Comparison of Time and Wavelet Domain Approaches In Nonparametric Detrending of fMRI Time-Series

Babak Afshinpour[1], Gholam Ali Hossein-Zadeh[2], Hamid Soltanian-Zadeh[3]
Control and Intelligence Processing Centre of Excellence, School of Elec. & Comp. Eng., University of Tehran, Tehran, Iran
b.afshinpour@engmail.ut.ac.ir

Abstract

fMRI time-series are contaminated with unknown low frequency fluctuations which are called ‘trend’ or ‘confounds’. Conventional methods of trend removal try to consider a model with a specific degree of smoothness for trend. In this paper, we estimate trend components using partially linear models (PLM). PLMs allow one to combine detrending and analysis of time-series in one scheme. In addition, we developed estimation procedures in time and wavelet domains for a nonparametric trend. We applied proposed methods on simulated and experimental data, and compared their performance with simple (linear) detrending through measuring the detection sensitivity, false alarm rate control, and variance of estimation.

Keywords: fMRI - Trend removal - drift - PLM(partially linear models)

Introduction

Functional magnetic resonance imaging (fMRI) is an important noninvasive technique for investigating the brain neural activity. Based on the coupling between local neural activity and regional changes in cerebral blood flow (CBF) and blood oxygenation level; it finds the activated regions of brain. The increment in CBF and blood oxygenation level; which is accompanied with neural activity, is evidenced in the form of increase in T_2^* decay rate of active region. Therefore rapid acquisition of T_2^*-weighted can reveal the variation of the blood oxygenation level.

One major limiting factor for fMRI is the existence of disturbances such as correlated head or eye movements, scanner calibration drifts, and effects of physiological processes such as vascular flow and respiratory.

Every time-series of fMRI is usually modeled as a linear convolution of stimulus pattern and with voxel-wise hemodynamic impulse response plus an error (Figure 1). This can be written in the following equation:

$$y_i(t) = p(t) * h_i(t) + \varepsilon(t) \qquad (1)$$

where $p(t)$ is temporal pattern of stimulus, $h_i(t)$ is the model of hemodynamic impulse response at voxel i, $\varepsilon(t)$ models

[1] PhD. student, School of Electrical and Computer Engineering, University of Tehran.
[2] Assistant Professor, School of Electrical and Computer Engineering, University of Tehran,
[3] Professor, School of Electrical and Computer Engineering, University of Tehran

undesired effects and $y_i(t)$ is measured time-series of the ith voxel. The linear model for describing BOLD effects in fMRI time-series is a very practical and common approach for analysis of fMRI time-series and detection of voxels. By modeling the hemodynamic response as a linear combination of definite basis functions [1], it is possible to rewrite the equation (1) in matrix-vector format as follow:

$$\mathbf{y} = \mathbf{X}^{\mathbf{T}}\beta + \boldsymbol{\varepsilon} \qquad (2)$$

Where, **y** is the fMRI time series observed at a single voxel, **X** is the design matrix containing basis functions in columns. β is the parameter to be estimated, and $\boldsymbol{\varepsilon}$ is the vector of residuals. The term $\boldsymbol{\varepsilon}$ in (2) can be written as a non-parametric component. So, equation (2) can be rewritten as follow:

$$\mathbf{y} = \mathbf{X}^{\mathbf{T}}\beta + \mathbf{f} + \boldsymbol{\varepsilon} \qquad (3)$$

This model is more flexible than (2). Because of its relation to the linear model (1), this model is called “partially linear model” (PLM) in literature. In this paper, we present two estimation approaches for $\mathbf{f}$ (or trends) in model (3). The first is based on time-domain analysis [2] and the second is based on wavelet domain approach [3]. Then we compare the efficiency of these methods.

