Chapter 3
SSA for Forecasting, Interpolation, Filtering and Estimation

3.1 SSA Forecasting Algorithms

3.1.1 Main Ideas and Notation

A reasonable forecast of a time series can be performed only if the series has a structure and there are tools to identify and use this structure. Also, we should assume that the structure of the time series is preserved for the future time period over which we are going to forecast (continue) the series. The last assumption cannot be validated using the data to be forecasted. Moreover, the structure of the series can rarely be identified uniquely. Therefore, the situation of different and even contradictory forecasts is not impossible. Hence it is important not only to understand and express the structure but also to assess its stability.

A forecast can be made only if a model is built. The model should be either derived from the data or at least checked against the data. In SSA forecasting, these models can be described through the linear recurrence relations (LRRs). The class of series governed by LRRs is rather wide and important for practical applications. This class contains the series that are linear combinations of products of exponential, polynomial and harmonic series.

Assume that $\mathbb{X}_N = \mathbb{X}_N^{(1)} + \mathbb{X}_N^{(2)}$, where the series $\mathbb{X}_N^{(1)}$ satisfies an LRR of relatively small order and we are interested in forecasting of $\mathbb{X}_N^{(1)}$. For example, $\mathbb{X}_N^{(1)}$ can be signal, trend or seasonality. The idea of recurrent forecasting is to estimate the underlying LRR and then to perform forecasting by applying the estimated LRR to the last points of the SSA approximation of the series $\mathbb{X}_N^{(1)}$. The main assumption allowing SSA forecasting is that for a certain window length L the series components $\mathbb{X}_N^{(1)}$ and $\mathbb{X}_N^{(2)}$ are approximately strongly separable. In this case, we can reconstruct the series $\mathbb{X}_N^{(1)}$ with the help of a selected set of the eigentriples and obtain approximations to both the series $\mathbb{X}_N^{(1)}$, its trajectory space and the true LRR.

N. Golyandina and A. Zhigljavsky, *Singular Spectrum Analysis for Time Series*, SpringerBriefs in Statistics,
https://doi.org/10.1007/978-3-662-62436-4_3

Let $\mathbb{X}_N = \mathbb{X}_N^{(1)} + \mathbb{X}_N^{(2)}$ and we intend to forecast $\mathbb{X}_N^{(1)}$. If $\mathbb{X}_N^{(1)}$ is a time series of finite rank $r < L$, then it generates an L-trajectory subspace of dimension r. This subspace reflects the structure of $\mathbb{X}_N^{(1)}$ and hence it can be taken as a base for forecasting.

Let us formally describe the forecasting algorithms in a chosen subspace. As we assume that estimates of this subspace are constructed by SSA, we shall refer to the algorithms as the algorithms of SSA forecasting.

Forecasting within a subspace means a continuation of the L-lagged vectors of the time series in such a way that they lie in or very close to the chosen subspace of R^L. We consider the following three forecasting algorithms: recurrent, vector and simultaneous.

Inputs in the forecasting algorithms:

(a) Time series $\mathbb{X}_N = (x_1, \ldots, x_N)$, $N > 2$.

(b) Window length L, $1 < L < N$.

(c) Linear space $\mathfrak{L}_r \subset \mathsf{R}^L$ of dimension $r < L$. We assume that $\mathbf{e}_L \notin \mathfrak{L}_r$, where $\mathbf{e}_L = (0, 0, \ldots, 0, 1)^{\mathrm{T}} \in \mathsf{R}^L$; in other terms, $\mathfrak{L}_r$ is not a 'vertical' space.

(d) Number M of points to forecast for.

Notation:

(a) $\mathbf{X} = [X_1 : \ldots : X_K]$ (with $K = N - L + 1$) is the trajectory matrix of $\mathbb{X}_N$.

(b) $P_1, \ldots, P_r$ is an orthonormal basis in $\mathfrak{L}_r$.

(c) $\widehat{\mathbf{X}} \stackrel{\mathrm{def}}{=} [\widehat{X}_1 : \ldots : \widehat{X}_K] = \sum_{i=1}^{r} P_i P_i^{\mathrm{T}} \mathbf{X}$. The vector $\widehat{X}_i$ is the orthogonal projection of X_i onto the space $\mathfrak{L}_r$.

(d) $\widetilde{\mathbf{X}} = \boldsymbol{\Pi}_{\mathcal{H}} \widehat{\mathbf{X}} = [\widetilde{X}_1 : \ldots : \widetilde{X}_K]$ is the result of hankelization of the matrix $\widehat{\mathbf{X}}$. The matrix $\widetilde{\mathbf{X}}$ is the trajectory matrix of some time series $\widetilde{\mathbb{X}}_N = (\widetilde{x}_1, \ldots, \widetilde{x}_N)$.

(e) For any vector $Y \in \mathsf{R}^L$, we denote by $\overline{Y} \in \mathsf{R}^{L-1}$ the vector consisting of the last $L - 1$ components of the vector Y and by $\underline{Y} \in \mathsf{R}^{L-1}$ the vector consisting of the first $L - 1$ components of Y.

(f) We set $\nu^2 = \pi_1^2 + \ldots + \pi_r^2$, where π_i is the last component of the vector P_i ($i = 1, \ldots, r$). As ν^2 is the squared cosine of the angle between the vector $\mathbf{e}_L$ and the linear space $\mathfrak{L}_r$, it is called the *verticality coefficient* of $\mathfrak{L}_r$. Since $\mathbf{e}_L \notin \mathfrak{L}_r$, $\nu^2 < 1$.

The following statement is fundamental.

Proposition 3.1 *In the notation above, the last component y_L of any vector $Y = (y_1, \ldots, y_L)^{\mathrm{T}} \in \mathfrak{L}_r$ is a linear combination of the first components $y_1, \ldots, y_{L-1}$:*

$$y_L = a_1 y_{L-1} + a_2 y_{L-2} + \ldots + a_{L-1} y_1,$$

where the vector $R = (a_{L-1}, \ldots, a_1)^{\mathrm{T}}$ can be expressed as

$$R = \frac{1}{1 - \nu^2} \sum_{i=1}^{r} \pi_i \underline{P_i} \tag{3.1}$$

and does not depend on the choice of the basis $P_1, \ldots, P_r$ in the linear space $\mathfrak{L}_r$.