A Solution to the Discrete-Time Linear Prediction Problem Using PCA*

Rosa Fernández-Alcalá, Jesús Navarro-Moreno, and Juan C. Ruiz-Molina

Department of Statistics and Operations Research
University of Jaén
23071 Jaén, Spain
(e-mail: rmfernan@ujaen.es, jnavarro@ujaen.es, jcruiz@ujaen.es)

Abstract. A recursive solution is given to the linear one-stage prediction problem in discrete-time systems involving correlated signal and noise. Using Principal Component Analysis of stochastic processes, a suboptimum filter is designed. The main advantage of this solution is that it can be computed through a Kalman-like filter in those situations in which the signal does not verify a state-space model. The efficiency of the proposed methodology lies in the possibility of representing adequately the processes involved by a sample of points not excessively large.
Keywords: Linear Prediction Problem, PCA.

1 Introduction

In this paper we treat the discrete linear one-stage prediction problem involving correlated signal and noise. This estimation problem is useful in applications to feedback control and feedback communications. Thus, let $\{x(t_i), t_1 \leq t_i \leq t_n\}$ be a signal process which is a real second-order stochastic process, with zero-mean and correlation function R_x. Let $\{z(t_i), t_1 \leq t_i \leq t_n\}$ be a second-order stochastic process with zero-mean and correlation function R_z.

We assume that the signal process is observed corrupted by an additive white noise through the equation

$$y(t_i) = x(t_i) + v(t_i), \qquad t_1 \leq t_i \leq t_n$$

where $v(t_i)$ is a zero-mean white noise process with $E[v(t_i)v(t_j)] = r_i\delta_{ij}$ and correlated with both the signal $x(t_i)$ and the process $z(t_i)$. Let $R_{x_1x_2}(t_i, t_j)$ denote the correlation function between any two processes $x_1(t_i)$ and $x_2(t_j)$.

Under the above hypotheses, we consider the problem of finding the linear minimum variance estimator $\hat{z}(t_{k+1}/t_k)$ of the process $z(t_{k+1})$, based on the set of observations $\{y(t_1), \ldots, y(t_k)\}$, with $k < n$.

* This work was supported in part by Project MTM2004-04230 of the Plan Nacional de I+D+I, Ministerio de Educación y Ciencia, Spain. This project is financed jointly by the FEDER.

According to the projection theorem, this element $\hat{z}(t_{k+1}/t_k)$ exists, is unique and can be expressed as a linear transform of the observations set $\{y(t_1), \ldots, y(t_k)\}$ of the form [Poor, 1994]

$$\hat{z}(t_{k+1}/t_k) = \mathbf{h}_k'(t_{k+1})\mathbf{y}_k \tag{1}$$

where $\mathbf{y}_k = [y(t_1), \ldots, y(t_k)]'$ and the vector of optimum coefficients $\mathbf{h}_k(t_{k+1}) = [h_1(t_{k+1}), \ldots, h_k(t_{k+1})]'$ satisfies the Wiener-Hopf equation

$$\boldsymbol{\sigma}_k(t_{k+1}) = \boldsymbol{\Sigma}_{k\times k}(t_k)\mathbf{h}_k(t_{k+1}) \tag{2}$$

where $\boldsymbol{\sigma}_k(t_{k+1}) = [R_{zy}(t_{k+1}, t_1), \ldots, R_{zy}(t_{k+1}, t_k)]'$, with $R_{zy}(t_{k+1}, t_i) = R_{zx}(t_{k+1}, t_i) + R_{zv}(t_{k+1}, t_i)$, and $\boldsymbol{\Sigma}_{k\times k}(t_k)$ is the correlation matrix of the vector $\mathbf{y}_k$ whose elements are $R_y(t_i, t_j) = R_x(t_i, t_j) + R_{xv}(t_i, t_j) + R_{vx}(t_i, t_j) + r_i\delta_{ij}$.

Then, the estimation problem is basically that of solving the equation (2) involving the correlation functions of the signal process and the process to be estimated. In principle, this equation is easy to solve and its solution is given by

$$\mathbf{h}_k(t_{k+1}) = \boldsymbol{\Sigma}_{k\times k}^{-1}(t_k)\boldsymbol{\sigma}_k(t_{k+1}) \tag{3}$$

Unfortunately, from the practical point of view, the determination of these optimum coefficients through the equation (3) can lead to a computational difficulty since the inversion of the matrix $\boldsymbol{\Sigma}_{k\times k}(t_k)$ makes that the number of basic computational operations grows linearly with the number of observations considered.

