



HAL
open science

Penalized partially linear models using orthonormal wavelet bases with an application to FMRI time series

Jalal M. Fadili, E. T. Bullmore

► **To cite this version:**

Jalal M. Fadili, E. T. Bullmore. Penalized partially linear models using orthonormal wavelet bases with an application to FMRI time series. IEEE International symposium on biomedical imaging, 2004, Arlington, United States. pp.1171-1174, 10.1109/ISBI.2004.1398752 . hal-00263558

HAL Id: hal-00263558

<https://hal.science/hal-00263558>

Submitted on 12 Mar 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PENALIZED PARTIALLY LINEAR MODELS USING ORTHONORMAL WAVELET BASES WITH AN APPLICATION TO FMRI TIME SERIES

M. J. Fadili

GREYC UMR CNRS 6072
ENSICAEN 14050 Caen France

E. T. Bullmore

University of Cambridge
Brain Mapping Unit, Cambridge UK

ABSTRACT

In this paper, we consider modeling the non-parametric component in partially linear models (PLM) using orthogonal wavelet expansions. We introduce a regularized estimator of the non-parametric part in the wavelet domain. The key innovation here is that the non-parametric part can be efficiently estimated by choosing an appropriate penalty function for which the hard and soft thresholding estimators are particular cases. This avoids excessive bias in estimating the parametric component. We give an efficient estimation algorithm. A large scale simulation study is also conducted to illustrate the finite sample properties of the estimator. The estimator is finally applied to real neurophysiological functional MRI data sets that are suspected to contain both smooth and transient drift features.

1. INTRODUCTION AND PROBLEM POSITION

In the last ten years, there has been an increasing interest and activity in the area of partially linear regression in the statistical community. Many methods and techniques have been studied. A very useful work bringing an up-to-date presentation of the state of the art of semi-parametric regression techniques in various statistical problems can be found in [1]. Let's assume a linear dynamical model of the form :

$$y_t = X_t^T \beta + g(t) + \varepsilon_t, 1 \leq t \leq N \quad (1)$$

where N is the number of observations, $\beta = (\beta_1, \dots, \beta_p)^T$ is a vector of unknown parameters, g is essentially an arbitrary and unknown (possibly nonlinear) function over \mathbb{R} , $X_t = (X_{1t}, \dots, X_{pt})^T$ are vectors of explanatory variables that are either random i.i.d. or fixed design points, ε_t are i.i.d. error processes with $E(\varepsilon_t) = 0$ and finite variance $0 < E(\varepsilon_t^2) = \sigma^2 < \infty$, and the ε_t and X_t are mutually independent. In practice, there is a growing number of enormous data sets for time series analysts in finance, economics, geophysics, biological signals, where the PLM is of a prime interest and allows one to investigate the structural relationships between factors with a high flexibility. PLMs are semi-parametric models since they contain both

parametric and non-parametric components. In the context of PLM, interest focuses on the construction of efficient estimation procedures for both the parametric and non-parametric parts. Much attention has been directed to estimating Eq.1, see [1] for a review. Most authors have constructed estimation procedures based on either the kernel method, the local linear method (local polynomial or trigonometric polynomial), the orthogonal series or the spline smoothing approach. Estimators have been proposed for both the i.i.d. and correlated observations.

Following [2], under suitable conditions, the non-parametric part can be characterized by $\mathcal{W}\gamma$, where \mathcal{W} is a $N \times k$ -matrix, γ is an additional unknown parameter vector and k is unknown. The PLM in (1) can be then rewritten in a matrix form:

$$\mathbf{Y} = \mathbf{X}\beta + \mathcal{W}\gamma + \varepsilon \quad (2)$$

For identifiability, we here consider that 1_N is not spanned by the vectors of \mathbf{X} , i.e. the mean of the vector \mathbf{Y} is explained by \mathcal{W} . Our concern in this paper is modeling the non-parametric component using orthonormal wavelet expansions. In this case the matrix \mathcal{W} is a $N \times N$, $N = 2^J$, and the estimation problem is ill-posed. One can then impose side constraints to narrow down the class of candidate solutions and produce regularized estimates. Recently, authors in [3] introduced a wavelet-hard thresholding estimator (nonlinear) and the Minimum Description Length (MDL) principle to automatically select a subset of coefficients in the optimal basis representing the non-parametric part. In [4], a linear least-squares wavelet estimator for the non-parametric part was proposed. In their work, choosing the detail coefficients retained in the reconstruction amounts to choosing the coarsest decomposition scale which equivalent to a simple low-pass filtering whose dyadic cutoff frequency is selected by means of a model complexity criterion (e.g. Akaike or Bayesian Information Criteria (AIC, BIC)). In [5], authors proposed an L_1 penalty function which is equivalent to the same procedure as in [3] replacing the hard by soft thresholding. We here propose to unify these approaches using regularization constraints on γ .

By orthonormality of \mathcal{W} we propose to minimize the

following penalized least-squares (PLS) problem to estimate β and γ :

$$\chi^2(\beta, \gamma) = \frac{1}{2}(y_{w_a} - X_{w_a}^T \beta - \gamma_a)^2 + \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \left[\frac{1}{2}(y_{w_{j,k}} - X_{w_{j,k}}^T \beta - \gamma_{j,k})^2 + \lambda p(|\gamma_{j,k}|) \right] \quad (3)$$

where y_{w_a} and $y_{w_{j,k}}$ are the approximation coefficient at coarsest scale 0 and the detail coefficients at scale $j = 0, \dots, J-1$ and location $k = 0, \dots, 2^j - 1$ of the DWT of \mathbf{Y} . $X_{w_{(\cdot)}}$ and $\gamma_{(\cdot)}$ are defined similarly for X and γ , where $X_{w_{(\cdot)}}$ is a row in the column-wise DWT (\mathbf{X}_w) of the design matrix \mathbf{X} . The regularizing penalty function p is not necessarily convex on \mathbb{R}^+ and irregular at point zero to produce sparse solutions. However, it remains a challenging task to optimize such nonconvex functionals. In particular, the PLM hard and soft thresholding rules ([3, 5]) are special cases that correspond respectively to the smooth clipped penalty defined in Eq.7 [6] and the L_1 penalty. These estimators have unappealing features (e.g. excessive bias for the soft thresholding) and much more efficient estimators can be obtained with appropriately designed function p (e.g. Sobolev H^s norm that will be treated in the following).

The PLS for PLM estimation is attractive for several reasons. The principle of adding a penalty term to a sum of squares or more generally to a -log-likelihood applies to a variety of statistical noise models. There is also a Bayesian interpretation to Eq.3. Minimizing Eq.3 corresponds to the MAP estimator with a prior density $\exp(-\lambda p(|\gamma_{j,k}|))$ imposed on each $\gamma_{j,k}$.

2. PENALIZED PLM WAVELET ESTIMATOR

We now turn to the minimization problem in Eq.3. We are able to state the following:

Theorem 1 *Let $p(\cdot)$ be a nonnegative, nondecreasing and differentiable function on \mathbb{R}^+ . $\mathbf{Y} \in L^2([0, N-1])$ and 1_N is not spanned by the vectors of \mathbf{X} . The solution to the minimization problem 3 exists and is given by:*

$$\hat{\gamma}_a = y_{w_a} \quad (4)$$

$$\hat{\beta} = (\mathbf{X}_w^T \mathbf{X}_w)^{-1} \mathbf{X}_w^T (\mathbf{Y}_w - \hat{\gamma}) \quad (5)$$

$$\hat{\gamma}_{j,k} = \begin{cases} 0 & \text{if } |r_{j,k}| \leq p_0 \\ r_{j,k} - \lambda \text{sign}(r_{j,k}) p'(|\hat{\gamma}_{j,k}|) & \text{if } |r_{j,k}| > p_0 \end{cases} \quad (6)$$

where $r_{j,k} = y_{w_{j,k}} - X_{w_{j,k}}^T \hat{\beta}$ and $p_0 = \min_{r \geq 0} (r + \lambda p'(r))$.

The proof follows mainly from differentiation of Eq.3 and arguments from [6]. The estimate of β is the least-squares estimate when γ is known. The estimation of the non-parametric part is essentially a thresholding rule. One can easily verify that soft thresholding [5] is a special case where $p(\gamma) = |\gamma|$

and hard thresholding [3] corresponds to the penalty function:

$$p(|\gamma|) = \lambda - \lambda^{-1}(\gamma - \lambda)^2 I(|\gamma| < \lambda) \quad (7)$$

The approximation coefficient (mean of the signal) is not thresholded and is always retained in the reconstruction. Other possible penalty functions include L_q norms ($q \geq 0$) and expressions for the threshold p_0 can be derived. Another possible choice is the classical regularization in the wavelet domain using the Sobolev H^s norm [7] assuming $g \in H^s$:

$$p(|\gamma|) = \sum_{j,k} 2^{2js} |\gamma_{j,k}|^2 \quad (8)$$

where s is the Sobolev smoothness degree. Actually, one can use more general norms since the norm equivalence in the wavelet domain has been proved for any Besov space $B_{p,q}^s$ and used here for $p = q = 2$. Thus, for this penalty function, the PLS problem can be written in a matrix form:

$$\chi^2(\beta, \gamma) = \frac{1}{2} \|\mathbf{Y}_w - \mathbf{X}_w \beta - \gamma\|_2^2 + \lambda \|D\gamma\|^2 \quad (9)$$

where $D = \text{diag}(2^{js})$. This minimization problem has a flavor of ridge regression. It has a unique closed-form solution:

Proposition 1 *Let $g \in H^s$, $\mathbf{Y} \in L^2([0, N-1])$ and 1_N is not spanned by the vectors of \mathbf{X} . Then, minimizers of Eq.9 are:*

$$\hat{\gamma}_a = y_{w_a} \quad (10)$$

$$\hat{\beta}_{LS} = (\mathbf{X}_w^T \tilde{\mathbf{X}}_w)^{-1} \mathbf{X}_w^T \tilde{\mathbf{Y}}_w \quad (11)$$

$$\hat{\gamma}_{j,k} = \frac{y_{w_{j,k}} - X_{w_{j,k}}^T \hat{\beta}_{LS}}{1 + 2\lambda 2^{2js}}, \text{ for each } (j, k) \quad (12)$$

where $\mathbf{P}^\perp = \text{diag}(\frac{2\lambda 2^{2js}}{1+2\lambda 2^{2js}})$, $\tilde{\mathbf{X}}_w = \mathbf{P}^\perp \mathbf{X}_w$ and $\tilde{\mathbf{Y}}_w = \mathbf{P}^\perp \mathbf{Y}_w$.

The proof can be obtained either directly by differentiating Eq.9 and back-substitution or as a corollary of Theorem 1. This estimator can be seen as an extension to the wavelet domain of the kernel smoothing estimators [2].

2.1. The regularized PLM estimation algorithm

From the structure of the updating equations in Theorem 1, we naturally see that we can use the iterative (backfitting-like) algorithm to estimate the parametric and the non-parametric part of the PLM:

Algorithm 1 The penalized PLM wavelet estimator

- 1: For any given wavelet, calculate the column-wise DWT of \mathbf{Y} and \mathbf{X} . Set $l = 0$, $\gamma^{(0)} = 0$,
 - 2: **repeat**
 - 3: Update $\beta^{(l+1)} = (\mathbf{X}_w^T \mathbf{X}_w)^{-1} \mathbf{X}_w^T (\mathbf{Y}_w - \gamma^{(l)})$.
 - 4: Update $\gamma^{(l+1)}$ using Eq.4 and 6.
 - 5: **until** $\|\beta^{(l+1)} - \beta^{(l)}\|_2 < \epsilon$
 - 6: Calculate the inverse DWT of $\gamma^{(l)}$ and get an estimate for g .
-

As far as the Sobolev penalty is concerned, there is no need to iterate the process and the estimates are obtained after one pass from Proposition 1.

2.2. Selection of the regularization parameter

An important issue in our penalized PLM estimator is the choice of the regularization parameter which influences the threshold p_0 . One can use the traditional universal threshold value $\lambda_{UD} = \sigma\sqrt{2\log N}$. However, this threshold is known to be somewhat too large. Alternatively, one can use a lower universal threshold $\lambda_{UA} = \sigma\sqrt{2\log N - \log(1 + c^2 \log N)}$ ($c \geq 1$) derived by [6] in wavelet non-parametric regression. Another popular thresholding scheme is 3σ which is very close to λ_{UA} when $1 \leq c \leq 1.7$ for $N \in [2^6, 2^{11}]$.

The generalized cross-validation criterion (GCV) has been already adopted for objectively choosing the regularizing parameter in penalized non-parametric regression (see e.g. [7]). GCV attempts to provide a data-driven estimate of λ which minimizes the unobservable risk estimate. The GCV function in our Sobolev-penalty wavelet estimator of the PLM is then taken to be:

$$\text{GCV}(\lambda) = \frac{(N-1)^{-1} \sum_{j,k} (y_{w_{j,k}} - \hat{y}_{w_{j,k}})^2}{[(N-1) - \text{trace}(\mathbf{P} + \mathbf{Q}_{\mathbf{X}_w} \mathbf{P}^\perp)]^2} \quad (13)$$

where $\mathbf{Q}_{\mathbf{X}_w} = \tilde{\mathbf{X}}_w (\mathbf{X}_w^T \tilde{\mathbf{X}}_w)^{-1} \mathbf{X}_w^T$. One must be aware that in the GCV, only the detail coefficients are taken into account in matrices \mathbf{X}_w and \mathbf{P} which justifies the factor $N-1$ used.

2.3. Asymptotic properties and inference

Here, we do not present any theoretical study about the asymptotic properties of our wavelet-domain regularized estimator of the PLM. Instead, we support our claims using arguments from the PLM kernel smoothing asymptotics literature [1]. Our conjectures will be confirmed by simulations. It has been established that PLS estimators of β of PLMs using kernel smoothing are asymptotically Gaussian. They are also biased even if the true λ is known. Therefore, our estimator is also expected to be Gaussian and subject to bias. However, this bias will be negligible. As far as the estimator of g is concerned, its bias and variance is expected

to be asymptotically of the same order as if β were known exactly. Our Monte-Carlo simulations, reported hereafter, support these unproved conjectures.

In most applications, e.g. fMRI, one is essentially interested in testing the significance of an effect corresponding to a covariate in the design matrix. Then, the asymptotic normality and the estimated vector $\hat{\beta}$ (from Theorem 1 or Proposition 1), justifies an approximate t or F statistics for inference purposes. For example, one can write the t -statistic for the Sobolev-penalty regularized wavelet estimator as:

$$t = \frac{\hat{\beta}}{\sqrt{\text{Var}(\beta)}} \quad (14)$$

where

$$\text{Var}(\beta) = \frac{\|\mathbf{Y} - \mathbf{A}\mathbf{Y}\|_2^2}{\nu} \text{diag} \left((\mathbf{X}_w^T \tilde{\mathbf{X}}_w)^{-1} \right)$$

here $\mathbf{A} = \mathbf{P} + \mathbf{Q}_{\mathbf{X}_w} \mathbf{P}^\perp$ is the so-called hat matrix. The t statistic has $\nu = \text{trace}((\mathbf{I}_N - \mathbf{A})^T (\mathbf{I}_N - \mathbf{A}))$ effective degrees of freedom. Similar formulae can be derived for any penalty function (e.g. soft or hard thresholding, etc).

3. LIBRARY OF ORTHONORMAL BASES

A *library* of bases in a linear space is a collection of waveforms whose elements are constructed by choosing individual subsets from a large set of vectors. In the context of PLMs, we are mainly concerned about decomposing the non-parametric component over family of waveforms that are well localized both in time and frequency. The most suitable basis for a particular signal is adaptively selected from the library. This approach leads to a vastly more efficient representation for g compared to confining ourselves to a single basis. We here restrict ourselves to complete dictionaries (bases), thus containing exactly N atoms such as the orthogonal wavelet transform. For example, one can use different wavelet bases each corresponding to a specific mother wavelet. We propose to minimize any model selection measure (MDL as in [3], AIC, BIC or GCV) to pick up the best atoms subset representing g from the best dictionary in the library.

4. RESULTS AND DISCUSSION

4.1. Simulation study

We first illustrate the finite-sample behavior of the PLS estimator by performing a simulation study on some test functions g that have been widely used in the literature [1]. These signals have different smoothness and time-frequency localization properties ranging from C^∞ functions to highly non-stationary signals with jumps, isolated singularities and time-dependent frequency content. Traditional spline-like

PLM estimators are expected to work poorly on the latter functions. The noise σ was 1 and the factors of the simulation were : the design matrix \mathbf{X} , β , g and its amplitude $\alpha = \text{SNR}\sigma$, and the number of samples N . For each combination of these factors, a simulation run was repeated 500 times. For each simulation run, we computed the bias and PSNR on \hat{g} , the bias, the variance as well as the χ^2 distance to normality of the $\hat{\beta}$. Fig.1 shows an example of the performance of our PLS estimator on the Corner function which is piecewise smooth with two big jumps in its first derivatives. Three penalty functions were compared, namely the hard thresholding penalty, the Sobolev and the L_1 norms.

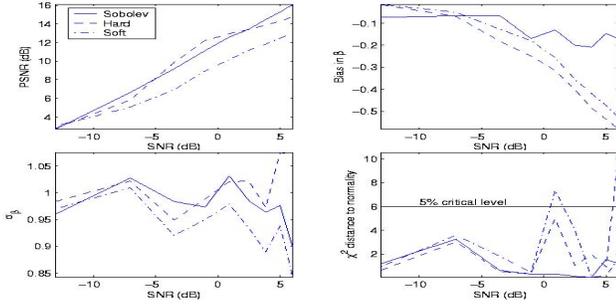


Fig. 1. Simulation results of the PLS estimator on the Corner function.

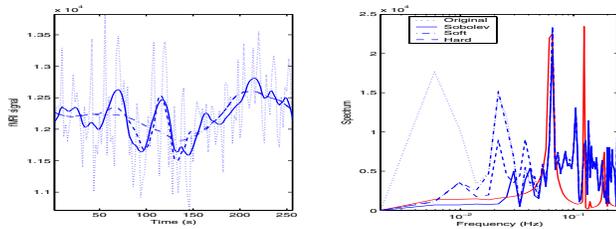


Fig. 2. Estimated drift and spectrum of detrended data

The output PSNR is comparable between the Sobolev and hard thresholding-based estimators but is higher than soft thresholding. This is consistent with non-parametric regression expectations. As expected from our claims, there is a bias (negative) in estimating β for all the penalty functions. At high SNRs, this bias is the smallest for the Sobolev norm and the highest for the hard penalty while the opposite tendency is observed at low SNRs. However, this bias was observed to decrease rapidly and becomes negligible with increasing N . For inference purposes, the statistical score in Eq.14 should be asymptotically normal with unit variance. Indeed, as shown by Fig.1 the standard deviation of this score is very close to one and tends to be somewhat underestimated in the case of the soft thresholding rule. The omnibus test χ^2 score to normality is also shown with the corresponding 5% critical level. Again, the Sobolev norm and the hard thresholding penalties give the best result. Therefore, if the main interest is inference on β (without underes-

timating g), then our results suggest that the Sobolev-based estimator is the most appropriate provided that the smoothness parameter s is known or can be efficiently estimated, e.g. by the GCV criterion. Otherwise, the hard thresholding constitutes a better alternative.

4.2. Application to fMRI data

To further illustrate our method, we also consider an application to real data sets in fMRI. Fig.2 shows the estimated trend superimposed on the original time course of a pixel in an fMRI event-related visual experiment ($N = 128$, stimulus at $f = 0.606\text{Hz}$ shown in red). Sobolev norm and hard thresholding rule give comparable estimates of the trend while soft thresholding provides an oversmooth estimate. For signals with isolated singularities (not shown), hard thresholding gave much better estimates unless the s was estimated accurately. After detrending, the low-frequency drift has been adaptatively removed without altering the stimulus component at 0.606Hz. Furthermore, we have observed that the residuals are Gaussian while serious departure from normality was observed before detrending.

5. CONCLUSION

In this paper, a flexible and powerful PLS estimator with orthonormal wavelet library was proposed for PLMs. The model can be easily extended to long-memory noise case. For overcomplete dictionaries, the Block Coordinate Relaxation algorithm can be used efficiently to solve the PLS minimization problem. This is the area of ongoing research in our group.

6. REFERENCES

- [1] W. Härdle, H. Liang, and J. Gao, *Partially Linear Models*, Springer Verlag, 2000.
- [2] P. Speckman, "Kernel smoothing in partial linear models," *Journal of the Royal Statistical Society Series B*, vol. 50, pp. 413–436, 1988.
- [3] M.J. Fadili and E.T. Bullmore, "Semiparametric time series regression using library of orthonormal bases and the MDL," Tech. Rep., GREYC CNRS, France, 2001.
- [4] F.G. Meyer, "Wavelet based estimation of a semi parametric generalized linear model of fmri time-series," *EEE Trans. on Med. Imaging*, vol. 22, no. 3, pp. 315–322, 2003.
- [5] X.W. Chang and L. Qu, "Wavelet estimation of partially linear models," Tech. Rep. SOCS 03.4, School of Computer Science, McGill University, 2003.
- [6] A. Antoniadis and J. Fan, "Regularization of wavelets approximations," *J. Amer. Statist. Assoc.*, vol. 96, no. 455, pp. 939–963, 2001.
- [7] U. Amato and D.T. Vuza, "Wavelet approximation of a function from samples affected by noise," *Revue Roumaine Math. Pures Appl.*, vol. 42, pp. 481–493, 1997.