C. R. Acad. Sci. Paris, Ser. I ●●● (●●●●) ●●●–●●●

Contents lists available at ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Statistics

A novel signal extraction approach for filtering and forecasting noisy exponential series

Une nouvelle approche dans l'extraction de signal pour le filtrage et la prévision par des séries exponentielles avec bruit

Hossein Hassani [a], Mahdi Kalantari [b]

[a] *Research Institute for Energy Management and Planning, University of Tehran, Tehran, Iran*
[b] *Department of Statistics, Payame Noor University, 19395-4697, Tehran, Iran*

A R T I C L E I N F O

Article history:
Received 27 January 2018
Accepted 12 March 2018
Available online xxxx

Presented by Paul Deheuvels

A B S T R A C T

The coefficients of Linear Recurrent Relations (LRR) play a pivotal role in many forecasting techniques. Precise and closed form of the LRR coefficients enables one to achieve more accurate forecasts. On account to the fact that, in real-world situations, a time series data is contaminated with noise, extracting the noiseless series is of great importance. This paper seeks to obtain a closed form, with less noise level, of LRR coefficients for noisy exponential time series. Improving the filtering performance through employing noiseless eigenvectors of the covariance matrix is another novelty of this study. Our simulation results confirm that the proposed approach enhances filtering and forecasting results.

R É S U M É

Les coefficients des relations récurrentes linéaires (RRL) jouent un rôle central dans beaucoup de techniques de prévision. Une formule exacte et close des coefficients d'une RRL permet d'obtenir des prévisions plus précises. Prenant en compte le fait que, dans la réalité, une suite temporelle de données est contaminée par du bruit, il est très important de pouvoir en extraire la série sans bruit. Ce texte vise à obtenir une forme close, avec un niveau de bruit moindre, des coefficients d'une RRL, pour les suites en temps exponentiel avec bruit. Une autre nouveauté de notre approche est l'amélioration de l'efficacité du filtrage par l'utilisation de vecteurs propres sans bruit de la matrice de covariance. Les résultats des simulations confirment que l'approche proposée améliore le filtrage et les prévisions.

E-mail address: hassani.stat@gmail.com (H. Hassani).

https://doi.org/10.1016/j.crma.2018.03.006

1. Introduction

A time series $Y_N = \{y_1, y_2, \ldots, y_N\}$ of length N is generated by a *linear recurrence relation* (LRR) of order $d > 0$, if there exists the coefficients $\alpha_1, \ldots, \alpha_d$ such that:

$$y_{i+d} = \sum_{j=1}^{d} \alpha_j y_{i+d-j}, \quad 1 \le i \le N-d, \quad \alpha_d \ne 0, \quad d < N. \tag{1}$$

The coefficients $\alpha_1, \ldots, \alpha_d$ are called the coefficients of the LRR or α-*coefficients*. The class of time series governed by LRRs are rather wide and important for practical applications (see, for example, [1–3,5,7–18,22,23]). The Singular Spectrum Analysis (SSA) technique is one of the powerful and non-parametric techniques with capability of both forecasting and filtering where LRR is used for forecasting new data points [6,17].

Suppose L be an integer called *Window Length* such that $2 \le L \le N/2$. The starting point of SSA is to construct the trajectory matrix $\mathbf{X} = [X_1 : \cdots : X_K]$ via vectors $X_i = (y_i, \ldots, y_{i+L-1})^\intercal \in \mathbf{R}^L$, $i = 1, \ldots, K$, called *lagged vectors*, where $K = N - L + 1$. The trajectory matrix $\mathbf{X}$ is a *Hankel* matrix in the sense that all elements on anti-diagonals $i + j = const$. are equal. The eigenvalues of $\mathbf{X}\mathbf{X}^\intercal$ are denoted by $\lambda_1, \ldots, \lambda_L$ in decreasing order of magnitude ($\lambda_1 \ge \cdots \ge \lambda_L \ge 0$), and the eigenvectors of $\mathbf{X}\mathbf{X}^\intercal$ corresponding to these eigenvalues are denoted by $U_1, \ldots, U_L$. It is assumed that the eigenvectors have unit length, i.e. $\|U_i\| = 1$, where $\|\cdot\|$ is the Euclidean norm.