Recently, an extensive literature concerning the design of a more efficient computational procedure has been developed. One of the most used techniques consists in imposing additional structural conditions on the correlations involved such as, stationarity [Poor, 1994], state-space models which lead to the Kalman filter [Kalman and Bucy, 1961], semi-degenerate kernel forms [Sugisaka, 1983], among others. Although this approach is widely applied, there is a great number of physical phenomena that do not satisfy these assumptions. In these situations, an alternative methodology is possible by using Principal Component Analysis (PCA) of stochastic processes [Aguilera *et al.*, 1995, Aguilera *et al.*, 1996].

In this paper, we propose a new recursive one-stage prediction procedure following this second perspective. In this framework, by considering any truncated series representation for the involved processes in terms of their principal components, the vector of optimum coefficients (3), and then the optimum one-stage predictor (1), can be approximated. Although a sub-optimum one-stage predictor is provided, the main advantage of this via of solution is that it can be efficiently computed through a recursive algorithm without imposing any structural assumption on the processes involved. In fact, they can be applied under the only hypothesis that the involved correlation functions are known. This occurs frequently in applications to system

identification problems or in statistical communication theory, where the relevant statistics of the problem are initially known in terms of correlation functions derived from measurements or mathematical models [Gardner and Franks, 1971]. In particular, these results can be applied in detection problems [Kailath, 1970] and in feedback communication systems [Gardner, 1975].

Then, the rest of the paper is structured as follows. In the next subsection, a brief description about the orthogonal representation of a stochastic process in terms of its principal components is included. The main characteristic of these series expansions is that they allow us to represent adequately a process through a short number of terms. Next, in Section 2, a new methodology based on these series representations is developed with the aim of designing a suboptimum one-stage predictor which can be efficiently computed through a Kalman-like recursive algorithm.

1.1 Approximate Series Expansions Using PCA

Let us consider the random vector $\mathbf{z}_{2n} = [z(t_1), \ldots, z(t_n), x(t_1), \ldots, x(t_n)]'$.

Let $\mathbf{a}_{2n}(i) = [a_1(i), \ldots, a_{2n}(i)]'$ and λ_i denote the principal values and the principal factors, respectively. Let also b_i be the principal components obtained from the principal factors as $b_i = \mathbf{a}'_{2n}(i)\mathbf{z}_{2n}$[1].

Then, $\mathbf{z}_{2n}$ admits the following orthogonal representation in terms of its principal components:

$$\mathbf{z}_{2n} = \sum_{i=1}^{2n} \mathbf{a}_{2n}(i) b_i \tag{4}$$

Moreover, this representation is optimal in the sense of being the best $2n$-dimensional linear model for $\mathbf{z}_{2n}$ in the least squares sense [Fukunaga and Koontz, 1970].

From (4), we have that the processes $z(t_j)$ and $x(t_j)$ can be expressed through finite series expansions in terms of their principal components as follows:

$$z(t_j) = \sum_{i=1}^{2n} a_j(i) b_i, \quad x(t_j) = \sum_{i=1}^{2n} a_{j+n}(i) b_i, \qquad j = 1, \ldots, n$$

On the other hand, the correlation functions involved in (2) can be expressed by the following product of matrices:

$$\begin{aligned} R_x(t_k, t_j) &= \mathbf{d}'_{2n}(t_k) \boldsymbol{\Lambda}_{2n \times 2n} \mathbf{d}_{2n}(t_j) \\ R_{xv}(t_k, t_j) &= \mathbf{d}'_{2n}(t_k) \mathbf{f}_{2n}(t_j) \\ R_{zx}(t_k, t_j) &= \mathbf{c}'_{2n}(t_k) \boldsymbol{\Lambda}_{2n \times 2n} \mathbf{d}_{2n}(t_j) \\ R_{zv}(t_k, t_j) &= \mathbf{c}'_{2n}(t_k) \mathbf{f}_{2n}(t_j) \end{aligned} \tag{5}$$

[1] Note that, $E[b_i] = 0$ and $E[b_i b_j] = \lambda_i \delta_{ij}$.

