DETERMINISTIC REGRESSION SMOOTHNESS PRIORS TVAR MODELLING

J.P. Kaipio and M. Juntunen

Department of Applied Physics, University of Kuopio
P.O.Box 1627, FIN-70211, Kuopio, FINLAND.
Email: jari.kaipio@uku.fi

ABSTRACT

In this paper we propose a method for the estimation of time-varying autoregressive processes. The approach is essentially to regularize the heavily underdetermined unconstrained prediction equations with a smoothness priors type side constraint. The implementation of nonhomogenous smoothness properties is straightforward. The method is compared to the usual determistic regression approach (TVAR) in which the coefficient evolutions are constrained to a subspace. It is shown that the typical transient oscillations of TVAR can be avoided with the proposed method.

1. INTRODUCTION

In the modelling and estimation of nonstationary processes the direct estimation of the nonstationary covariance matrix usually requires much more data than is available. For this reason parametric models with time-varying coefficients are often employed. If the coefficients at each time are treated as (structurally) independent parameters, the number of parameters is usually greater than the amount of data. There are two main approaches to overcome this problem.

The first approach is to constrain the corresponding time-varying coefficients to a subspace, that is, they are constrained to be linear combinations of a set of basis functions [1, 2, 3]. This is called the deterministic regression approach. The main problem in deterministic regression is the choice of the basis functions (and the corresponding subspace). The resulting estimation problem acquires also ill-posed character when the number of basis functions is increased. On the other hand, the basis functions determine the approximation capabilities of the method. In most cases the selection of the basis functions has been more or less ad hoc. Popular choices for the basis functions include Fourier bases (sinusoids), polynomial bases and prolate spheroidal wave bases. If something is known of the characteristics of the nonstationarity, the basis functions have to be constructed specially for each case. In [4] a principal component type approach was used in the construction of a set of optimal basis functions for a special nonstationarity model (a single smooth transition). This construction is, however, rather involved and it is not applicable to such cases in which the information on the nonstationarity is ambiguous.

The other approach is to use adaptive algorithms such as the LMS and RLS algorithms and the Kalman filter [5]. This is called the stochastic regression approach. The usual implementation is to "connect" the adaptive algorithms to yield one step predictors. In the LMS and RLS algorithms only a coarse adjustment of the tracking properties can be made. With the Kalman filter more sophisticated evolution models can be used, although the simplest possible model, the random walk model, is the most popular one. In the random walk model the implicit assumption is that the rate of change of the coefficient evolutions are small. However, in many applications, such as EEG analysis, all expectable changes are not necessarily slow.

In [6, 7, 8] the Kalman filter was employed with the smoothness prior (SP) evolution model. In the SP model the norms of some higher differences of the coefficient evolutions are assumed to be small. This approach allows relatively fast changes without introducing excessive noisiness of the estimates. However, the stochastic regression approaches are not always preferred since they they sometimes necessitate several runs to obtain initial (and possibly final) estimates for the coefficients and certain covariance matrices. Such situations occur especially with short data segments.

In this paper we propose a nonparametric deterministic regression scheme that does not utilize subspace constraints. Instead, the approach is based on interpreting the heavily underdetermined unconstrained parameter estimation problem as an ill-posed inverse problem and using Tikhonov-type regularization with smoothness priors type regularizing side constraints. Furthermore, we suggest the use of nonhomogenous weighting of the associated regularization operators and the use of several difference side constraints simultaneously. Although the approach is well suited to more general time series models, we consider here only the time-varying AR(p) estimation case.

2. METHODS

The general form of a TVAR(p) model for a process x_t is

$$x_t = \sum_{k=1}^{p} a_k(t) x_{t-k} + e_t \ , \tag{1}$$

where $a_k(t)$ are the coefficient evolutions and e_t is the prediction error process. If the process x_t can be approximated with the TVAR(p) model, the prediction error process can be used to approximate the associated innovation process. However, if no constraints are imposed on the coefficient evolutions, the estimation of the coefficient evolution is clearly a meaningless task.

2.1. Deterministic regression TVAR models

In the deterministic regression TVAR problem the coefficients are constrained to

$$a_k(t) = \sum_{\ell=0}^{M} c_{k\ell} \phi_\ell(t) \ , \tag{2}$$

where $\phi_\ell(t)$, $\ell = 0, \ldots, M$ are the basis functions.

The minimization of the 2-norm of the residuals in (1) with the constraints (2) leads to a quadratic problem with the $(M+1)p$ parameters $c_{k\ell}$. The TVAR problem was introduced first in [1] and has thereafter been partially reformulated and applied to EEG and speech modelling for example in [2, 9, 4, 10].

The traditional method for the solution of the constrained LS problem (1–2) is the covariance formulation [2]. The LS solution can be accessed more conveniently by writing $c = (c_{10}, \ldots, c_{1M}, \ldots, c_{p0}, \ldots, c_{pM})^{\mathrm{T}}$, $X = (x_{p+1}, \ldots, x_T)^{\mathrm{T}}$, $E = (e_{p+1}, \ldots, e_T)^{\mathrm{T}}$ and regressor matrix $H = (H_{10}, \ldots, H_{1M}, \ldots, H_{p0}, \ldots, H_{pM})$ where $H_{k\ell} = (\phi_\ell(p+1) x_{p+1-k}, \ldots, \phi_\ell(T) x_{T-k})^{\mathrm{T}}$ and $(\cdot)^{\mathrm{T}}$ denotes transpose [11]. The least squares problem can then be stated as

$$\min_c \|X - Hc\|_2 \ . \tag{3}$$

2.2. Deterministic regression smoothness priors TVAR model

Define the matrices $K_\ell \in \mathbb{R}^{(T-p-1)\times T}$ with elements

$$K_\ell(i,j) = x_{i+p-\ell} \delta_{j-i-p} \ , \quad \ell = 1, \ldots, p$$

where $\delta_{(\cdot)}$ denotes the Kronecker symbol. The unconstrained LS problem that corresponds to (1) can then be written in the form

$$\min_a \|X - Ka\|_2 \ , \tag{4}$$

where

$$\begin{aligned} K &= (K_1, \ldots, K_p) \\ a &= (a_1(1), \ldots, a_1(T), \ldots, a_p(1), \ldots, a_p(T))^{\mathrm{T}} \end{aligned}$$

and where X is as above. It is clear that the null space of the matrix K has at least dimension $pT - T + p$ and that the pseudoinverse solution has usually little to do with the actual coefficient evolutions. Thus the problem (4) exhibits the characteristics of ill-posed inverse problems.

One of the most popular approaches in the solution of inverse problems is to use Tikhonov regularization [12] that involves the solution of the problem

$$\min_a \left\{ \|X - Ka\|_2^2 + W(a) \right\} \ , \tag{5}$$

where $W(a) \geq 0$ is the side constraint functional, that should be small for all feasible solutions a. The most common side constraints are

$$\begin{aligned} W(a) &= \alpha^2 \|a\|_2^2 \\ W(a) &= \alpha^2 \|\mathcal{D}^\ell a\|_2^2 \end{aligned}$$

where α is the regularization parameter and $\mathcal{D}^\ell$ denotes the ℓ'th difference operator.

We seek thus the solution of the modified problem

$$\min_a \left\{ \|X - Ka\|_2^2 + \sum_{k=1}^{p} \sum_{\ell=0}^{M} \|\alpha_{k,\ell}(t) \mathcal{D}^\ell a_k(t)\|_2^2 \right\} \tag{6}$$

where $\alpha_{k,\ell}(t)$ are weight functions that are constructed based on the a priori assumptions on the evolution properties of the time-varying coefficients. We call the solutions corresponding to (6) as deterministic regression smoothness priors (DRSP) estimates.

Usually we set $\alpha_{k,\ell}(t) \equiv \alpha_{k,\ell}$ but sometimes other feasible selections are possible, for instance, when a sharp transition is expected at an approximately known instant.

The problem (6) can be written in the augmented least squares form as $\min_a \|\widetilde{X} - \widetilde{K}a\|_2$ where

$$\begin{aligned} \widetilde{K} &= \begin{pmatrix} K \\ L \end{pmatrix} \\ \widetilde{X} &= (X^{\mathrm{T}}, 0, \ldots, 0)^{\mathrm{T}} \end{aligned}$$

$$L = \begin{pmatrix} \widetilde{\alpha}_{1,0}\mathcal{D}^0 & & & \\ \vdots & & & \\ \widetilde{\alpha}_{1,M}\mathcal{D}^M & & & \\ & \widetilde{\alpha}_{2,0}\mathcal{D}^0 & & \\ & \vdots & & \\ & \widetilde{\alpha}_{2,M}\mathcal{D}^M & & \\ & & \ddots & \\ & & & \widetilde{\alpha}_{p,0}\mathcal{D}^0 \\ & & & \vdots \\ & & & \widetilde{\alpha}_{p,M}\mathcal{D}^M \end{pmatrix}$$

or if (as is usually appropriate) $\alpha_{k,\ell}(t) = \alpha_\ell(t)$, we can write

$$L = I_p \otimes \begin{pmatrix} \widetilde{\alpha}_0\mathcal{D}^0 \\ \vdots \\ \widetilde{\alpha}_M\mathcal{D}^M \end{pmatrix}$$

where $\widetilde{\alpha}_{k,\ell}$ are diagonal weighting matrices corresponding to $\alpha_{k,\ell}(t)$, I_p is a $p\times p$ identity matrix and $\otimes$ denotes the Kronecker (tensor) product.

The matrix $\widetilde{K}$ is very sparse which means that specialized least squares algorithms can be used. The density of $\widetilde{K}$ is of the order of $(M+1)/T$. See e.g. [13] for computational issues of sparse least squares problem. Thus the computational burden in the solution of (6) is not very big.

3. A SIMULATION STUDY

We study the estimation of the time-varying coefficients of two nonstationary AR(2) processes. The coefficients of the first example process vary smoothly while the coefficients of the second process contain an abrupt jump that is clearly visible from the data, see Figs. 1 and 2.

The TVARLS and DRSP estimates are computed for 20 realizations of the test processes and the means and standard deviations for the coefficient estimates are calculated. In both cases the weighting functions $\alpha_{k,\ell}(t)$ were given constant values except for an expected coefficient jump at about $t = 128$ where a notch was introduced.

In TVARLS a Fourier basis with $M = 11$ basis functions was used (the use of more basis functions resulted in rank deficiency of the matrix H. In both cases the relevant parameters of DRSP were $N = 3$, $\alpha_{k,1} = 10$, $\alpha_{k,1} = 1000$, $k = 1, 2$ (except for the notch), $\alpha_{k,\ell} = 0$, $\ell \neq 1, 3$.

The results for the coefficient $a_1(t)$ are shown in Figs. 3–6. The main differences between the TVARLS and DRSP estimates are that 1) the bias of the DRSP coefficient estimates is smaller and 2) that the DRSP solutions have considerably smaller tendency to yield oscillatory behaviour near the end points and the discontinuities of the coefficient evolutions. Without the nonhomogenous weighting the DRSP estimates exhibit almost as severe oscillating properties as the TVARLS estimates.

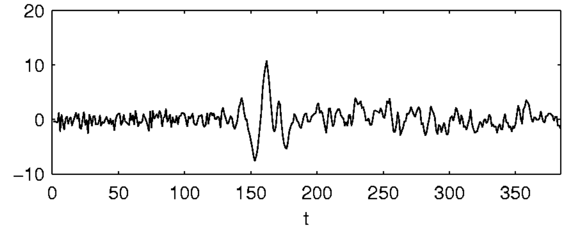

Figure 1: a) A realization of the process x_t with smoothly evolving coefficients.

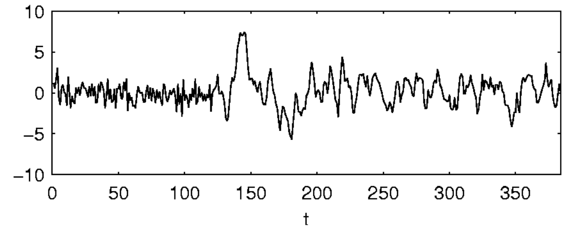

Figure 2: A realization of the process x_t with an abrupt jump in the coefficients. The step occurs at time $t = 128$.

4. CONCLUSIONS

We have proposed an approach to the estimation of time-varying parametric models. The approach is based on Tikhonov-type regularization of the unconstrained underdetermined parameter estimation problem. As the regularization side constraint we use a "weighted Sobolev norm" generalization of the smooth-

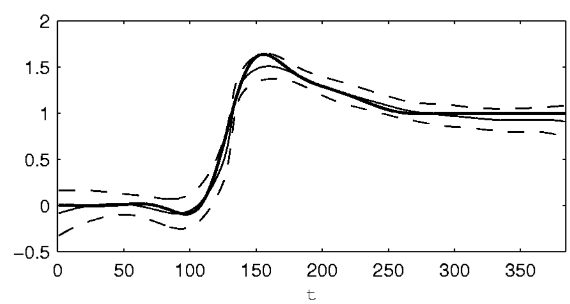

Figure 3: The true coefficient evolution $a_1(t)$ (bold), the mean SP estimate and the corresponding standard deviation interval: smooth evolution case.

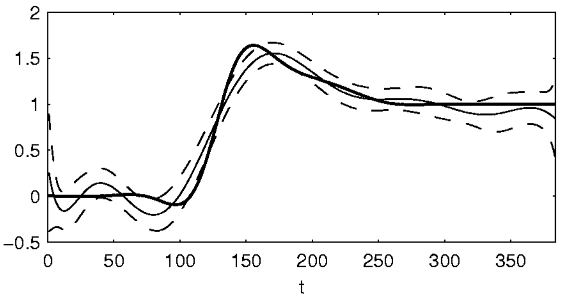

Figure 4: The true coefficient evolution $a_1(t)$ (bold), the mean TVARLS estimate and the corresponding standard deviation interval. $M = 11$ Fourier basis functions $\phi_m(t)$ were used: smooth evolution case.

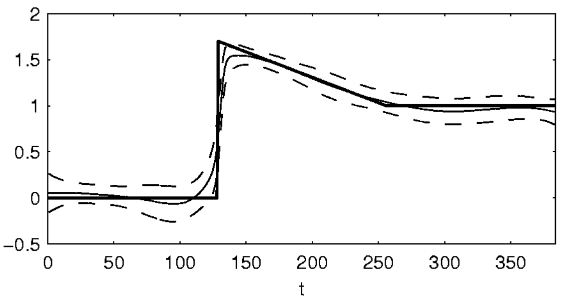

Figure 5: The true coefficient evolution $a_1(t)$ (bold), the mean SP estimate and the corresponding standard deviation interval: nonsmooth evolution case.

ness priors idea, that Kitagawa and Gersch used as evolution model in an earlier stochastic regression approach. The proposed method bears also a connection to more general Bayesian time series models since the side constraint can be shown to correspond to the prior distribution of the parameters [14]. This connection is especially clear if all associated distributions can be assumed to be Gaussian.

5. REFERENCES

[1] T. S. Rao, "The fitting of non-stationary time series models with time-dependent parameters," *J R Statist Soc*, vol. B 32, pp. 312–322, 1970.

[2] L. Liporace, "Linear estimation of nonstationary signals," *J Acoust Soc Am*, vol. 58, pp. 1288–1295, 1975.

[3] W. Gersch, A. Gevins, and G. Kitagawa, "A multivariate time varying autoregressive modeling of nonstationary covariance time series," in *IEEE Conf. CDC-83*, pp. 579–584, 1983.

[4] J. Kaipio and P. Karjalainen, "Estimation of event

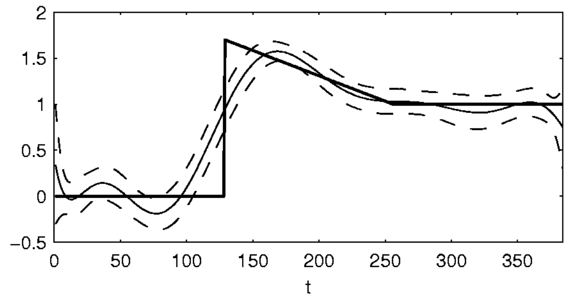

Figure 6: The true coefficient evolution $a_1(t)$ (bold), the mean TVARLS estimate and the corresponding standard deviation interval. $M = 11$ Fourier basis functions $\phi_m(t)$ were used: nonsmooth evolution case.

related synchronization changes by a new TVAR method," *IEEE Trans Biomed Eng*, vol. 44, no. 8, 1997.

[5] A. Isaksson, A. Wennberg, and L. Zetterberg, "Computer analysis of EEG signals with parametric models," *Proc IEEE*, vol. 69, pp. 451–461, 1981.

[6] G. Kitagawa and W. Gersch, *Smoothness Priors Analysis of Time Series.* Springer-Verlag, 1996.

[7] G. Kitagawa and W. Gersch, "A smoothness priors time-varying AR coefficient modeling of nonstationary covariance time series," *IEEE Trans Automat Contr*, vol. 30, pp. 48–56, 1985.

[8] W. Gersch, "Smoothness priors," in *New Directions in Time Series Analysis, Part II*, pp. 113–146, Springer-Verlag, 1991.

[9] W. Gersch and J. Yonemoto, "Parametric time series models for multivariate EEG analysis," *Comput Biomed Res*, vol. 10, pp. 113–125, 1977.

[10] J. Hiltunen, P. Karjalainen, and J. Kaipio, *Estimation of Medium Rate EEG Transitions.* EMBS97 Conference, 1997.

[11] J. Kaipio, *Simulation and Estimation of Nonstationary EEG.* PhD thesis, University of Kuopio, 1996.

[12] C. Groetsch, *Inverse Problems in the Mathematical Sciences.* Vieweg, 1993.

[13] Å. Björck, *Numerical methods for least squares problems.* SIAM, 1996.

[14] A. Tarantola, *Inverse Problem Theory.* Elsevier, 1987.