Notes on Signal Processing

Hossein Noorazar

March 6, 2021

A Map of This Paper: a View from 30,000 feet

How do we extract usable information from data, especially data that is full of noise and or information that has nothing to do with the information we are interested in extracting?

That this is an old challenge is underlined by the fact that the first appearance of the phrase "Looking for a needle in a haystack" is nearly 500 years old. But the fact that life introduces confounding data – entropy, noise, etc – has made this idiom a part of the working wisdom of anyone trying to make careful inferences, implying the idea is much, much older.

First known appearance of the idiom "Looking for a needle in a haystack", in written English, is from St. Thomas More, in 1532. The actual quote is "to go looking for a needle in a meadow".

In this paper, we introduce a relatively new approach to the extraction of information from signals of all sorts – *overcomplete dictionary methods* – as well as the background or context that helps in the effort to begin making these methods instinctively available.

Signals and Vectors: We use the term *signal* very generally to cover any measured signal – a time series of real values, an image of $N \times N$ pixel values, a movie of $N \times N$ images, etc. We will also refer to each of these as *vectors* because they are equally well described as points in vector spaces.

After a **gentle introduction to finding needles in haystacks**, illustrating how the closely related components of sparsity, low-dimensional models and correctly chosen representations solves the inverse problem of signal recovery from noisy measurements (i.e. finding needles in haystacks), we move on to explain that the well known and ubiquitous **Fourier methods are just one of an infinite family of orthogonal transformation methods**.

The **heart of this paper** – an introduction to sparse, non-orthogonal representations via the **overcomplete dictionary methods** assisted by the matching pursuit algorithm – is illustrated on the problem of stock price prediction.

Prediction vs. thinking in bets: Actually, we use the models to predict odds, which is more in line with the idea of thinking in bets than it is with prediction of the future.

We then fill out the context a little more with an **introduction to inverse problems and sparsity**, and some comments on, and illustrations of, the power of sparsity assumptions and other related priors. The **importance of metrics and priors** is explained, as is their intimate relation to each other.

We close with **an invitation to explore** through links to code that we provide for those that want to explore the ideas on their own.

A (somewhat) Technical Introduction

Any evolving quantity – prices in some market, temperatures in Moab, Utah, population of salmon in the Columbia river, or expected commute times in Seattle WA – can be described as a continuous or discrete time series:

$$s(t) \text{ with } t \in [0, T] \text{ or } t \in \{t_1, t_2, t_3, ...\},$$

where we have used s – for *signal* – to denote the quantity of interest. It will also almost always be the case that we do not have access to $s(t)$, but instead have a version of $s(t)$, $m(t)$ – the *measured* signal – which is $s(t)$ corrupted by noise or some distorting process $\eta(t)$:

$$m(t) = s(t) + \eta(t).$$

While it is absolutely true that to actually fully understand any given signal $s(t)$ we will usually need more information than the measured signal $m(t)$ contains, having the ability to extract information from $m(t)$ and $s(t)$ is almost always a very high priority.

The geometric ideas behind $m^{\eta}_{\tau}(t) \equiv s_{\tau}(t) + \eta(t)$ is illustrated in Fig. 1

This paper focuses on the extraction of cycles and cycle-like components from signals through the use of tools from the relatively new area of sparse data analysis. In particular, we will look at the power of overcomplete dictionaries for the generation of *sparse signal representations* and use that to improve the prediction of probabilities for $s(t)$ for future times t.

We begin with two examples illustrating the power that having the right signal representations gives analysts in their quest to extract information from signals. In the first example, we use the *low dimensional signal manifolds* – a very close cousin to sparse signal subspaces – to denoise a signal. In the second example, we show a classic example of signal denoising illustrating the power of *Fourier methods*.

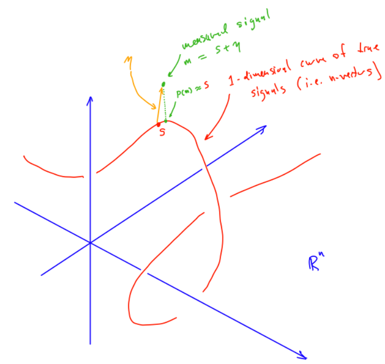

Figure 1: This shows the curve of true signals that is corrupted by noise and then recovered due to the sparsity of the true signals in the space of possible signals.

