Denoising physiological signals using greedy methods

Summary

Alain Soltani, Loïc Tudela,
Adil Salim, Laurent Lambert

1 Introduction

One of the growing techniques in functional brain imaging is the study of recordings provided by magneto, electro-encephalographic methods (often called EEG, MEG). These techniques, of the latest fashion in functional neuroimaging, record an electromagnetic signal over the scalp of a patient, using non-invasive devices. Unlike many other invasive procedures, which directly disturb the normal functioning of the brain, we get here a signal which directly emanates from neural activity. Furthermore, these methods offer great temporal resolution : it is possible to perform measurements on very small time intervals, and one can follow neural activity step by step.

But this does not come without drawbacks : measurements cannot be acquired without any artifacts. As a matter of fact, this physiological data, of a very low order of magnitude when compared to the quantities produced by Earth, is extremely sensitive and can be corrupted by many sources of additive noise : background brain activity, electrical heart activity, eye-blinks, etc. It is therefore necessary to denoise it - a difficult but essential task. In many biophysical and medical areas, clear, easy-to-use signals are needed for further physical and behavioral analysis.
In this project, we aimed at forming the best possible reconstruction of these signals, without noise ; the final object should be a concise approximation of the true signal, containing the most information possible.

2 Framework

We used MEG recordings from experiments conducted at the Martinos Center of Massachusetts General Hospital, using a set of 203 sensors that would capture a signal sampled at the millisecond scale. In the experiment, auditory or visual stimuli were presented to the patient in a random sequence.

We placed ourselves in a linear framework, using a single set of relevant explanatory variables : each reconstruction $\tilde{y}$ of the signal y should be a linear combination of these variables, called *atoms* - all coming from the same explanatory *dictionary*.

The difference between the acquired and the reconstructed signals $y - \tilde{y}$ is the noise we aim at extracting ; the error or *residual* term. For those preferring a matrix notation :

$$y = \tilde{y} + \epsilon = D\lambda + \epsilon,$$

where D is our dictionary. λ contains the weights we attribute to the atoms selected to explain the signal.

We assume that this extracted noise would satisfy suitable statistical requirements : it should be a normal random variable (i.e. the noise occurred randomly during the experiment), independent from the true brain signal.

We focused first on a single-channel model : we developed all our denoising algorithms for acquisitions made with a unique sensor. It becomes easier to remove the idiosyncratic component, while still being easily generalizable to a a multi-channel model (taking multiple sensor acquisitions into account, detailed afterwards).

3 Matching Pursuit

Our idea was to denoise a signal using as few as possible atoms from our dictionary. The constitution of the weight vector λ was performed iteratively, using greedy optimization methods ; we centered our study around the classic signal processing algorithm called *Matching Pursuit* (MP), first designed by Mallat and Zang.
Given a fixed dictionary D, the Matching Pursuit will first find the atom that has the biggest correlation with the signal, weight it into λ, and then subtract the contribution due to that atom from the error term, and repeat the process until the signal is satisfactorily decomposed. To stop including explanatory variables in our model, we test the significance of each new atom *vis-à-vis* the temporary model. Figure 1 shows the effects of our denoising on the 100^{th} acquisition :

