in \Sigma_s$, 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 have 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 (1.2), guaranteeing $\|A(x_1 - x_2)\|_2^2 > 0$, and requiring the RIP order of A to be $2s$.

Remark 1.5 (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 1.6 (δ -stable embedding). *We say that a mapping is a δ -stable embedding of $U, V \subset \mathbb{R}^n$ if*

$$(1 - \delta) \|u - v\|_2^2 \leq \|Au - Av\|_2^2 \leq (1 + \delta) \|u - v\|_2^2 \quad (1.3.18)$$

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

Remark 1.7. *Note that a matrix A , satisfying the RIP of order $2s$ is a δ -stable embedding of Σ_s, Σ_s .*

Remark 1.8. *Definition 1.6 has a simple interpretation: the matrix A must approximately preserve Euclidean distances between all points in the signal model Σ_s .*

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:

$$m \geq C\mu^2(A)S \log\left(\frac{n}{\delta}\right) \quad (1.3.19)$$

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:

$$\arg \min_x \frac{1}{2} \|y - Ax\|_2^2 + \lambda \|x\|_1 \quad (1.3.20)$$

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.

Reconstruction isn't the only signal processing task: it is often part of a chain of operations. For example, we may require high frequency components of the received signal to be filtered out, or for timing and phase synchronisation in demodulation. Using Compressive sensing in these chains is less attractive: current methodology is to reconstruct the signal and then process the Nyquist samples using traditional techniques. The processing chain is then limited by the speed of the reconstruction step.

The foundational ideas of compressive sensing suggest, however, that reconstruction may be unnecessary. Compressive samples contain (almost) all the information about the signal, and are acquired via a linear process. This suggests that there may be some operations directly on the samples which can be used to extract information about the original signal, but which avoids costly reconstruction. I.e. for signal processing tasks where some other metric is the end result of the processing chain (e.g. classification), reconstructing Nyquist samples is optimal. This has

an added benefit: it may be possible to further reduce the sampling rate, as we are no longer constrained by the requirement to reconstruct the samples.

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

1.4 Random Matrix Constructions

To construct matrices satisfying definition 1.6, 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). $\mathbb{E}(e^{A_{ij}t}) \leq e^{C^2 t^2 / 2}$

Lemma 1.9 (Concentration). *There exists constants, for all t , such that random variables A_{ij} satisfying conditions (1) and (2) satisfy the following concentration inequality [1], [6][Lemma 6.1]:*

$$\mathbb{P}\left(\left|\|Ax\|_2^2 - \|x\|_2^2\right| \geq \varepsilon \|x\|_2^2\right) \leq 2e^{-cM\varepsilon^2} \quad (1.4.21)$$

Then in [1] the following theorem is proved:

Theorem 1.10. *Suppose that m, n and $0 < \delta < 1$ are given. If the probability distribution generating A satisfies condition (1.4.21), then there exist constants c_1, c_2 depending only on δ such that the RIP (1.2) 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 $A_{ij} \sim \mathcal{N}(0, 1/m)$, then the matrix A will satisfy the RIP

1.5 Wishart Matrices

Let X be a an $n \times p$ random matrix, each row of which is drawn from the p -variate normal distribution with mean 0 and covariance matrix H .

$$X_i = (x_1^{(i)}, \dots, x_p^{(i)}) \sim N(0, H) \quad (1.5.22)$$

Definition 1.11 (Wishart Matrix [9]). *Let*

$$W_n = X^T X \quad (1.5.23)$$

Then $W \in \mathbb{R}^{p \times p}$ has the Wishart distribution with parameters

$$W_p(H, n) \quad (1.5.24)$$

where p is the number of degrees of freedom.

Remark 1.12. *This distribution is a generalisation of the Chi-squared distribution: let $p = H = 1$.*

Theorem 1.13 (Expected Value).

$$\mathbb{E}(W) = rH \quad (1.5.25)$$

Proof.