Partially linear model Estimation

Since the trend component in fMRI time series is a smooth function of time, we adopt a sort of penalty function in the estimation which guarantees a smooth answer for non-parametric part ($\mathbf{f}$). Therefore the cost function is the sum of log-likelihood function plus a penalty term that tunes the smoothness of non-parametric parts:

$$P(\boldsymbol{\beta},\mathbf{f}) = \left\|\mathbf{Wy} - \mathbf{WX}^{\mathbf{T}}\beta - \mathbf{Wf}\right\|_2^2 + \lambda p(\mathbf{Wf}) \qquad (4)$$

Where $p(.)$ is a specific function that adjusts the smoothness of $\mathbf{f}$ and λ is a smoothing parameter. **W** is a linear and orthonormal transform. The most important characteristic of this transform is representing the signal with a few numbers of coefficients. The cost function has two terms. The first term is a norm of error which preserves closeness of fit; and the second term is a measure of entropy by which smoothing of trend could be controlled. This penalty function can be estimated in either time or wavelet domains.

Time Domain Estimation of Trends

In the time domain estimation, we assume that **W**=**I**, (**I** is n×n identity matrix). The most well-known solution for time-domain is called natural cubic spline in which the adopted penalty function in time domain is as follows:

$$p(\mathbf{f}) = \mathbf{f}^T\mathbf{K}\mathbf{f} \qquad (5)$$

where $\mathbf{K}$ is the nonnegative definite smoothing matrix. Green and Silverman [4] proposed a method to define a proper smoothing matrix. In this paper, we use a simplified version of their work for equally-spaced time points. The smoothing matrix is considered as $\mathbf{K} = \mathbf{Q}\mathbf{R}^{-1}\mathbf{Q}^T$. The size considered for **Q** is n×(n-2) and the entries are Q_{ij} for i = 1,…,n and j = 2,…,n-1, where

$$Q_{j-1,j} = Q_{j+1,j} = T_R^{-1} \quad and \qquad (6)$$
$$Q_{j,j} = -2T_R^{-1}$$

With $Q_{ij}=0$ for $|i-j| \geq 2$. T_R is the sampling time of image acquisition. Note that the columns are indexed with an unusual convention starting with j = 2. The (n -2)×(n-2) matrix R is given by

$$R_{ii} = 2T_R/3 \quad for \quad i=2,\dots,n-1 \qquad (7)$$
$$R_{i,i+1} = R_{i+1,i} = T_R/6 \quad for \quad i=2,\dots,n-2$$
$$R_{i,j} = 0 \quad for \quad |i-j| \geq 2$$

It is now easy to show that the minimum of (4) with the penalty of (5) is;

$$\beta = (\mathbf{X}^T \mathbf{W}_\mathbf{x} \mathbf{X})^{-1} \mathbf{X}^\mathbf{T} \mathbf{W}_\mathbf{x} \mathbf{y} \quad (8)$$

$$\mathbf{f} = (\mathbf{W}_f + \lambda \mathbf{K})^{-1} \mathbf{W}_f \mathbf{y},$$

where

$$\mathbf{W}_x = \mathbf{I} - (\mathbf{I} + \lambda \mathbf{K})^{-1} \quad (9)$$

$$\mathbf{W}_f = \mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$$

The implementation of spline estimation is computationally extensive because it contains inverse calculation of some matrices.

Wavelet Domain Estimation of Trends

In wavelet domain estimation, we assume that **W** is n×n wavelet matrix operator.

If $p(\mathbf{f})$ is defined as L_1 norm ($p(\mathbf{f}) = \|\mathbf{Wf}\|_1$), minimization of $P(\beta, f)$ is turned out to soft-thresholding [3]. So

$$\mathbf{f} = sign(\mathbf{W}y - \mathbf{WX}^\mathbf{T}\beta) \circ (|\mathbf{Wy} - \mathbf{WX}^\mathbf{T}\beta| - \lambda)_+ \quad (10)$$

The L_1 entropy is continuous but this comes at the price of shifting the resulting estimator by a constant.