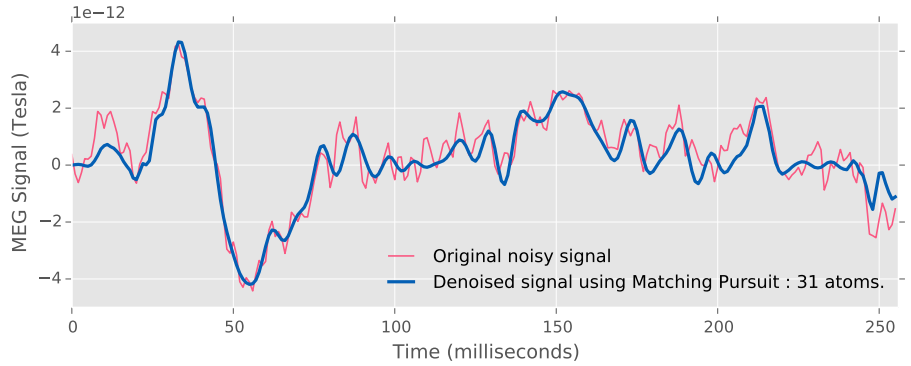

FIGURE 1 – *Original & denoised signals, via Matching Pursuit.*

Using this algorithm, the variables selected have a decreasing significance, and we stop the inclusion procedure when there is no more significant variable to add (here, we only have 31 atoms in our model, for instance).

Note that the error extracted from the noisy signal, using this algorithm, meets all the statistical requirements made earlier. In addition, this algorithm provides very stable results : for example, one can artificially add noise to our reconstructed signals, and restart the algorithm to see if we obtain the same denoised reconstruction ; indeed, the new output is extremely close to the previous output.

4 Refinements

Satisfied with the results provided by our algorithm, we have made some improvements : for example, we added a step in our iterative process that consists in projecting the weights in λ on the atoms we select.
This is called *Orthogonal Matching Pursuit* (OMP), and improves many aspects of the MP - such as the error term, that becomes smaller.

Besides, in certain specific cases, these algorithms become equivalent to a simple, one-step procedure : the user deletes, in one fell swoop, all non-significant variables in the model - called *hard thresholding*.

5 Multi-channel Matching Pursuit

Now that we have the Matching Pursuit procedure, we can ideally get a total denoised acquisition, by denoising individually each signal measured by a single sensor. But in this case, each single denoising doesn't necessarily refer to the same atoms, to explain the signals ; furthermore, it can be very costly and tedious to perform all these denoisings, one after another.

Taking into account the multi-channel nature of MEG - recall that two hundred sensors were used to record the brain signal -, we needed methods that can denoise at the same time, using a single dictionary of atoms, all the noisy signals we possess.

Hence we developed a multi-channel version of the Matching Pursuit ; in that framework, the weights given to the atoms may vary between sensors, but the atoms chosen are the same for all sensors. This multi-channel denoising is stable and compelling : for example, the residuals obtained here present the same qualities as in the single-channel section.

As a graphic overview, Figure 2 presents a superposition of several brain signals, on a short range of sensors, before and after denoising them. The signal becomes smoother and doesn't take into account the environmental noise.

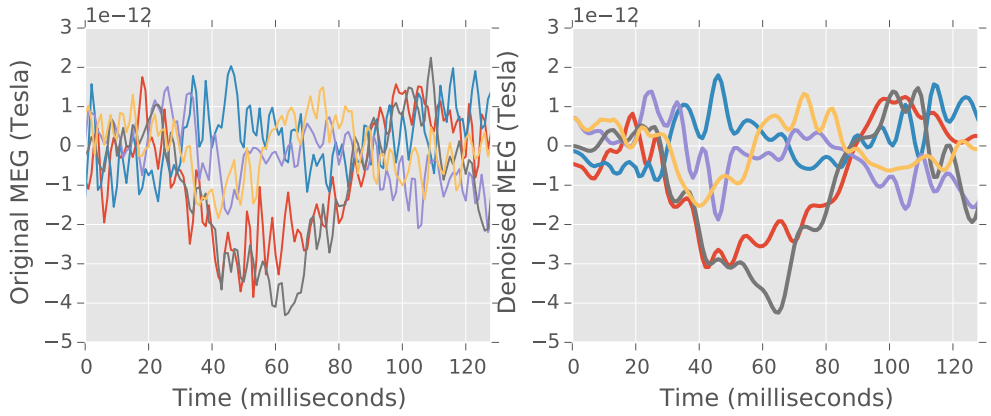

FIGURE 2 – FROM LEFT TO RIGHT : *Original, denoised acquisitions ; sensors 50-55.*

6 Dictionary selection

At this stage, the experimenter can perform a mapping of the brain by representing respective areas of high and low noise. This would allow to distinguish brain zones stimulated by the experiment from the intact ones ; it becomes possible to draw conclusions about the physical nature of the experiments, and their influence on brain performance.
However, these conclusions are dependent on the choice of explanatory dictionary ; thus we still have to select the optimal dictionary for our study.

In order to select a well-suited dictionary for our study, we focused on two main criteria :
· firstly, the Akaike information criterion (AIC). This statistical criterion allows us to discriminate between models : the model with the lower AIC value will provide the best trade-off between a good fitting of the data, and a small model complexity (i.e. less weights given to the atoms) ; with too few atoms, the data won't be explained well and a lot of information will be lost ; with too many, the model won't be concise and efficient enough.

This criterion has the advantage, compared to a conventional fitting coefficient like R^2, of truly penalizing models with too many atoms.

· secondly, a cross-validation : this is a technique for estimating the performance of a predictive model. We split the experiment signal in two parts, and obtain the atoms that explain the first part. Then we use the same atoms on the second part, to see how well they fit the data. We repeat again the whole procedure by switching roles between the two parts.
A good dictionary should explain the bulk of the signal using the same atoms.

To give graphical insights of the dictionary selection procedure, Figure 3 presents two graphs, for the criteria detailed above. Simply note that we seek, for each criterion, the dictionary giving the lowest values.

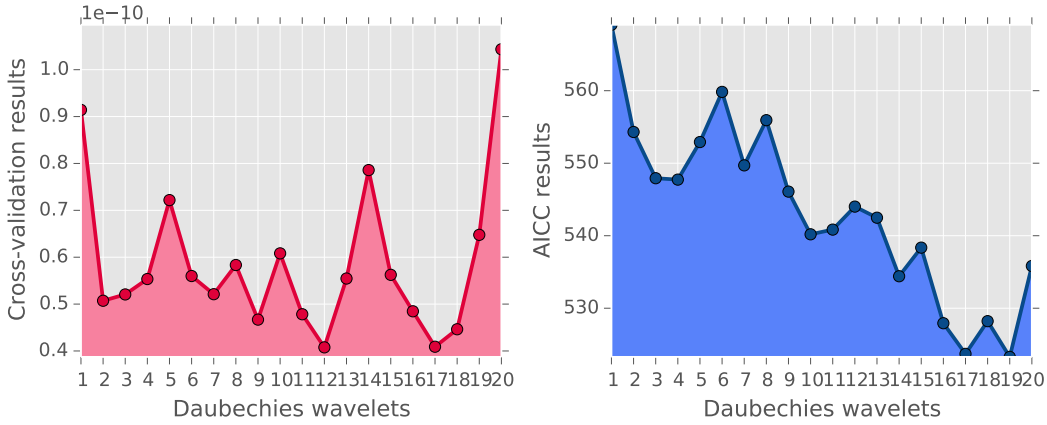

FIGURE 3 – FROM LEFT TO RIGHT : *Criteria results for selecting the best dictionary. Research has been made among the Daubechies family, a certain family containing 20 different dictionaries.*

All these criteria allow us to choose an optimal dictionary from a list of available dictionaries. For example, from the two previous figures, we can see that the dictionary called `Daubechies 17` is the optimal dictionary available from the list.

In this study, we used classical dictionaries called *wavelet* dictionaries ; there is a golden rule when choosing a dictionary from this family. With this detailed study, we were able to prove that the traditional rule comes from a statistical reality, but we must remain cautious with its use. The use of true criteria is strongly preferable.
We conclude this part by presenting an algorithm for optimizing the dictionary chosen, called K-SVD, whose reconstructive power is very strong.

7 Conclusion

We finally obtained quality signals that can be analyzed quickly by doctors and physicists, to draw conclusions on the influence of stimuli on brain behavior. It becomes possible to distinguish brain areas stimulated during the experiment and develop brain mappings to medically quantify the patient behavior during the study ; a purpose beyond the scope of this study, but raising scientific issues of great interest.