where $\mathbf{d}_{2n}(t_j) = [a_{j+n}(1), \ldots, a_{j+n}(2n)]'$, $\mathbf{c}_{2n}(t_j) = [a_j(1), \ldots, a_j(2n)]'$, $\boldsymbol{\Lambda}_{2n\times 2n}$ is the $2n$-dimensional diagonal matrix whose elements are the principal values λ_i, and $\mathbf{f}_{2n}(t_j)$ is the $2n$-dimensional vector with elements $f_i(t_j) = E[v(t_j)b_i]$, for $i = 1, \ldots, 2n$.

Finally, note that a suitable representation of any stochastic process is possible without taking all the samples but that it is sufficient to select an adequate subset of them [Fukunaga and Koontz, 1970]. Then, we can select $m < n$ instants of times, $t_1 \leq t_{i_1} < t_{i_2} <, \ldots, < t_{i_m} < t_n$, and consider the vector $[z(t_{i_1}), \ldots, z(t_{i_m}), x(t_{i_1}), \ldots, x(t_{i_m})]'$. Next, using the principal values $\tilde{\lambda}_i$, the principal factors $\tilde{\mathbf{a}}_{2m}(i) = [\tilde{a}_1(i), \ldots, \tilde{a}_{2m}(i)]'$ and the principal components $\tilde{b}_i$ associated with this vector, $\mathbf{z}_{2n}$ can be approximated by the series expansion

$$\mathbf{z}_{2n} \approx \tilde{\mathbf{z}}_{2n} = \sum_{i=1}^{2m} \tilde{\mathbf{g}}_{2n}(i)\tilde{b}_i \tag{6}$$

where $\tilde{\mathbf{g}}_{2n}(i)$ is a $2n$-dimensional vector whose elements are of the form

$$\tilde{g}_j(i) = \frac{1}{\tilde{\lambda}_i} E\left[z(t_j)\tilde{b}_i\right] = \frac{1}{\tilde{\lambda}_i} \sum_{k=1}^{m} \left(\tilde{a}_k(i) R_z(t_j, t_{i_k}) + \tilde{a}_{m+k}(i) R_{zx}(t_j, t_{i_k})\right)$$

$$\tilde{g}_{j+n}(i) = \frac{1}{\tilde{\lambda}_i} E\left[x(t_j)\tilde{b}_i\right] = \frac{1}{\tilde{\lambda}_i} \sum_{k=1}^{m} \left(\tilde{a}_k(i) R_{xz}(t_j, t_{i_k}) + \tilde{a}_{m+k}(i) R_x(t_j, t_{i_k})\right)$$

for $j = 1, \ldots, n$.

The main advantage of the series expansion (6) with respect to (4) is the reduction of the computational burden. In fact, the amount of computation required depends on the number of points selected, m, and a criterion for determining a suitable m can be found in [Fukunaga and Koontz, 1970].

Now, the processes $z(t_j)$ and $x(t_j)$ can be approximated by finite series expansions with less number of terms as follows:

$$z(t_j) \approx z_m(t_j) = \sum_{i=1}^{2m} \tilde{g}_j(i)\tilde{b}_i, \quad x(t_j) \approx x_m(t_j) = \sum_{i=1}^{2m} \tilde{g}_{j+n}(i)\tilde{b}_i, \quad j = 1, \ldots, n$$

Moreover, the correlation functions given in (5) can be approximated by the product of matrices of reduced dimension. Specifically,

$$\begin{aligned} R_x(t_k, t_j) \approx R_{x_m}(t_k, t_j) &= \tilde{\mathbf{d}}'_{2m}(t_k)\tilde{\boldsymbol{\Lambda}}_{2m\times 2m}\tilde{\mathbf{d}}_{2m}(t_j) \\ R_{xv}(t_k, t_j) \approx R_{x_m v}(t_k, t_j) &= \tilde{\mathbf{d}}'_{2m}(t_k)\tilde{\mathbf{f}}_{2m}(t_j) \\ R_{zx}(t_k, t_j) \approx R_{z_m x_m}(t_k, t_j) &= \tilde{\mathbf{c}}'_{2m}(t_k)\tilde{\boldsymbol{\Lambda}}_{2m\times 2m}\tilde{\mathbf{d}}_{2m}(t_j) \\ R_{zv}(t_k, t_j) \approx R_{z_m v}(t_k, t_j) &= \tilde{\mathbf{c}}'_{2m}(t_k)\tilde{\mathbf{f}}_{2m}(t_j) \end{aligned} \tag{7}$$