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

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

Remark 1.14. The matrix $M = A^T A$, where A is constructed by the methods from section 1.4, will have a Wishart distribution. In particular, it will have $\mathbb{E}M = \frac{1}{m}I_n$

The joint distribution of the eigenvalues is given by [10]:

$$p(\lambda_1, \dots, \lambda_r) = c_r \prod_{i=1}^r e^{-\lambda_i} \prod_{i < j} (\lambda_i - \lambda_j)^2 \quad (1.5.26)$$

The eigenvectors are uniform on the unit sphere in \mathbb{R}^r .

2 Compressive Estimation

In this section, we develop some intuition into constructing estimators for the signal s directly on the compressive measurements:

Theorem 2.1. Given a set of measurements of the form:

$$y = As + n \quad (2.0.27)$$

where $A \in \mathbb{R}^{m \times n}$, $A_{ij} \sim \mathcal{N}(0, 1/m)$, and $n \in \mathbb{R}^n$ is AWGN i.e. $\sim N(0, \sigma^2 I)$. We again assume that s comes from a fixed set of models, parametrised by some set Θ .

Then, the maximum likelihood estimator of s , for the case where s can be expanded in an orthonormal basis $s = \sum_{i=1}^n \alpha_i \phi_i$:

$$\hat{s} = \sum_{i=1}^n m \langle y, A \phi_i \rangle \phi_i \quad (2.0.28)$$

Proof. The likelihood for this model is, (as y is a normal random variable):

$$f(y | s) = \left(\frac{1}{(2\pi)^{n/2}} \right) \exp \left(-\frac{(y - As)^T (y - As)}{2} \right) \quad (2.0.29)$$

Taking the logarithm and expanding, we find

$$\ln f(y | s) = -y^T y - s^T A^T A s + 2 \langle y, As \rangle + c \quad (2.0.30)$$

which is equal to:

$$\ln f = -\|y\|_2^2 - \|As\|_2^2 + 2 \langle y, As \rangle \quad (2.0.31)$$

(where the constant has been dropped). The first term of (2.0.31) is constant, for the same reasons as in section (2). The term

$$\|As\|_2^2 = \langle As, As \rangle \quad (2.0.32)$$

can be written as

$$\langle A^T As, s \rangle \quad (2.0.33)$$

We will assume that $A^T A$ concentrates in the sense of 1.4.21 and replace 2.0.33 with its expectation $\mathbb{E}(\langle A^T As, s \rangle)$

$$\begin{aligned} \mathbb{E}(\langle A^T As, s \rangle) &= \mathbb{E} \sum_{i=1}^n (A^T As)_i^T s_i \\ &= \sum_{i=1}^n \mathbb{E}(A^T As)_i s_i \\ &= \sum_{i=1}^n \left(\frac{1}{m} e_i s_i \right)_i^T s_i \\ &= \frac{1}{m} \langle s, s \rangle \end{aligned}$$

because

$$\mathbb{E} A^T A = \frac{1}{m} I \quad (2.0.34)$$

as it is a Wishart matrix (see section 1.3).
So we can further approximate (2.0.31):

$$\ln f(y | s) = -\|y\|_2^2 - \frac{1}{m} \|s\|_2^2 + 2\langle y, As \rangle \quad (2.0.35)$$

The only, non-constant part of (2.0.35) is the third term and so we define the estimator:

$$\hat{s} = \arg \max_{\Theta} \langle y, As(\Theta) \rangle \quad (2.0.36)$$

□

Corollary 2.2. *Consider the case where $y = As$ (no noise). Then*

$$y^T A \phi_j = \sum_i \alpha_i \phi_i^T A^T A \phi_j$$

So

$$y^T A \phi_j = \sum_i \alpha_i \phi_i^T A^T A \phi_j \sim \frac{\alpha_i}{m} \delta_{ij}$$

giving

$$\hat{\alpha}_i = m (y^T A \phi_j) \quad (2.0.37)$$

Remark 2.3. The matrix $M = A^T A$ is the projection onto the row-space of A . It follows that $\|Ms\|_2^2$ is simply the norm of the component of s which lies in the row-space of A . This quantity is at most $\|s\|_2^2$, but can also be 0 if s lies in the null space of A . However, because A is random, we can expect that $\|Ms\|_2^2$ will concentrate around $\sqrt{m/n} \|s\|_2^2$ (this follows from the concentration property of sub-Gaussian random variables (1.4.21)).