Smoothing Parameter

The optimal smoothing parameter λ is determined with GCV [5]. For a given λ, the GCV score is ;

$$GCV(\lambda) = \frac{1}{n} \frac{\|\mathbf{Wy} - \mathbf{WX}^\mathbf{T}\beta - \mathbf{Wf}\|_2^2}{(1 - trace(\mathbf{H})/n)^2} \quad (11)$$

Where, H is called ‘hat matrix’ which is defined as $\mathbf{HWy} = \mathbf{WX}^\mathbf{T}\beta + \mathbf{Wf}$. The optimum λ minimizes GCV criterion. GCV criterion has a logic interpretation. The nominator of (11) expresses the fitting quality, and the denominator control the over fitting. If $\mathbf{f}$ estimated so that over fitted to the observation, the nominator of (11) will decrease, but simultaneously the denominator will decrease and GCV may increase.

Activation Detection after Estimation

In fMRI analysis, one tries to specify whether a time-series is active or non-active. The null hypothesis (H_0) means that the time-series belongs to a non-active voxel and alternative hypothesis (H_1)—means that the time-series includes activation. The parameter β qualifies the activation contents of a time-series and therefore can be used for detection. The estimated β can not be directly used for hypothesis testing. In fMRI literature usually a t-test is used. If it is assumed that estimated β has a normal probability density function under null hypothesis, then $\beta / \sqrt{\mathrm{var}(\beta)}$ has a t distribution under H_0. Variance of β is calculated as follow [6]:

$$\mathrm{var}(\beta) = \frac{\|\mathbf{Wy} - \mathbf{WX}^\mathbf{T}\beta - \mathbf{Wf}\|_2^2}{trace((\mathbf{I} - \mathbf{H})^\mathbf{T}(\mathbf{I} - \mathbf{H}))} \quad (12)$$

Computational Steps

In order to implement the method, it is necessary to run the following algorithm for each time-series:

1. Transform the time-series to a new domain, for example wavelet domain (optional).
2. Initialization, choose proper values for β.
3. assume a value for λ
4. For k=1,2,… until convergence (for the case **W=I,** there is a close form response and no iteration is needed)
    a. Compute β, $\mathbf{f}$ via (8,9) or (10)

5. Go to 3, until the GCV minimized. (The minimization here is done by optimization toolbox of MATLAB).

6. Calculate t-value using t-test approach, form detecting active voxels.

Experiments

Simulated Data

In order to produce simulation data, the Gaussian noise is simulated using random generator of MATLAB software packages. A standard deviation of 10 for the density function was chosen. A Gamma HRF is used for hemodynamic system impulse response [7]:

$$h(t;\tau,\sigma)=\begin{cases} e^{-t/\sqrt{\sigma\tau}}\left(\frac{et}{\tau}\right)^{\sqrt{\tau/\sigma}} & \text{t} > 0 \\ 0 & \text{t} < 0 \end{cases} \quad (13)$$

Here, we consider, $\sigma = 0.063$ and $\tau = 4.73$.

A linear combination of a sin function ($f_1(n) = 5\sin(2\pi n T_R / 256)$), a line ($f_2(t) = 0.02 n T_R$) and a second-degree polynomial ($f_3(n) = 0.00008(nT_R)^2$) are used for simulation of trends. We assume T_R=2s, and $0 \le n \le 255$, which means each time series has a length of 256 samples. Different activation intensities (β_{sim} = 0.5, 1, 1.5, and 2) in different sizes of clusters (64, 16, 4, and 1) were added to dataset. Therefore, the simulation time-series is produced as follows:

$$y_{sim}(n) = \beta_{sim}(h(nT) * p(n)) + f(n) + \varepsilon(n) \quad (14)$$
$$f(n) = q_1 f_1(n) + q_2 f_2(n) + q_3 f_3(n)$$

which q_1, q_2 and q_3 are random numbers between [0,1] and their sum is always kept equal to one. $\varepsilon(n)$ is the white Gaussian noise with the standard deviation of 10. $p(n)$ is the stimulation profile which is a series of randomly positioned impulses corresponding to our experimental fMRI data acquisition described in Oddball task session. The spatial pattern of activity is shown in Figure (2). The simulation dataset is used for investigation of sensitivity of the proposed PLM method in activation detection comparing with other methods.