where $\tilde{\mathbf{d}}_{2m}(t_j) = [\tilde{g}_{j+n}(1), \ldots, \tilde{g}_{j+n}(2m)]'$, $\tilde{\mathbf{c}}_{2m}(t_j) = [\tilde{g}_j(1), \ldots, \tilde{g}_j(2m)]'$, $\tilde{\boldsymbol{\Lambda}}_{2m\times 2m}$ is the $2m$-dimensional diagonal matrix with i-th entry $\tilde{\lambda}_i$, and

$\tilde{\mathbf{f}}_{2m}(t_j)$ is the $2m$-dimensional vector with elements $\tilde{f}_i(t_j)$, for $i = 1, \ldots, 2m$, of the form

$$\tilde{f}_i(t_j) = E[v(t_j)\tilde{b}_i] = \sum_{k=1}^{m} \left(\tilde{a}_k(i) R_{vz}(t_j, t_{i_k}) + \tilde{a}_{m+k}(i) R_{vx}(t_j, t_{i_k}) \right)$$

2 Suboptimum Predictor

In this section, a recursive suboptimum solution to the linear least mean-square one-stage prediction problem in discrete-time systems involving correlated signal and noise is devised. For that, the following approximate version of (2) is considered by taking the approximate representations (7) for the correlation functions involved:

$$\tilde{\boldsymbol{\sigma}}_k(t_{k+1}) = \tilde{\boldsymbol{\Sigma}}_{k\times k}(t_k)\tilde{\mathbf{h}}_k(t_{k+1}) \tag{8}$$

where $\tilde{\boldsymbol{\sigma}}_k(t_{k+1}) = [R_{z_m y_m}(t_{k+1}, t_1), \ldots, R_{z_m y_m}(t_{k+1}, t_k)]'$, and $\tilde{\boldsymbol{\Sigma}}_{k\times k}(t_k)$ is the correlation matrix of $\tilde{\mathbf{y}}_k = [y_m(t_1), \ldots, y_m(t_k)]'$, with $y_m(t_i) = x_m(t_i) + v(t_i)$.

From (7), we obtain that

$$\tilde{\boldsymbol{\sigma}}_k(t_{k+1}) = \mathbf{L}_{k\times 4m}(t_k)\mathbf{A}_{4m\times 4m}\mathbf{q}_{4m}(t_{k+1})$$

and

$$\tilde{\boldsymbol{\Sigma}}_{k\times k}(t_k) = \mathbf{L}_{k\times 4m}(t_k)\mathbf{A}_{4m\times 4m}\mathbf{L}'_{k\times 4m}(t_k) + \mathbf{R}_{k\times k}(t_k)$$

where $\mathbf{q}_{4m}(t_k) = [\tilde{\mathbf{c}}'_{2m}(t_k), \mathbf{0}'_{2m}]'$, with $\mathbf{0}_{2m}$ the $2m$-dimensional vector whose elements are zero, $\mathbf{L}_{k\times 4m}(t_k) = \left[\mathbf{D}'_{2m\times k}(t_k), \mathbf{F}'_{2m\times k}(t_k)\right]$ with $\mathbf{D}_{2m\times k}(t_k) = \left[\tilde{\mathbf{d}}_{2m}(t_1), \ldots, \tilde{\mathbf{d}}_{2m}(t_k)\right]$ and $\mathbf{F}_{2m\times k}(t_k) = \left[\tilde{\mathbf{f}}_{2m}(t_1), \ldots, \tilde{\mathbf{f}}_{2m}(t_k)\right]$, $\mathbf{R}_{k\times k}(t_k)$ is a diagonal matrix with i-th entry r_i, and

$$\mathbf{A}_{4m\times 4m} = \left[\begin{array}{c|c} \tilde{\boldsymbol{\Lambda}}_{2m\times 2m} & \mathbf{I}_{2m\times 2m} \\ \hline \mathbf{I}_{2m\times 2m} & \mathbf{0}_{2m\times 2m} \end{array}\right]$$

being $\mathbf{I}_{2m\times 2m}$ the $2m \times 2m$-dimensional identity matrix and $\mathbf{0}_{2m\times 2m}$ the $2m \times 2m$-dimensional matrix with zero elements.

