

Machine Learning Spring 2016, Homework 5 – final project

We continue to work on data provided through a student project which has been designed by Lee Cheng (<http://www.masys.url.tw/AU/AU.htm>) and is nicely documented at http://www.masys.url.tw/AU/2015SP/BMSD-D/HW/HWfinal-Maternal_Fetal_ECG/ProjectDescription.htm. This project concerns isolating an unborn child's ECG signal from ECG recordings taken from the mother's thorax and abdomen. As you can see from the references list at the end of that webpage, this is a heavily investigated signal processing task. We will apply the (comparatively modest) powers of adaptive linear signal processing to this task.

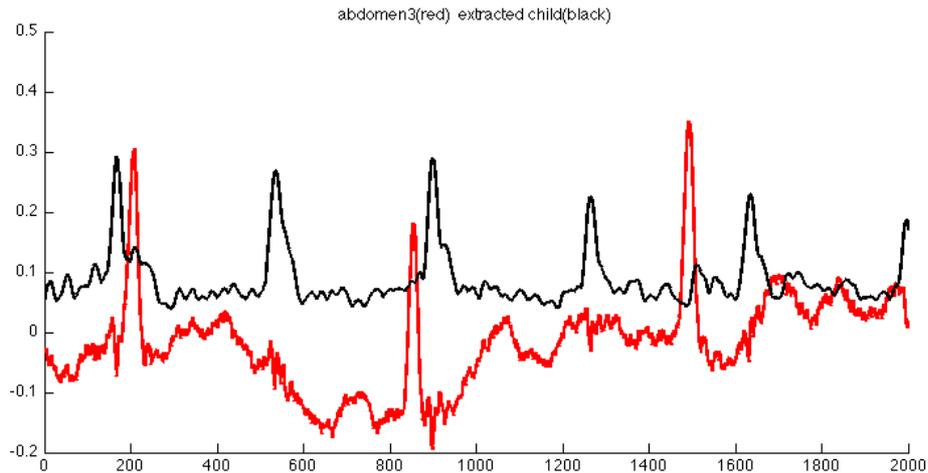
After the preparatory HW #4, we now address the much more relevant task of isolating the child's ECG from the mother's abdominal ECG signal(s), which constitute a sum of (distorted, noisy versions of) the mother's own heart ECG and the child's ECG signals. We use the five signals `abdomen1`, `abdomen2`, `abdomen3`, `thorax1`, `thorax2` contained in the Matlab data file `fecg.mat` that can be downloaded from that webpage. When you visually inspect plots of those five signals (such a visual data inspection is always the best thing to do at the beginning of any data modeling project), you will find that the abdominal signals look much less well-behaved than the sample abdominal signal shown on that webpage (Section 4 on that webpage). If you look with very sharp eyes you will find that only the signal `abdomen3` contains a visible, feeble trace of some pulse that is about twice as frequent as the dominating mother's ECG pulses. So I suggest that you try to extract the child's signal from `abdomen3`, on the basis of mother references `thorax1`, `thorax2`, and ignore `abdomen1` and `abdomen3`.

The basic architecture to use is the one shown in Figure 2 of the paper

Fanaswala, M (2008), *Fetal Electrocardiogram Extraction: A Case-Study in Non-linear System Identification*.

This is the same architecture as in Figure 6.10 in our lecture notes, -- it is the standard classical system for "denoising" a signal. (*Note: there exist other, more recently developed and more powerful architectures for the same task. If you are interested, check out https://en.wikipedia.org/wiki/Independent_component_analysis. These methods are however significantly more involved.*)

You will find that a straightforward implementation of this denoising filter does not yield useful results. A crippling issue is the random, slow drift in `abdomen3`. You can get rid of it by using a high-pass filtered version of `abdomen3` instead of the raw signal (high-pass filtering removes slow signal components – here, the drifts). You get a simple high-pass filtering effect by subtracting from `abdomen3` a moving local average. In terms of computation a moving average is a convolution operation, for which the Matlab function `conv` will be helpful. Some more insights and "tricks" will be needed until you achieve a good-looking result. I tried it on my own, it is possible to get something like in the figure below (the two curves are individually rescaled for nice visual appearance).



Note that the black line in this graphics doesn't look like a typical ECG signal. This cannot be expected since the electric signal pathways will introduce arbitrary transformations on the original child ECG signal; furthermore, the "true" ECG is not even well-defined as one would get different signals when (hypothetically) placing electrodes directly at the heart at different locations.

Task: extract from `abdomen3` a signal that you can interpret as a transform of the child's ECG. It should be roughly periodic with a period length about half as long as the mother's heartbeat.

Deliverables: A report and your code. The code should contain a runnable Matlab or Python script `myFinalScript` which, when executed, produces the results and graphics of the solution that you document in the report. As usual, the report should be self-contained, that is, formally and fully specify all the pre- and postprocessing tricks and routines that you found, and also contain a discussion of difficulties that you encountered and suggestions for further work.

Bonus points may be awarded for extra efforts, like for example

- tackling other datasets (a database of mother/child ECGs is linked at the bottom of that project webpage)
- analyses of the less easy-to-use signals `abdomen2` and `abdomen1`,
- very clever and insightful pre/postprocessing,
- very beautiful-professional report writing,
- applying other filtering methods besides the classical denoising filter.

As usual, send your code and document to x.he@jacobs-university and myself (h.jaeger@jacobs-university.de). **Deadline** is Thursday May 12, 23:59 hrs.