Technical details: projections work because random signals (like noise) will typically be orthogonal to any fixed direction. See Figure Fig. 4.

Example 1: Finding Low Dimensional Signals

We consider temporal signals $m^{\eta}_{\tau}(t)$ which are very noisy measurements of simple step signals $s_{\tau}(t)$ which step from -1 to 1 at some time $0 < t = \tau < T$:

$$m^{\eta}_{\tau}(t) \equiv s_{\tau}(t) + \eta(t)$$

where the noise $\eta(t)$ is a Gaussian random variable with mean at time t $\mu(t) = 0$, variance $\sigma^2(t) = C \gg 1$, and the property that $\eta(t)$ and $\eta(s)$ are independent if $t \neq s$. See Fig. 2 showing an example $m^{\eta}_{\tau}(t)$ in which the noise makes the original signal impossible to see:

It turns out that even if the true signal is buried in so much noise, we cannot by eye pick out the true transition time τ in the pure signal.

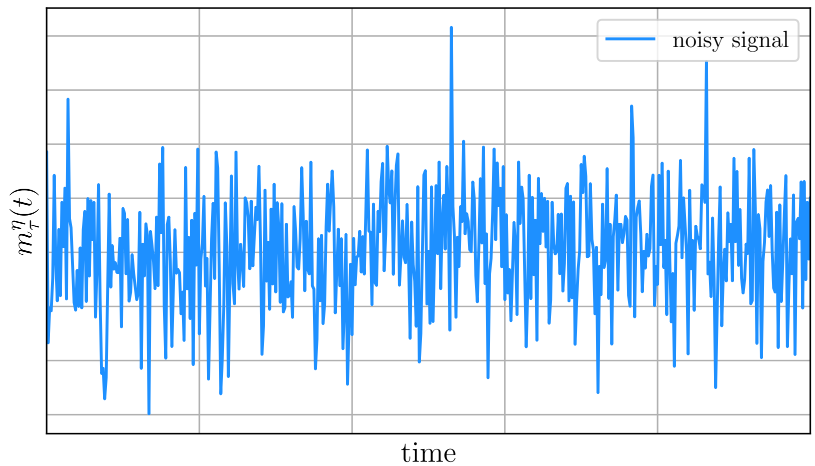

Figure 2: The level of noise makes the visual determination of which true signal we are measuring.

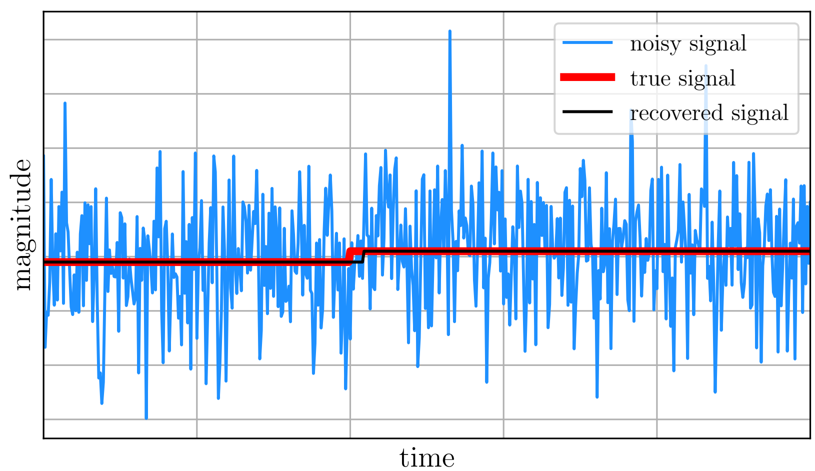

Figure 3: Because of the low dimensionality – the sparsity of the set of true signals in the space of possible signals – a projection to the set of true signals finds the true signal with a small error.