Then, the solution of (8) is of the form

$$\tilde{\mathbf{h}}_k(t_{k+1}) = \left[\mathbf{L}_{k\times 4m}(t_k)\mathbf{A}_{4m\times 4m}\mathbf{L}'_{k\times 4m}(t_k) + \mathbf{R}_{k\times k}(t_k)\right]^{-1} \times \mathbf{L}_{k\times 4m}(t_k)\mathbf{A}_{4m\times 4m}\mathbf{q}_{4m}(t_{k+1}) \tag{9}$$

From (9) we can define the suboptimum one-stage predictor

$$\hat{z}_m(t_{k+1}/t_k) = \tilde{\mathbf{h}}'_k(t_{k+1})\mathbf{y}_k \tag{10}$$

At a first sight, in comparison with the optimum one-stage predictor (1), the proposed solution (10) does not show an improvement from the computational standpoint since both estimates require the computation of the product of two k-dimentional vectors. However, the suboptimum coefficients (9) lead to a reduction in the computational burden with respect to solving directly the Wiener-Hopf equation.

In the following result, a recursive algorithm similar to the Kalman filter is designed for the computation of the proposed suboptimum one-stage predictor (10).

Theorem 1

$$\hat{z}_m(t_{k+1}/t_k) = \mathbf{q}'_{4m}(t_{k+1})\mathbf{e}_{4m}(t_k) \tag{11}$$

where $\mathbf{e}_{4m}(t_k)$ *is recursively computed through the equation*

$$\mathbf{e}_{4m}(t_k) = \mathbf{e}_{4m}(t_{k-1}) + \mathbf{k}_{4m}(t_k)\left[y(t_k) - \mathbf{l}'_{4m}(t_k)\mathbf{e}_{4m}(t_{k-1})\right]$$

with the initialization $\mathbf{e}_{4m}(t_0) = \mathbf{0}_{4m}$*, and where* $\mathbf{l}'_{4m}(t_k) = \left[\tilde{\mathbf{d}}'_{2m}(t_k), \tilde{\mathbf{f}}'_{2m}(t_k)\right]$ *and the vector* $\mathbf{k}_{4m}(t_k)$ *is given by*

$$\mathbf{k}_{4m}(t_k) = \mathbf{P}_{4m\times 4m}(t_{k-1})\mathbf{l}_{4m}(t_k)\left[\mathbf{l}'_{4m}(t_k)\mathbf{P}_{4m\times 4m}(t_{k-1})\mathbf{l}_{4m}(t_k) + r_j\right]^{-1} \tag{12}$$

with

$$\mathbf{P}_{4m\times 4m}(t_k) = \mathbf{P}_{4m\times 4m}(t_{k-1}) - \mathbf{k}_{4m}(t_k)\mathbf{l}'_{4m}(t_k)\mathbf{P}_{4m\times 4m}(t_{k-1})$$

where $\mathbf{P}_{4m\times 4m}(t_0) = \mathbf{A}_{4m\times 4m}$.

Proof. From (9), we have that the suboptimum one-stage predictor (10) is given by

$$\begin{aligned}\hat{z}_m(t_{k+1}/t_k) = {} & \mathbf{q}'_{4m}(t_{k+1})\mathbf{A}_{4m\times 4m}\mathbf{L}'_{k\times 4m}(t_k) \\ & \times \left[\mathbf{L}_{k\times 4m}(t_k)\mathbf{A}_{4m\times 4m}\mathbf{L}'_{k\times 4m}(t_k) + \mathbf{R}_{k\times k}(t_k)\right]^{-1}\mathbf{y}_k\end{aligned}$$

Then, introducing the vector

$$\begin{aligned}\mathbf{e}_{4m}(t_k) = {} & \mathbf{A}_{4m\times 4m}\mathbf{L}'_{k\times 4m}(t_k) \\ & \times \left[\mathbf{L}_{k\times 4m}(t_k)\mathbf{A}_{4m\times 4m}\mathbf{L}'_{k\times 4m}(t_k) + \mathbf{R}_{k\times k}(t_k)\right]^{-1}\mathbf{y}_k\end{aligned} \tag{13}$$

the equation (11) for $\hat{z}_m(t_k)$ is obtained.