Example 2.4. *Example: Single Spike* We illustrate these ideas with a simple example: estimate which of n frequencies s is composed of.

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

$$s = \alpha_i \delta_i \quad (2.0.38)$$

with

$$a_i \sim \mathcal{N}(100, 1) \quad (2.0.39)$$

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} was calculated:

$$\hat{\alpha}_j = m \langle (A\alpha_i \delta_i), A\delta_j \rangle \quad (2.0.40)$$

We plot the $\hat{\alpha}_j$ below, figure 2.1, (red circles), with the original signal (in blue, continuous line). Note how the maximum of the $\hat{\alpha}_j$, coincides with the true signal.

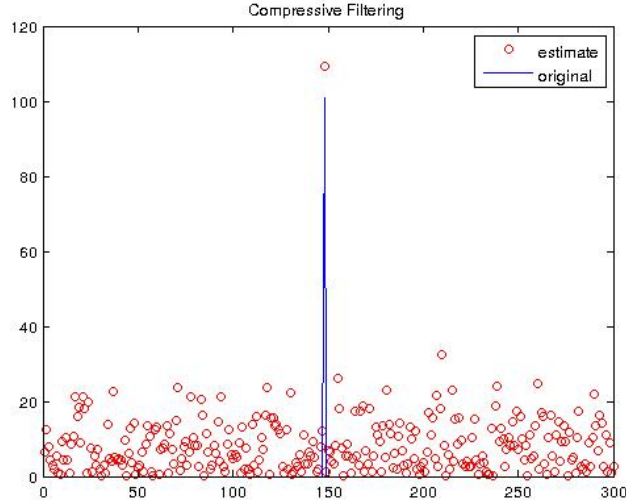


Figure 2.1:

3 Estimating Frequency Spectra

3.1 Estimating a single rectangle

We show how to estimate a signal, composed of a single rectangle (3.2) expanded in the following basis

$$l_i(x) = \begin{cases} 1 & \text{if } x \leq i \\ 0 & \text{otherwise} \end{cases} \quad (3.1.41)$$

That is, l_i is a left-hand step function.

The basis (3.1.41) can be expressed as a matrix in $\mathbb{R}^{n \times n}$ as:

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \dots 0 \\ 1 & 1 & 0 & 0 & 0 \dots 0 \\ 1 & 1 & 1 & 0 & 0 \dots 0 \\ \dots & & & & \\ 1 & 1 & 1 & 1 & 1 \dots 1 \end{pmatrix} \quad (3.1.42)$$

By direct computation, this inverse of L is:

$$D = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \dots 0 \\ -1 & 1 & 0 & 0 & 0 \dots 0 \\ 0 & -1 & 0 & 0 & 0 \dots 0 \\ \dots & & & & \\ 0 & 0 & 0 & 0 \dots -1 & 1 \end{pmatrix} \quad (3.1.43)$$

We model our PSD signal g as a linear combination of the basis functions (3.1.41):

$$g(x) = \sum_i a_i l_i(x) = L^T a \quad (3.1.44)$$

where $a = (a_1, \dots, a_n)$ are the coefficients in this basis expansion, and l_i are the rows of L . Note that as defined, g is a column vector.

Proposition 3.1.

$$D^T g = a \quad (3.1.45)$$

Proof.

$$D^T g = D^T L^T a \quad (3.1.46)$$

$$= (LD)^T a \quad (3.1.47)$$

$$= a \quad (3.1.48)$$

as $LD = I$. □

We define the inner product between two vectors as follows:

Definition 3.2.

$$\langle x, y \rangle = x^T y = \sum_i x_i y_i \quad (3.1.49)$$

where x_i, y_i are the components of the vectors x, y in the i^{th} direction with respect to some basis vectors e_i .