Rest State fMRI data

The rest state data is acquired using a T_2^*-weighted gradient echo single-shot echo-planner (EPI) sequence with $T_R = 1648ms$, $T_E = 45ms$, Flip Angle=90 and $FOV = 250 \times 250\,\text{mm}^2$. A total 256 EPI volumes were scanned from each subject. During data collection, the subject was at rest.

Oddball task fMRI data

The task given to the subjects in this experiment is known as the 'classic visual oddball paradigm' in the EEG literature. In this task, a train of equally spaced visual stimuli is presented to the subjects. There are two types of stimuli: the standard stimuli and the target stimuli. The standard events occur more frequently than the target events. The subjects are instructed to silently count the target stimuli and report the total number at the end of the experiment. In the present study, the standard visual stimulus was an image consisting of the string of white characters 'OOOOO' on a dark background, while the target image was the string of characters 'XXXXX'. Visual stimuli were delivered to the subject via a liquid crystal display (LCD) mounted on the MRI scanner's radio frequency (RF) head coil. The target events were distributed randomly among the four runs and 1024 trials, but it was ensured that there were at least eight frequent events between every pair of target events. Two dataset were acquired in same condition from two subjects with the same temporal stimulation patterns.

Results

Both methods of trend removal — in time and wavelet domain are applied to the simulated data. We also applied common linear detrending method to fMRI and copared it with our proposed methods. As it shows in Figure (3), the detected voxel in simulated data for wavelet domain PLM is more than time domain PLM but they both have a significant difference with linear detrending method which just removes a linear trend from time-series. Linear detrending can not detect active voxels as well as PLM approaches.

The rest dataset is used for calculating the empirical mean and variance of β .

It is also used for exploring the ability of methods in control of false alarm rate. As it shows in figure (4) the linear detrending is detected less false active voxels. Time domain PLM has better false alarm rate control than wavelet domain PLM. Real false alarm rate in all methods was higher than expected false alarm rate. In table (1) the estimation variance of β in wavelet domain and time domain PLM approaches is less than linear detrending. The estimation variances achieved by PLM approaches are approximately the same.

The PLMs and linear detrending methods are applied to oddball task data. The detected active areas are shown in figure (5). The linear detrending could not detect the 'Insula' and 'the posterior cingulated gyrus'. The PLM approaches are detected mentioned areas but detected active areas in wavelet domain PLM are wider than time domain PLM.

Conclusion

In this paper, we compared two methods of estimating a non parametric trend estimation and a model based detrending method (linear detrending). It is shown that wavelet domain PLM is more sensitive to detect active voxel, but this comes in price of weaker false alarm rate control.

Reference

[1] Hossein-Zadeha G. A., Ardekani BA., Soltanian-Zadeh H. "A signal subspace approach for modeling the hemodynamic response function in fMRI," Magnetic Resonance Imaging, Vol. 21, pp 835–843, 2003.

[2] Zhang, D., Lin, X., Raz, J. and Sower, M. F "Semiparametric stochastic mixed models for longitudinal data". J. Am. Statist. Ass., Vol. 93, pp 710-719, 1998.

[3] Chang X. and Qu L. , "Wavelet estimation of partially linear models" , Computational Statistics & Data Analysis, Vol. 47, No. 1, pp 31-48, 2004.

[4] Green, P.J., and Silverman, B.W. "Nonparametric regression and generalized linear models. A roughness penalty approach", Chapman & Hall, London.

[5] Craven, P., and Wahba, G. "Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation" Numer. Math. 31: pp 377-403, 1979.

[6] Hastie, T.J. and Tibshirani, R.J. "Generalized additive models", London: Chapman & Hall, 1990.

[7] Knuth KH, Ardekani BA, Helpern JA. "Bayesian estimation of a parameterized hemodynamic response functions in an event-related fMRI experiment". Proc. of ISMRM 2001; 3:1732.