Next, applying the matrix inversion lemma [Anderson and Moore, 1979, p. 138] in (13), we have that

$$\mathbf{e}_{4m}(t_k) = \mathbf{P}_{4m\times 4m}(t_k)\mathbf{L}'_{k\times 4m}(t_k)\mathbf{R}^{-1}_{k\times k}(t_k)\mathbf{y}_k$$

where

$$\mathbf{P}_{4m\times 4m}(t_k) = \left[\mathbf{A}^{-1}_{4m\times 4m} + \mathbf{L}'_{k\times 4m}(t_k)\mathbf{R}^{-1}_{k\times k}(t_k)\mathbf{L}_{k\times 4m}(t_k)\right]^{-1} \tag{14}$$

Finally, taking into account the matrix inversion lemma in (14) and defining the vector $\mathbf{k}_{4m}(t_k)$ as in (12), the theorem holds.

Remark 1 *Note that, from* (5)*, a similar recursive algorithm can be designed for the optimum one-stage predictor. However, the amount of computation required with the resulting recursive formulas makes that this algorithm loss interest in practical applications.*

Remark 2 *From the PCA, the convergence of the proposed suboptimum predictor toward the optimum one is guaranteed. Then, the suboptimum one-stage predictor becomes a better approximation of the optimum one as the number* m *increases. On the other hand, a suitable* m *must be selected in order to reduce the computational burden. In fact, the efficiency of the proposed suboptimum estimate will be more relevant when the signal can be represented by a short series expansion. Some examples of such signals can be found in [Ghanem and Spanos, 1991].*

References

[Aguilera *et al.*, 1995]A.M. Aguilera, R. Gutiérrez, F.A. Ocaña, and M.J. Valderrama. Computational Approaches to Estimation in the Principal Component Analysis of a Stochastic Process. *Applied Stochastic Models and Data Analysis*, 11(4):279–299, 1995.

[Aguilera *et al.*, 1996]A.M. Aguilera, R. Gutiérrez, and M.J. Valderrama. Approximation of Estimators of the PCA of a Stochastic Process Using B-Splines. *Communications in Statistics (Simulation and Computation)*, 25(3):671–690, 1996.

[Anderson and Moore, 1979]B.D.O. Anderson and J.B. Moore. *Optimal Filtering.* Prentice-Hall, New Jersey, 1979.

[Fukunaga and Koontz, 1970]K. Fukunaga and L.G. Koontz. Representation of Random Processes Using the Finite Karhunen-Loève Expansion. *Informat. and Control*, 16:85–101, 1970.

[Gardner and Franks, 1971]W.A. Gardner and L.E. Franks. An Alternative Approach to Linear Least Squares Estimation of Continuous Random Processes. In *5th Ann. Princeton Conf. Information Sciences and Systems*, 1971.

[Gardner, 1975]W.A. Gardner. A Series Solution to Smoothing, Filtering, and Prediction Problems Involving Correlated Signal and Noise. *IEEE, Trans. Information Theory*, IT-21:698–699, 1975.

[Ghanem and Spanos, 1991]R.G. Ghanem and P.D. Spanos. *Stochastic Finite Elements: A Spectral Approach.* Springer-Verlag, New York, 1991.

[Kailath, 1970]T. Kailath. Likelihood Ratios for Gaussian Processes. *IEEE, Trans. Information Theory*, 16(3):276–288, 1970.

[Kalman and Bucy, 1961]R. E. Kalman and R. S. Bucy. New Results in Linear Filtering and Prediction Theory. *Trans. ASME, J. Basic Engineering, Ser. D*, 83:95–108, 1961. In: Epheremides, A. and Thomas, J.B. (Ed.) 1973. Random Processes. Multiplicity Theory and Canonical Decompositions.

[Poor, 1998]H.V. Poor. *An Introduction to Signal Detection and Estimation.* Springer-Verlag, New York, 2nd edition, 1998.

[Sugisaka, 1983]M. Sugisaka. The Design of On-line Least-Squares Estimators Given Covariance Specifications Via an Imbedding Method. *Applied Mathematics and Computation*, (13):55–85, 1983.