We can define a matrix representation for the set of basis vectors l_i , by taking all inner products between all pairs basis vectors:

Definition 3.3.

$$F_{n,ij} = \langle l_i, l_j \rangle \quad (3.1.50)$$

This matrix has the representation:

$$F_{ij} = \min(i, j) \quad (3.1.51)$$

An example of such a matrix is:

$$F_n = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 & 3 \\ 1 & 2 & 3 & 4 & 4 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix} \quad (3.1.52)$$

This matrix is invertible.

Theorem 3.4.

$$\det(F_n) = 1 \quad (3.1.53)$$

Proof. Consider the matrix F_n . Subtract the $n-1$ th column from the n th. We obtain a matrix with 0 on the final column except the entry $F_n(n, n) = 1$. Since the top $(n-1) \times (n-1)$ is F_{n-1} we find that

$$\det(F_n) = 1 \times \det(F_{n-1}) \quad (3.1.54)$$

By recursion and $\det(F_1) = 1$ we have $\det(F_n) = 1$. □

This matrix can be factorised as $F = LL^T$.

From this it follows that

$$F^{-1} = \begin{pmatrix} 2 & -1 & 0 & 0 & 0 \dots 0 \\ -1 & 2 & -1 & 0 & 0 \dots 0 \\ 0 & -1 & 2 & -1 & 0 \dots 0 \\ \dots & & & & \\ 0 & 0 & 0 & 0 \dots -1 & 1 \end{pmatrix} \quad (3.1.55)$$

To find the a_i , we correlate (take the inner product of) the signal against the basis (3.1.41).

Definition 3.5.

$$h_j = \langle g, l_j \rangle \quad (3.1.56)$$

$$= \sum_j g(x) l_j(x) \quad (3.1.57)$$

$$= \sum_j a_i l_i(x) l_j(x) \quad (3.1.58)$$

$$= a_i \langle l_i, l_j \rangle \quad (3.1.59)$$

$$\left(= \sum_{x=1}^j g(x) \right) \quad (3.1.60)$$

In matrix language $h = Fa^T$. This is the inner product between the signal g and the basis functions l_i .

Given a set of compressive measurements:

$$y = Ag \quad (3.1.61)$$

where $A \in \mathbb{R}^{m \times n}$, $A_{ij} \sim \mathcal{N}(0, 1/m)$.

We then compute

$$\langle y, Al_i \rangle = a_j l_j^T A^T A l_j \sim \frac{a_j}{m} \langle l_j, l_i \rangle \quad (3.1.62)$$

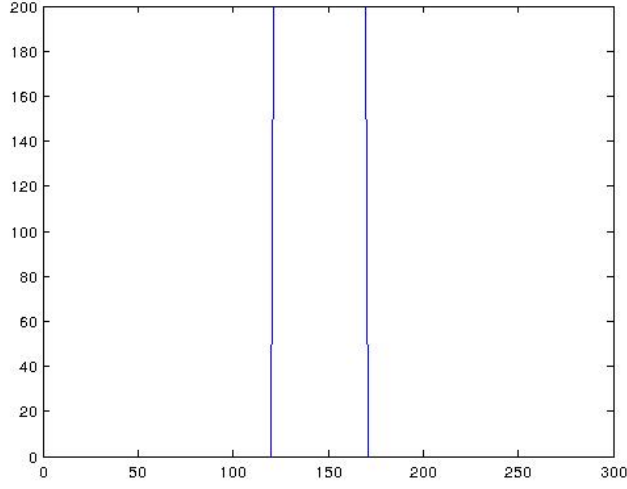


Figure 3.2:

for the set of basis vectors $l_1 \dots l_n$ i.e. the estimator from the previous section, corresponding to this set of basis functions (2.0.36).

We then form the vector

$$\hat{h} = m \sum_i \langle y, Al_i \rangle l_i \sim h \quad (3.1.63)$$

An example can be seen in figure 3.3, for a matrix $A \in \mathbb{R}^{200 \times 300}$.