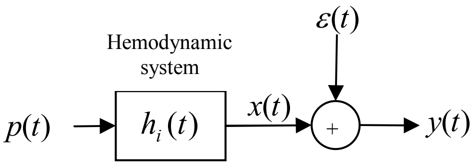

Figure 1. fMRI time-series linear model, *p(t)*is the profile of stimuli, $h_i(t)$ is the hemodynamic response function at voxel *i*, *ε(t)* is noise and artifact model and *y(t)* is observed data.

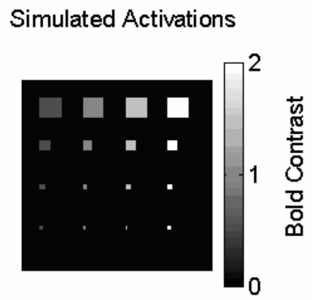

Figure 2. Activation levels profile in simulated time-series

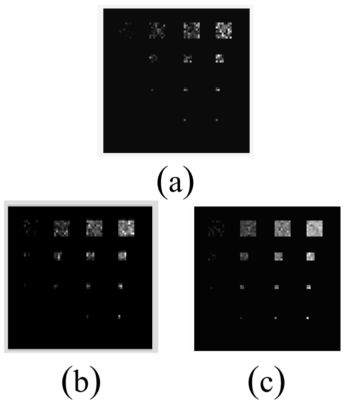

Figure 3. Detected active voxels using different methods of deternding. (a) Linear detrending (b) time domain PLM (c) wavelet domain PLM

Table 1. Estimation variance of β in rest data set.

PLM		Linear Detrending
Wavelet	Time-domain	
0.5285	0.5274	06512

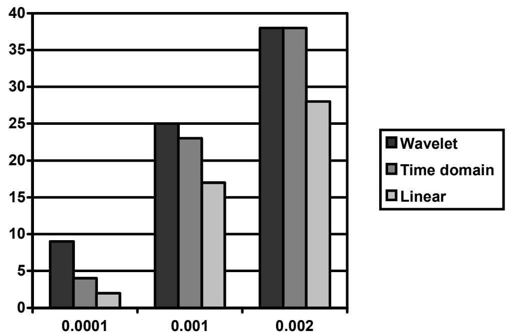

Figure 4. False Detected voxel in Rest state data

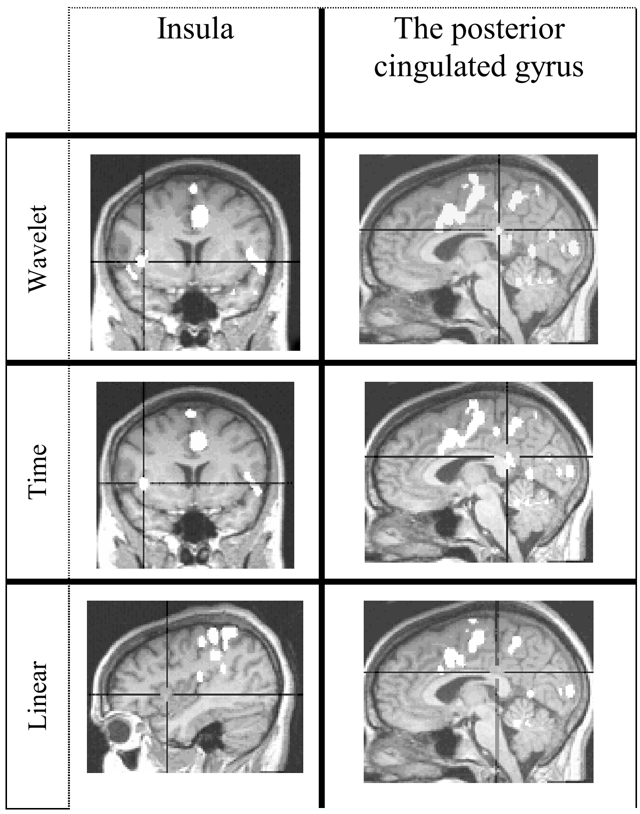

Figure 5. Detected active voxels using different approaches.