The eigenvectors of $\mathbf{X}\mathbf{X}^\intercal$ play a very pivotal role in the reconstruction stage of SSA. Let I be the chosen set of eigentriples attained at the grouping step of SSA and $U_i \in \mathbf{R}^L$, $i \in I$, be the corresponding eigenvectors. Denote by $\mathcal{L} \subset \mathbf{R}^L$ the linear space spanned by the vectors $U_i, i \in I$; i.e. $\mathcal{L} = \text{span}\{U_i, i \in I\}$. Note that the set $\{U_i, i \in I\}$ is an orthonormal basis in $\mathcal{L}$. To reconstruct the time series Y_N by set I, all lagged vectors X_i are first orthogonally projected onto $\mathcal{L}$ through $\widehat{\mathbf{X}} = \sum_{j \in I} U_j U_j^T \mathbf{X}$, where $\mathbf{X}$ is the trajectory matrix of series Y_N, and where the matrix $\widehat{\mathbf{X}}$ consists of column vectors $\widehat{X}_i$, $\widehat{X}_i = \sum_{j \in I} U_j U_j^T X_i$. Then the matrix $\widehat{\mathbf{X}}$ is diagonally averaged to get the reconstructed series $\widetilde{Y}_N = \{\widetilde{y}_1, \ldots, \widetilde{y}_N\}$.

The eigenvectors $U_i, i \in I$, are also employed in the forecasting methods of SSA. Let $\underline{U_i} \in \mathbf{R}^{L-1}$ be the vector consisting of the first $L-1$ components of the vector U_i, π_i be the last component of the vector U_i and $v^2 = \sum_{i \in I} \pi_i^2$. The last component z_L of any vector $Z = (z_1, \ldots, z_L)^\intercal \in \mathcal{L}$ is a linear combination of the first components $z_1, \ldots, z_{L-1}$, i.e. $z_L = \alpha_1 z_{L-1} + \cdots + \alpha_{L-1} z_1$, (see [6]), where the vector $A = (\alpha_{L-1}, \ldots, \alpha_1)^\intercal$ is obtained as follows:

$$A = \frac{1}{1 - v^2} \sum_{i \in I} \pi_i \underline{U_i}. \tag{2}$$

The α-coefficients $\{\alpha_j, j = 1, \ldots, L-1\}$ in (2), which are made by eigenvectors $U_i, i \in I$, play a fundamental role in SSA forecasting. For example, in *Recurrent* forecasting (R-forecasting) if the time series $Z_{N+h} = \{z_1, \ldots, z_{N+h}\}$ is defined by:

$$z_i = \begin{cases} \widetilde{y}_i & \text{for } i = 1, \ldots, N, \\ \sum_{j=1}^{L-1} \alpha_j z_{i-j} & \text{for } i = N+1, \ldots, N+h, \end{cases} \tag{3}$$

then the numbers $z_{N+1}, \ldots, z_{N+h}$ are the h step ahead recurrent forecasts. It is clear that R-forecasting is performed by the direct use of LRR in (1) and of α-coefficients in (2).

The same approach can be used for *Vector* forecasting (V-forecasting). Consider the matrix $\Pi = \underline{V}\,\underline{V}^\intercal + (1 - v^2) A A^\intercal$, where the matrix $\underline{V}$ consists of column vectors $\underline{U_i}, i \in I$. If the vectors W_i defined as:

$$W_i = \begin{cases} \widehat{X}_i & \text{for} \quad i = 1, \ldots, K, \\ \mathcal{P}_{\text{Vec}} W_{i-1} & \text{for} \quad i = K+1, \ldots, K+h+L-1, \end{cases} \tag{4}$$

where $\mathcal{P}_{\text{Vec}} W_{i-1} = \begin{pmatrix} \Pi \overline{W}_{i-1} \\ A^\intercal \overline{W}_{i-1} \end{pmatrix}$ and $\overline{W}_{i-1}$ is the vector consisting of the last $L-1$ components of the vector W_{i-1}, then by constructing the matrix $\mathbf{W} = [W_1 : \cdots : W_{K+h+L-1}]$ and making its diagonal averaging the series $\{z_1, \ldots, z_{N+h+L-1}\}$ is obtained. The numbers $z_{N+1}, \ldots, z_{N+h}$ are the h step ahead vector forecasts. It is clear that α-coefficients have also key role in V-forecasting through the matrix Π and the linear operator $\mathcal{P}_{\text{Vec}}$.

It can be assumed that Y_N is the sum of a noise free series (signal) and noise, i.e.:

$$y_t = s_t + n_t, \quad t = 1, \ldots, N, \tag{5}$$

where s_t and n_t represent the signal and noise components, respectively. Equation (5) can be expressed in the following matrix form:

$$\mathbf{X} = \mathbf{S} + \mathbf{N}, \tag{6}$$

Please cite this article in press as: H. Hassani, M. Kalantari, A novel signal extraction approach for filtering and forecasting noisy exponential series, C. R. Acad. Sci. Paris, Ser. I (2018), https://doi.org/10.1016/j.crma.2018.03.006