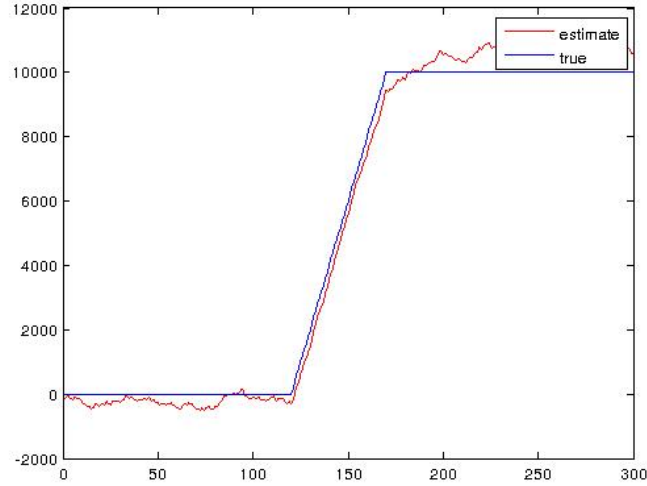


Figure 3.3:

3.2 Estimating Frequency spectra

Continuing from the previous (sub)-section we can create an estimate of g , by the following procedure:

- Estimate the coefficients of the basis \hat{a} using $\hat{a} = F^{-1}\hat{h}$
- Choose the k largest (for some k to be determined later).
- Between the indices of the k \hat{a} take the average of the signal $L^{-1}\hat{h}$.
- For each piece between the k changepoints, do a t-test to determine whether the signal is noise, or a transmission.

Some examples of the output of this procedure are shown below, for synthetic and real (Ofcom) data.

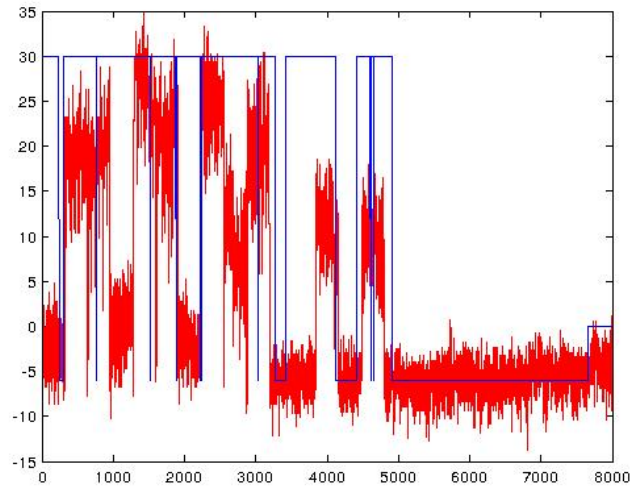


Figure 3.4:

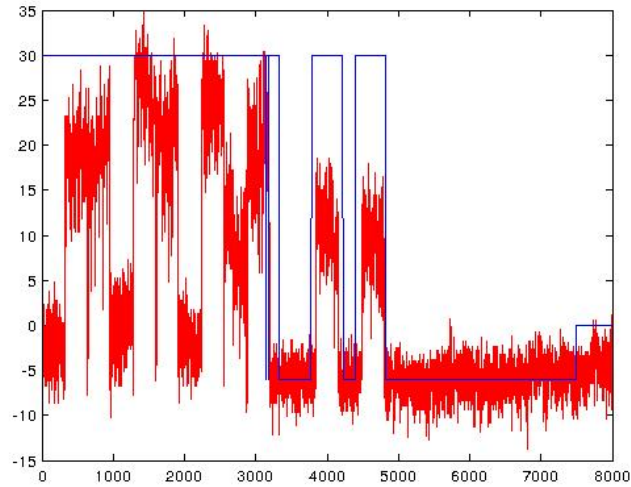


Figure 3.5:

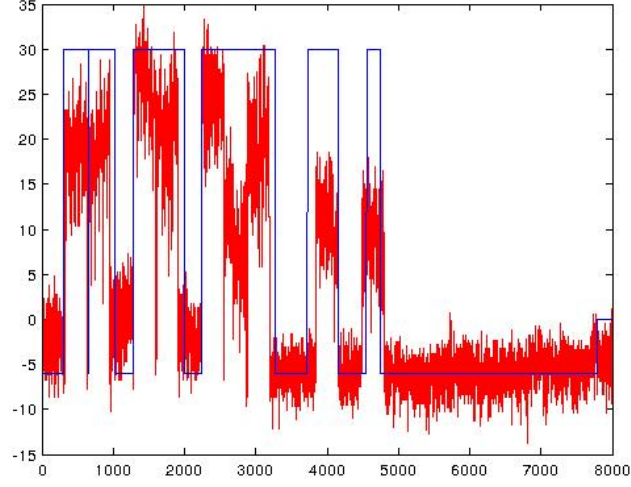


Figure 3.6:

4 Time-Frequency Model

In this section we extend the model from section 3, to jointly estimate a collection of signals sensed over time.

In this case, we have a series frequency spectra $\in \mathbb{R}^n$, sensed at T time instants. We extend the model, again in the same basis:

$$l_{i,t}(x, \tau) = \mathbb{I}(t \leq \tau) \mathbb{I}(x \leq i) \quad (4.0.64)$$

We can express these basis functions in terms of the basis functions (3.1.41), and similarly defined basis functions over time:

$$l_t(\tau) = \begin{cases} 1 & \text{if } t \leq \tau \\ 0 & \text{otherwise} \end{cases} \quad (4.0.65)$$

That is, l_t is again a left-hand step function.

Proposition 4.1.

$$l_{i,t}(x, \tau) = l_i(x) \otimes l_t(\tau) \quad (4.0.66)$$

Proof. Consider a fixed frequency, w.l.o.g $i = 1$. Then

$$l_{1,t}(x, \tau) = \mathbb{I}(x \leq 1) \mathbb{I}(t \leq \tau) \quad (4.0.67)$$

There are T possible vectors, each of length n , for this combination of (i, t) . Therefore $l_{1,t} \in \mathbb{R}^{1 \times nT}$. As there are n possible frequency vectors, one for each fixed frequency basis function, we have $n \times nT$ vectors. By symmetry of time and frequency, the same argument applies for a fixed time, and we have $T, 1 \times nT$ basis vectors in total, one for each fixed time instant. Therefore the joint time-frequency basis is in $\mathbb{R}^{nT \times nT}$. As time and frequency are orthogonal in this representation, the proposition follows. \square

Thus this basis can be represented with the following matrix:

$$L = L_T \otimes L_n \quad (4.0.68)$$

where L_k is the matrix (3.1.42) of size $k \times k$. From (4.0.68), it follows that:

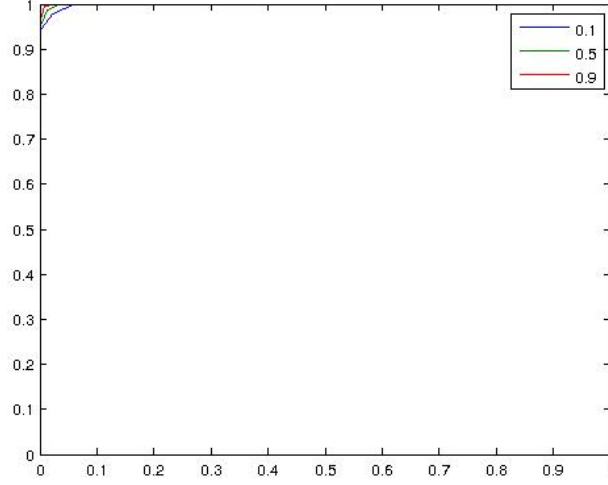


Figure 3.7:

$$D = L^{-1} = L_T^{-1} \otimes L_n^{-1} = D_T \otimes D_n \quad (4.0.69)$$

where D_k is the same as (3.1.43), and that:

$$F = LL^T \quad (4.0.70)$$

$$= (L_n \otimes L_T) (L_n \otimes L_T)^T \quad (4.0.71)$$

$$= L_T L_T^T \otimes L_n L_n^T \quad (4.0.72)$$

$$= F_T \otimes F_n \quad (4.0.73)$$

where F_k is defined as in (3.1.52).