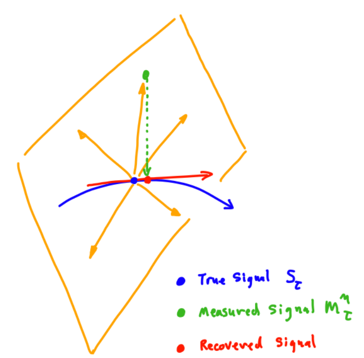

Figure 4: Technical Details behind The Signal Recovery: Concentration of measure (and sparsity) implies that the noise moves the true signal mostly orthogonally to the curve making the projection back to the curve a good strategy for signal recovery.

The fact that we have a **low-dimensional** family of true signals, parameterized here by a single (one-dimensional) value τ, allows us to extract this information from the noisy measurements. See Fig. 3 showing the recovered signal that is close to the true signal, in spite of the huge amount of noise.

In contrast, if we did not know the shape of the true signal in the general case in which s(t) can have any shape and $\eta(t)$ is still the Gaussian noise, the extraction of the true signal from the noisy measurement

$$m^\eta(t) \equiv s(t) + \eta(t),$$

is harder.

Knowing the right representation in a low dimensional family of signals (the family is parameterized by a single value, τ) allows us to determine the true signal through a simple projection.

Example 2: Fourier Filtering

An even simpler example is that of a true signal possessing only low-frequency Fourier modes, corrupted with high frequency noise.

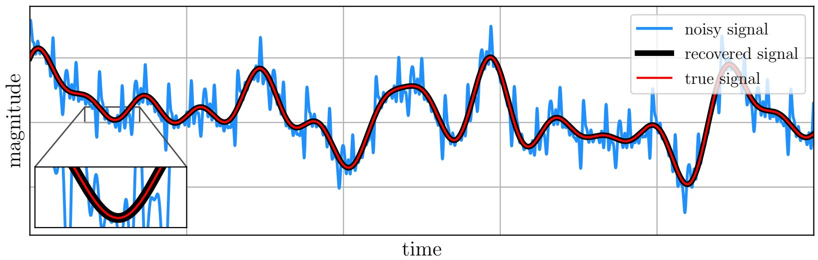

Figure 5: Recovery of the true, denoised signal becomes trivial in the transformed coordinates. Using the right representation for the task makes the problem easy to solve! The inset plot shows the true signal and the recovered one coincide.

Technical Details: The recovery of a signal in which the noise frequencies are higher than the signal frequencies is easy.The idea is simple. Transforming to the frequency domain, we can simply through away components that are higher than some cutoff and then compute the inverse Fourier transform. If we have picked the cutoff correctly, the result is the true signal. See Fig. 6 and Fig. 7

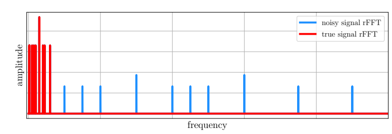

Figure 6: **Fourier Transform of Noisy Signal**: The red shows the part from the true signal and blue, the part from the corrupting noise.

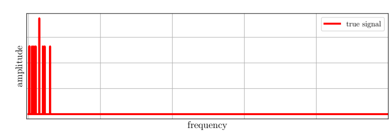

Figure 7: **Signal Recovery is Easy**: Denoting the cutoff function by H_c, the measured noisy signal by m^{η}, and the Fourier transform by F we get that the recovered signal is $F^{-1} \circ H_c \circ F(m^{\eta})$.

In spite of the fact that it is tricky to denoise the measured, noisy signal in the time domain, the problem becomes extremely easy when transformed to the frequency domain. See Fig. 5 for an example of a signal that can be recovered exactly because the noise and signal separate in the transformed representation, making the recovery computationally fast and extremely accurate.

While real signals are often more complex than this, without a nice clear separation in frequencies between signal and noise, this idea is at the core of a lot of noise reduction methods: Finding a representation in which the noise and signal are clearly separated is a central thread in methods aimed at signal understanding.

Key Task: Choosing the Right Representation!

While the operational principle/perspective of *The Efficient extraction of information from signals depends on having the right representation* is not a surprise to anyone who has even merely dabbled with Fourier analysis, we believe that the potential of this perspective, when it is instinctively part of the analysts toolbox, is far from fully realized. In this paper, we introduce the power of *sparse representations*, with a special focus on *sparse representations through the construction and use of overcomplete dictionaries and matching pursuit*.