We sense the frequency spectrum at T instants, collecting signals g_1, \dots, g_T .

Definition 4.2. We define

$$G = [g_1, \dots, g_T] \quad (4.0.74)$$

and

$$g = \text{vec}(G) = \begin{pmatrix} g_1 \\ \vdots \\ g_T \end{pmatrix} \quad (4.0.75)$$

we also define:

Definition 4.3.

$$a = \text{vec}(a_t) = \sum_{t=1}^T e_t \otimes a_t \quad (4.0.76)$$

These definitions are related by the formula:

$$g = L^T a = (L_T^T \otimes L_n^T) a \quad (4.0.77)$$

We can find an expression for a by taking the inner product of g with $l_{j,s}$:

$$\begin{aligned}
h &= \langle g, l_{j,s}(x, \tau) \rangle \\
&= \sum_{i,t} g(x, \tau) \langle l_{i,t}, l_{j,s} \rangle \\
&= Fa
\end{aligned}$$

where we have defined $F = (F_T \otimes F_n)$.

The measurements in this model can be written as

$$y_T = A_T g \quad (4.0.78)$$

with

$$A_T = I_T \otimes A \quad (4.0.79)$$

and $A \sim N(0, 1/m)$.

We consider the case where $g_t = g$ for each time slot (i.e. the signal is unchanging).

Proposition 4.4. *We can write the i^{th} column of the matrix L_k as, for $k \leq j$:*

$$L_k^T e_i = \sum_{j=1}^i e_j \quad (4.0.80)$$

where e_j is the j^{th} canonical basis vector in \mathbb{R}^k , with has a 1 in position j and is zeros everywhere else.

Proof. The i^{th} column of L_k will have i ones, and $k - i$ zeros. It can be written in terms of the canonical basis e_j , simply as sum up to the last 1 in position i . \square

In this case, we can write g as:

$$\begin{aligned}
g &= L^T a = (L_T^T \otimes L_n^T) \left(\sum_{t=1}^T e_t \otimes a_t \right) \\
&= \sum_{t=1}^T L_T^T e_j \otimes L_n^T a_t \\
&= \sum_{t=1}^T \left(\sum_{j=1}^n e_j \right) \otimes L_n^T a_t \\
&= \sum_{j=1}^T e_j \otimes \sum_{t=j}^T L_n^T a_t \\
&= \sum_{j=1}^T e_j \otimes b_j
\end{aligned}$$

where we define $b_j = \sum_{t=j}^T L_n^T a_t$. We can now write

$$\begin{aligned}
y_T &= A_T g = (I_T \otimes A) \left(\sum_{j=1}^T e_j \otimes b_j \right) \\
&= \sum_{j=1}^T e_j \otimes A b_j
\end{aligned}$$

Example 4.5. Now consider the case where all the a_t are identical. In this case:

$$\begin{cases} a_T = a & \text{if } t = T \\ a_t = 0 & \text{if } t < T \end{cases} \quad (4.0.81)$$

so

$$b_j = \sum_{t=j}^T L_n^T a_t = L_n^T a = b \quad (4.0.82)$$

Then

$$y_T = \left(\sum_{j=1}^T \right) \otimes Ab = 1_T \otimes Ab \quad (4.0.83)$$

Let the s^{th} column of L_T^T be denoted by B_s , then

$$(L_T^T)^{-1} B_s = D_T^T B_s = e_s$$

As before, we form the estimator

$$\hat{h}_T = \sum_{i,t} \langle y_T, A_T l_{i,t} \rangle l_{i,t} \sim \frac{1}{mT} F a$$

and run the algorithm defined in the last section on the results.
Consider

$$\begin{aligned} \langle y_T, A_T l_{i,t} \rangle &= \langle A_T^T y_T, l_{i,t} \rangle \\ &= \langle (I_T \otimes A^T) \left(\sum_{j=1}^T e_j \otimes A b_j \right), (l_t \otimes l_i) \rangle \\ &= \langle \sum_{j=1}^T e_j \otimes A^T A b_j, (l_t \otimes l_i) \rangle \\ &= \sum_{j=1}^T e_j^T l_t \otimes b_j^T A^T A l_i \\ &\sim \frac{1}{m} \sum_{j=1}^T e_j^T l_t \otimes b_j^T l_i \\ &\stackrel{?}{=} \frac{1}{m} \sum_{j=1}^T j \otimes F_n a \end{aligned}$$

where we used the expectation of a Wishart matrix between the 5th and 6th lines.

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.

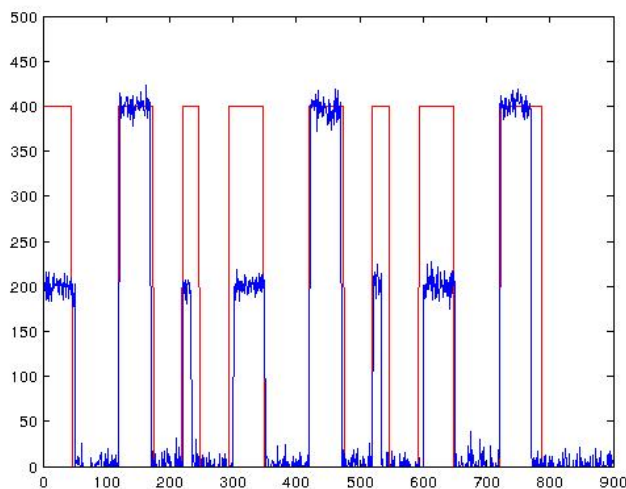


Figure 4.8:

- [2] Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):21–30, 2008.
- [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 ell-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.
- [9] Morris L Eaton et al. The wishart distribution. In *Multivariate Statistics*, pages 302–333. Institute of Mathematical Statistics, 2007.
- [10] LÉVÊQUE Olivier. Random matrices and communication systems, 2011-2012 - Spring Semester.
- [11] M. Unser. Sampling—50 Years after Shannon. *Proceedings of the IEEE*, 88(4):569–587, April 2000.

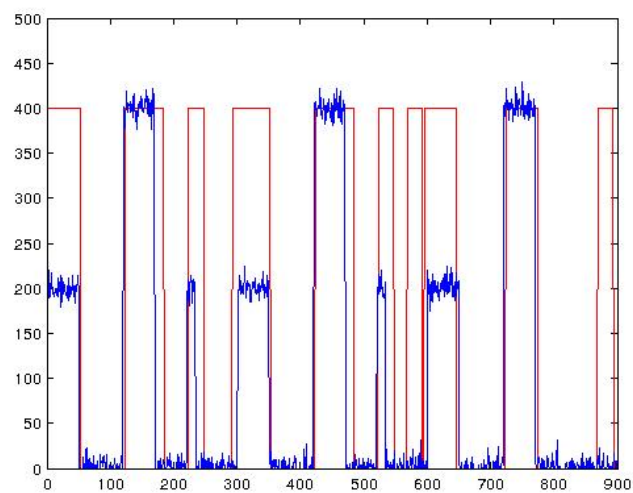


Figure 4.9:

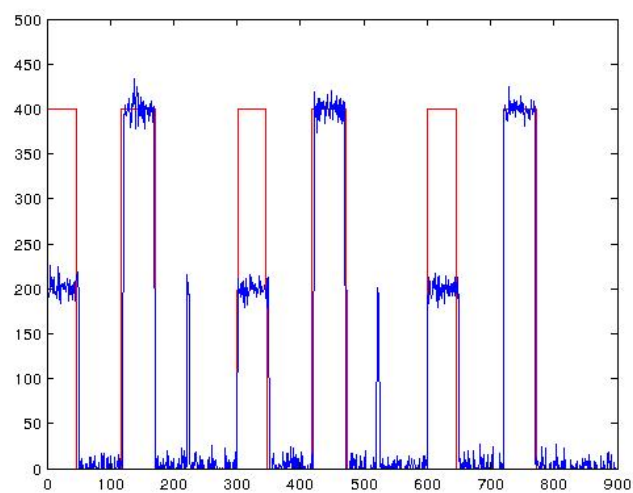


Figure 4.10: