Then the measure m_f is concentrated at the points $\pm\omega_k$, $\omega_k \in (0, 1/2)$, with the weights $(a_k^2 + b_k^2)/4$. The weight of the point 1/2 equals $a^2(1/2)$.

A series of the form (1.14) will be called *almost periodic* (see Section 6.4.2 for the precise definition). *Periodic* series correspond to a spectral measure m_f concentrated at the points $\pm j/T$ ($j = 1, \ldots, [T/2]$) for some integer T. In terms of the representation (1.14), this means that the number of terms in this representation is finite and all the frequencies ω_k are rational.

Almost periodic series that are not periodic are called *quasi-periodic*. For these series the spectral measure is discrete, but it is not concentrated on the nodes of any grid of the form $\pm j/T$. The *harmonic* $f_n = \cos 2\pi\omega n$ with an irrational ω provides an example of a quasi-periodic series.

Aperiodic (in other terminology — *chaotic*) series are characterized by a spectral measure that does not have atoms. In this case one usually assumes the existence of the *spectral density*: $m_f(d\omega) = p_f(\omega)d\omega$. Aperiodic series serve as models for *noise*; they are also considered in the theory of chaotic dynamical systems. If the spectral density exists and is constant, then the aperiodic series is called *white noise*.

Almost periodic and chaotic series have different asymptotic behaviour of their covariance functions: in the aperiodic case this function tends to zero, but the almost periodic series are (generally) characterized by almost periodic covariance functions.

As a rule, real-life stationary series have both components, periodic (or quasi-periodic) and noise (aperiodic) components. (The series 'White dwarf' – Section 1.3.2 – is a typical example of such series.)

Note that it is difficult, or even impossible when dealing with a finite series, to distinguish between a periodic series with a large period and a quasi-periodic series. Moreover, on finite time intervals aperiodic series are hardly distinguished from a sum of harmonics with wide spectrum and small amplitudes.

For a description of finite, but reasonably long, stationary series, it is convenient to use the language of the *Fourier expansion* of the initial series. This is the expansion

$$f_n = c_0 + \sum_{k=1}^{[N/2]} \Big(c_k \cos(2\pi n\, k/N) + s_k \sin(2\pi n\, k/N)\Big), \tag{1.15}$$

where N is the length of the series, $0 \le n < N$, and $s_{N/2} = 0$ for even N. The zero term c_0 is equal to the average of the series, so that if the series is centred, then $c_0 = 0$.

For a series of a finite length, the *periodogram* of the series is an analogue of the spectral measure. By definition (see Section 6.4.5) the periodogram $\Pi_f^N(\omega)$ of the series $F = (f_0, \ldots, f_{N-1})$ is

$$\Pi_f^N(\omega) = \frac{1}{N}\left|\sum_{n=0}^{N-1} e^{-i2\pi\omega n} f_n\right|^2, \quad \omega \in (-1/2, 1/2]. \tag{1.16}$$

Since the elements of the series F are real numbers, $\Pi_f^N(-\omega) = \Pi_f^N(\omega)$ for $|\omega| < 1/2$, and therefore we can consider only the interval $[0, 1/2]$ for ω. If the series F is represented in the form (1.15), then it is not difficult to show that

$$\Pi_f^N(k/N) = \frac{N}{2}\begin{cases} 2c_0^2 & \text{for } k = 0, \\ c_k^2 + s_k^2 & \text{for } 0 < k < N/2, \\ 2c_{N/2}^2 & \text{for } k = N/2. \end{cases} \tag{1.17}$$

The last case is, of course, possible only when N is odd.

Let us consider the Fourier expansions (1.15) of two series $F^{(1)}$ and $F^{(2)}$ of length N and denote the corresponding coefficients by $c_k^{(j)}$ and $s_k^{(j)}$, $j = 1, 2$. Using the notation

$$d_k = \begin{cases} c_k^{(1)}c_k^{(2)} + s_k^{(1)}s_k^{(2)} & \text{for } k \neq 0 \text{ and } N/2, \\ 2c_k^{(1)}c_k^{(2)} & \text{for } k = 0 \text{ or } N/2, \end{cases} \tag{1.18}$$

we can easily see that the inner product of two series is

$$\begin{aligned} \left(F^{(1)}, F^{(2)}\right) &\stackrel{\text{def}}{=} \sum_{k=0}^{N-1} f_n^{(1)} f_n^{(2)} \\ &= \frac{N}{2}\Big(2d_0 + \sum_{0<k<N/2} d_k + 2d_{N/2}\Big), \end{aligned} \tag{1.19}$$

where $d_{N/2} = 0$ for odd N.

This immediately yields that the norm $||F|| = \sqrt{(F, F)}$ of the series (1.15) is expressed through its periodogram as follows:

$$||F||^2 = \sum_{k=0}^{[N/2]} \Pi_f^N(k/N). \tag{1.20}$$

The equality (1.20) implies that the value (1.17) of the periodogram at the point k/N describes the influence of the harmonic components with frequency $\omega = k/N$ into the sum (1.15). Moreover, (1.20) explains the normalizing coefficient N^{-1} in the definition (1.16) of the periodogram.

Some other normalizations of the periodograms are known in literature and could be useful as well. In particular, using below the periodogram analysis of the time series for the purposes of SSA, we shall plot the values of $\Pi_f^N(k/N)/N$ (for fixed k this is called *power* of the frequency k/N), but we shall keep the name 'periodogram' for the corresponding line-plots.

The collection of frequencies $\omega_k = k/N$ with positive powers is called the *support of the periodogram.* If the support of a certain periodogram belongs to some interval $[a, b]$, then this interval is called the *frequency range of the series.*

Asymptotically, for the stationary series, the periodograms approximate the spectral measures (see Theorem 6.4 of Section 6.4.5).

Thus, a standard model of a stationary series is a sum of a periodic (or quasi-periodic) series and an aperiodic series possessing a spectral density.

If the length N of a stationary series is large enough, then the frequencies j/N, which are close to the most powerful frequencies of the almost periodic component of the series, have large values in the periodogram of the series.

For short series the grid $\{j/N,\ j=0,\dots,[N/2]\}$ is a poor approximation to the whole range of frequencies $[0,1/2]$, and the periodogram may badly reflect the periodic structure of the series components.

(b) Amplitude-modulated periodicities

The nature of the definition of stationarity is asymptotic. This asymptotic nature has both advantages (for example, the rigorous mathematical definition allows illustration of all the concepts by model examples) as well as disadvantages (the main one is that it is not possible to check the stationarity of the series using only a finite-time interval of it).

At the same time, there are numerous deviations from stationarity. We consider only two classes of the nonstationary time series which we describe at a qualitative level. Specifically, we consider amplitude-modulated periodic series and series with trends. The choice of these two classes is related to their practical significance and importance for the SSA.

The trends are dealt with in the next subsection. Here we discuss the *amplitude-modulated* periodic signals, that is, series of the form $f_n = A(n)g_n$, where g_n is a periodic sequence and $A(n) \geq 0$. Usually it is assumed that on the given time interval ($0 \leq n \leq N-1$) the function $A(n)$ varies much more slowly than the low-frequency harmonic component of the series g_n.

Series of this kind are typical in economics where the period of the harmonics g_n is related to seasonality, but the amplitude modulation is determined by long-term tendencies.

An explanation of the same sort is suitable for the example 'War' of Section 1.3.7, where the seasonal component of the combat deaths (Fig. 1.14, bottom line) seems to be modulated by the intensity of the military activities.

Let us discuss the periodogram analysis of the amplitude-modulated periodic signals, for the moment restricting ourselves to the amplitude-modulated harmonic

$$f_n = A(n)\cos(2\pi\omega + \theta), \quad n = 0,\dots,N-1. \tag{1.21}$$

As a rule, the periodogram of the series (1.21) is supported on a short frequency interval containing ω. This is not surprising since, for example, for large $\omega_1 \approx \omega_2$ the sum

$$\cos(2\pi\omega_1 n) + \cos(2\pi\omega_2 n) = 2\cos\Big(\pi(\omega_1-\omega_2)n\Big)\cos\Big(\pi(\omega_1+\omega_2)n\Big)$$

is a product of a slowly varying sequence

$$A(n) = 2\cos\left(\pi(\omega_1 - \omega_2)n\right)$$

and the harmonic with the high frequency $(\omega_1 + \omega_2)/2$.

Note that for $n \leq 1/2(\omega_1 - \omega_2)$ the sequence $A(n)$ is positive and its oscillatory nature cannot be seen for small n.

Fig. 1.16 depicts the periodogram of the main seasonal (annual plus quarterly) component of the series 'War' (Section 1.3.7). We see that the periodogram is supported around two main seasonal frequencies, but is not precisely concentrated at these two points. For the 'War' series, this is caused by the amplitude modulation.

However, the above discussion implies that in the general case the appearance of exactly the same modulation can be caused by two different reasons: either it can be the 'true' modulation, which can be explained by taking into account the nature of the signal, or the modulation is spurious, with its origin in the closeness of the frequencies of the harmonic components of the original series.

The other possible reason of spreading around the main frequencies is the discreteness of the periodogram grid $\{k/N\}$: if a frequency ω of a harmonic does not belong to the grid, then it spreads over it.

Note that since the length of the 'War' series is proportional to 12, the frequencies $\omega = 1/12$ and $\omega = 1/3$, which correspond to annual and quarterly periodicities, fall exactly on the periodogram grid $\{k/36,\ k = 1, \ldots, 18\}$.

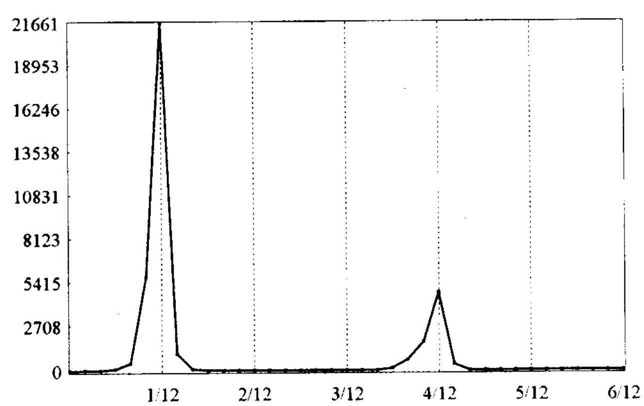

Figure 1.16 *War: periodogram of the main seasonality component.*

Evidently, not only periodic series can be modulated by the amplitude; the same can hold for quasi-periodic and chaotic sequences. However, identification of these cases by means of the periodogram analysis is more difficult.

(c) Trends

It seems that there is no commonly accepted definition of the concept 'trend'. Certainly, the main tendency of the series can be postulated with the help of a parametric model, and subsequent estimation of the parameters would allow us to talk about, say, linear, exponential, or logistic trends; see, for instance, Anderson (1994, Chapter 3.8). Very often the problem of trend approximation is stated directly, as a pure approximation problem, without any worry concerning the tendencies. The most popular approximation is polynomial; see, for example, Otnes and Enochson (1978, Chapter 3.8) or Anderson (1994, Chapter 3.2.1).

For us, this meaning of the notion 'trend' is not suitable just because Basic SSA is a model-free, and therefore nonparametric method.

Under the assumption that the series F is a realization of a certain random discrete-time process, $\xi(n)$, trend is often defined as $\mathbf{E}\xi(n)$, the expectation of the random process (see, for instance, Diggle, 1990, Section 1.4). We cannot use this definition since we are working with only one trajectory and do not have an ensemble of trajectories for averaging.

In principle, the trend of a time series can be described as a function that reflects slow, stable, and systematic variation over a long period of time; see Kendall and Stuart (1976, Section 45.12). The notion of trend in this case is related to the length of the series — from the practical point of view this length is exactly the 'long period of time'.

Moreover, we have already collected oscillatory components of the series into a separate class of (centred) stationary series and therefore the term 'cyclical trend' (see Anderson, 1994, Chapter 4) does not suit us. In general, an appropriate definition of trend for SSA defines the trend as an additive component of the series which is (i) not stationary, and (ii) 'slowly varies' during the whole period of time that the series is being observed (compare Brillinger, 1975, Chapter 2.12).

At this point, let us mention some consequences of this understanding of the notion 'trend'. The most important is the nonuniqueness of the solution to the problem 'trend identification' or 'trend extraction' in its nonparametric setup. This nonuniqueness has already been illustrated by the example 'Production'; see Section 1.3, where Figs. 1.1-1.2 depict two forms of trend: the trend that describes the general tendency of the series (Fig. 1.1) and the detailed trend (Fig. 1.2).

Furthermore, for a finite time series, a harmonic component with a low frequency is practically indistinguishable from a trend (it can even be monotone on a finite time interval). In this case, auxiliary subject-related information about the series can be decisive for the problem of distinguishing trend from the periodicity.

For instance, even though the reconstructed trend in the example 'War' (see Section 1.3.7 and Fig. 1.14) looks like a periodicity observed over a time interval that is less than half of the period, it is clear that there is no question of periodicity in this case.

In the language of frequencies, trend generates large powers in the low-frequency range of the periodogram.

Finally, we have assumed that any stationary series is centred. Therefore, the average of all the terms f_n of any series F is always added to its trend. On the periodogram, a nonzero constant component of the series corresponds to an atom at zero.

Therefore, a general descriptive model of the series that we consider in the present monograph is the additive model where the components of the series are trends, oscillations, and noise components. In addition, the oscillatory components are subdivided into periodic and quasi-periodic, while the noise components are, as a rule aperiodic series. Both stationarity and amplitude modulation of the oscillatory and noise components are allowed. The sum of all the additive components, except for the noise, will be called the *signal*.

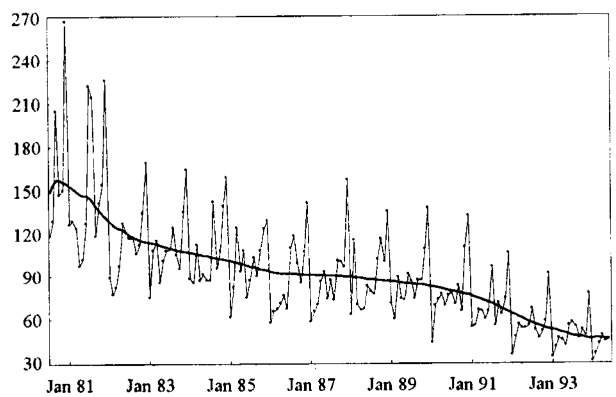

Figure 1.17 *Rosé wine: initial time series and the trend.*

Example 1.1 *Additive components of time series*
Let us consider the 'Rosé wine' series (monthly rosé wine sales, Australia, from July 1980 to June 1994, thousands of litres). Fig. 1.17 depicts the series itself (the thin line) and Fig. 1.18 presents its periodogram.

Fig. 1.17 shows that the series 'Rosé wine' has a decreasing trend and an annual seasonality of a complex form. Fig. 1.18 shows the periodogram of the series; it seems reasonable that the trend is related to large values at the low-frequency range, and the annual periodicity is related to the peaks at frequencies $1/12$, $1/6$, $1/4$, $1/3$, $1/2.4$, and $1/2$. The nonregularity of the powers for these frequencies indicates a complex form of the annual periodicity.

Fig. 1.19 depicts two additive components of the 'Rosé wine' series: the seasonal component (top graph), which is described by the eigentriples 2-11, 13 and the residual series ($L = 84$). The trend component (thick line in Fig. 1.17) is reconstructed from the eigentriples 1, 12, and 14.

Periodogram analysis demonstrates that the expansion of the series into three parts is indeed related to the separation of the spectral range into three regions: low

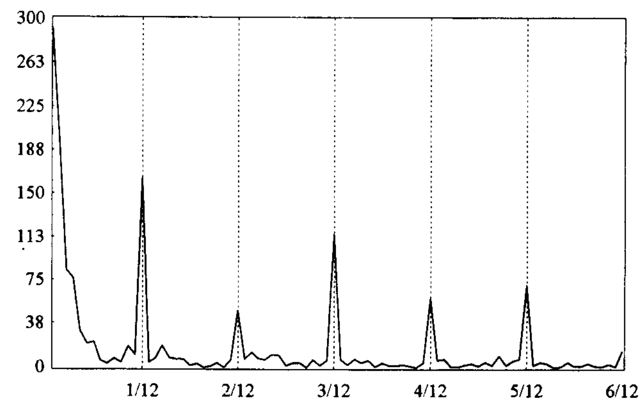

Figure 1.18 *Rosé wine: periodogram for the series.*

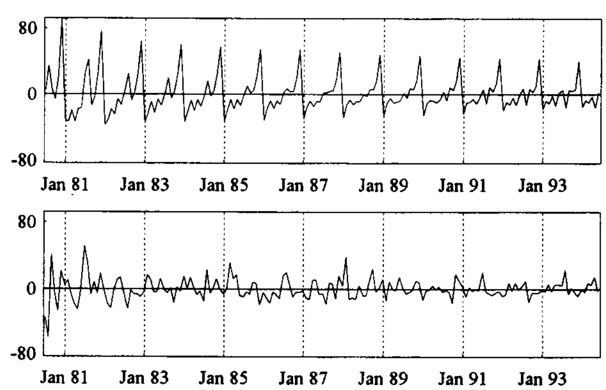

Figure 1.19 *Rosé wine: two components of the series.*

frequencies correspond to the trend (the thick line in Fig. 1.20), the frequencies describing the seasonalities correspond to the periodic component (Fig. 1.20, the thin line), and the residual series (which can be regarded as noise) has all the other frequencies (Fig. 1.21).

The periodograms of the whole series (see Fig. 1.18), its trend and the seasonal component (see Fig. 1.20) are presented in the same scale.

1.4.2 Basic SSA: Classification of the main tasks

Classification of the main tasks, which Basic SSA can be used for, is naturally related to the above classification of the time series and their components. It is, of

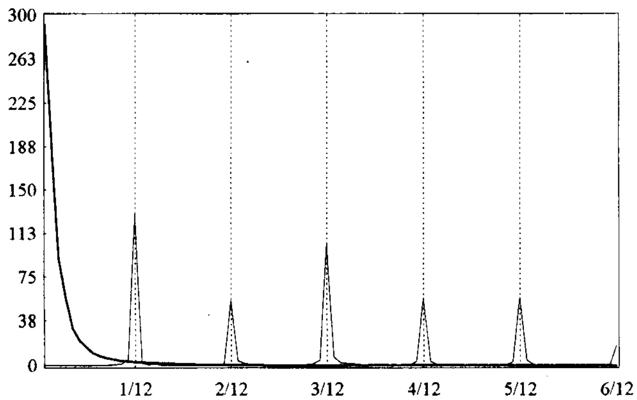

Figure 1.20 *Rosé wine: periodograms of the trend and the seasonal component.*

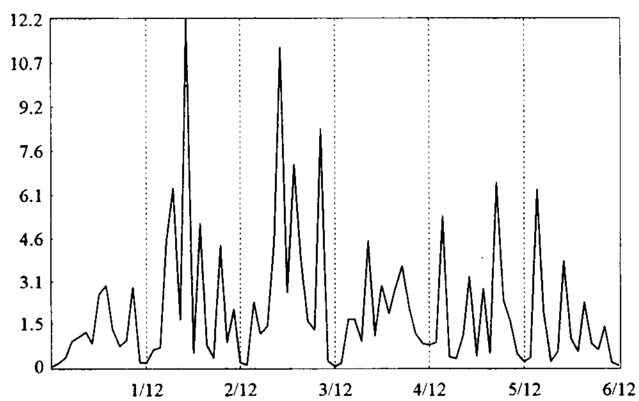

Figure 1.21 *Rosé wine: periodogram of the residuals.*

course, neither rigid nor exact, but it helps to understand features of the method and the main principles for the choice of its parameters in analyzing specific data sets; see Section 1.6.

1. *Trend extraction and smoothing*
   These two problems are in many ways similar and often cannot be distinguished in practice. None of these problems has an exact meaning, unless a parametric model is assumed. Therefore, a large number of model-free methods can be applied to solve each of them. Nevertheless, it is convenient to distinguish trend extraction and smoothing, at least on a qualitative level.

- *Trend extraction*

  This problem occurs when we want to obtain more or less refined non-oscillatory tendency of the series. The trend being extracted, we can investigate its behaviour, approximate it in a parametric form, and consider its continuation for forecasting purposes.

  Note that here we are not interested in whether the residual from the trend extraction has 'structure' (for example, it can contain a certain seasonality) or is a pure noise series.

  Results of trend extraction with the help of Basic SSA can be demonstrated by the examples 'Production' (Section 1.3.1, Figs. 1.1 and 1.2), 'Unemployment' (Section 1.3.6, Fig. 1.12) and 'War' (Section 1.3.7, Fig. 1.14).

  In the language of periodograms, trend extraction means extraction of the low-frequency part of the series that could not be regarded as an oscillatory one.

- *Smoothing*

  Smoothing a series means representing the series as a sum of two series where the first one is a 'smooth approximation' of it. Note that here we do not assume anything like existence of the trend and do not pay attention to the structure of the residuals: for example, the residual series may contain a strong periodicity of small period. In the language of frequencies, to smooth a series we have to remove all its high-frequency components.

  Methods that use weighted moving averages (see Anderson, 1994, Chapter 3.3) or weighted averages depending on time intervals, including the local polynomial approximation (see Kendall and Stuart, 1976, Chapter 36), perfectly correspond to the meaning of the term 'smoothing'. The same is true for the median smoothing; see Tukey (1977, Chapter 7).

  If a series is considered as a sum of a trend and a noise, a smoothing procedure would probably lead to a trend extraction.

  The example 'Snake' (Section 1.3.2, Fig. 1.3) shows the smoothing capabilities of Basic SSA. There is no distinct border between the trend extraction and smoothing, and the example 'Unemployment' (Section 1.3.6, Fig. 1.12) can be considered both for the refined trend extraction and for the result of a certain smoothing.

2. *Extraction of oscillatory components*

   The general problem here is identification and separation of the oscillatory components of the series that do not constitute parts of the trend. In the parametric form (under the assumptions of zero trend, finite number of harmonics, and additive stochastic white noise), this problem is extensively studied in the classical spectral analysis theory (see, for example, Anderson, 1994, Chapter 4).

   The statement of the problem in Basic SSA is specified mostly by the model-free nature of the method. One of the specifics is that the result of Basic SSA extraction of a single harmonic component of a series is not, as a rule, a purely harmonic sequence. This is a general feature of the method, it was thoroughly discussed in the Introduction. From the formal point of view, this means that in practice we deal with an approximate separability rather than with the exact one (see Section 1.5).

Also, application of Basic SSA does not require rigid assumptions about the number of harmonics and their frequencies. For instance, the example 'Births' (Section 1.3.4) illustrates simultaneous extraction of two (approximately) periodic components in daily data (the annual and weekly periodicities).

Certainly, auxiliary information about the initial series always makes the situation clearer and helps in choosing the parameters of the method. For example, the assumption that there might be an annual periodicity in monthly data suggests that the the analyst must pay attention to the frequencies $j/12$ $(j = 1, \dots, 6)$. The presence of sharp peaks on the periodogram of the initial series leads to the assumption that the series contains periodic components with these frequencies.

Finally, we allow the possibility of amplitude modulation for the oscillatory components of the series. In examples 'War' (Section 1.3.7), 'Drunkenness' (Section 1.3.5) and 'Unemployment' (Section 1.3.6) the capabilities of Basic SSA for their extraction have been demonstrated.

The most general problem is that of finding the whole structure of the series, that is splitting it into several 'simple' and 'interpretable' components, and the noise component.

3. *Obtaining the refined structure of a series*

   According to our basic assumption, any series that we consider can be represented as the sum of a signal, which itself consists of a trend and oscillations, and noise.

   If the components of the signal are expressed in a parametric form (see, for instance, Ledemann and Lloyd, 1984, Chapter 18.2) for a parametric model of seasonal effects), then the main problem of the decomposition of the series into its components can be formalized. Of course, for the model-free techniques such as Basic SSA, this is not so.

   The previously discussed problems are similar in the sense that we want to find a particular component of a series without paying much attention to the residuals. Our task is now to obtain the whole structure of the signal, that is to extract its trend (if any), to find its seasonal components and other periodicities, and so on. The residuals should be identified as the noise component.

   Therefore, we must take care of both the signal and noise. The decomposition of the signal into components depends, among the other things, on the interpretability of these components (see the 'War' example of Section 1.3.7).

   As for the residual series, we have to be convinced that it does not contain parts of the signal. If the noise can be assumed stochastic, then various statistical procedures may be applied to test the randomness of the residuals. Due to its simplicity, the most commonly used model of the noise is the model of stochastic white noise. From the practical point of view, it is usually enough to be sure that the residual series has no evident structure.

These considerations work for the formally simpler problem of *noise reduction,* which differs from that discussed above in that here the series is to be split into two components only, the signal and the noise, and a detailed study of the signal is not required.

In practice, this setup is very close to the setup of the problem of 'smoothing', especially when the concept 'noise' is understood in a broad sense.

For a series with a large signal-to-noise ratio, the signal generates sharp peaks at the periodogram of the series. It is sometimes important to remember that if the frequency range of the noise is wide (as for white noise), then the powers of the frequencies, relating to these peaks, include the powers of the same frequencies of the noise. The exception is only when the frequency ranges of the signal and noise are different.

1.5 Separability

As mentioned above, the main purpose of SSA is a decomposition of the original series into a sum of series, so that each component in this sum can be identified as either a trend, periodic or quasi-periodic component (perhaps, amplitude-modulated), or noise.

The notion of separability of series plays a fundamental role in the formalization of this problem (see Sections 1.2.3 and 1.2.4). Roughly speaking, SSA decomposition of the series F can be successful only if the resulting additive components of the series are (approximately) separable from each other.

This raises two problems that require discussion:

1. Assume that we have obtained an SSA decomposition of the series F. How do we check the quality of this decomposition? (Note that the notion of separability and therefore the present question are meaningful when the window length L is fixed.)

2. How, using only the original series, can we predict (at least partially and approximately) the results of the SSA decomposition of this series into components and the quality of this decomposition?

Of course, the second question is related to the problem of the choice of the SSA parameters (window length and the grouping manner). We shall therefore delay the corresponding discussion until the next section. Here we consider the concept of separability itself, both from the theoretical and the practical viewpoints.

1.5.1 Weak and strong separability

Let us fix the window length L, consider a certain SVD of the L-trajectory matrix $\mathbf{X}$ of the initial series F of length N, and assume that the series F is a sum of two series $F^{(1)}$ and $F^{(2)}$, that is, $F = F^{(1)} + F^{(2)}$.

In this case, separability of the series $F^{(1)}$ and $F^{(2)}$ means (see Section 1.2.3) that we can split the matrix terms of the SVD of the trajectory matrix $\mathbf{X}$ into two different groups, so that the sums of terms within the groups give the trajectory matrices $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ of the series $F^{(1)}$ and $F^{(2)}$, respectively.

The separability immediately implies (see Section 6.1) that each row of the trajectory matrix $\mathbf{X}^{(1)}$ of the first series is orthogonal to each row of the trajectory matrix $\mathbf{X}^{(2)}$ of the second series, and the same holds for the columns.

Since rows and columns of trajectory matrices are subseries of the corresponding series, the orthogonality condition for the rows (and columns) of the trajectory matrices $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ is just the condition of orthogonality of any subseries of length L (and $K = N - L + 1$) of the series $F^{(1)}$ to any subseries of the same length of the series $F^{(2)}$ (the subseries of the time series must be considered here as vectors).

If this orthogonality holds, then we shall say that the series $F^{(1)}$ and $F^{(2)}$ are *weakly separable*. A finer (and more desirable in practice) notion of separability is the notion of strong separability which, in addition to the orthogonality of the subseries of the two series, puts constraints on the singular values of the matrices $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$. This notion is discussed in Section 1.5.4. Note that if all the singular values of the trajectory matrix $\mathbf{X}$ are different, then the conditions for weak separability and strong separability coincide. Below, for brevity, we shall use the term 'separability' for 'weak separability'.

The condition of (weak) separability can be stated in terms of orthogonality of subspaces as follows: the series $F^{(1)}$ and $F^{(2)}$ are separable if and only if the subspace $\mathfrak{L}^{(L,1)}$ spanned by the columns of the trajectory matrix $\mathbf{X}^{(1)}$, is orthogonal to the subspace $\mathfrak{L}^{(L,2)}$ spanned by the columns of the trajectory matrix $\mathbf{X}^{(2)}$, and similar orthogonality must hold for the subspaces $\mathfrak{L}^{(K,1)}$ and $\mathfrak{L}^{(K,2)}$ spanned by the rows of the trajectory matrices.

Example 1.2 *Weak separability*

Let us illustrate the notion of separability in the language of geometry. Consider the series $F = F^{(1)} + F^{(2)}$ with elements $f_n = f_n^{(1)} + f_n^{(2)}$ where $f_n^{(1)} = a^n$, $f_n^{(2)} = (-1/a)^n$, $a = 1.05$ and $0 \le n \le 27$. Let $L = 2$, then the series $F^{(1)}$ and $F^{(2)}$ are (weakly) separable.

The two-dimensional phase diagram of the series F, that is the plot of the vectors $X_{n+1} = (f_n, f_{n+1})^{\mathrm{T}}$, is shown in Fig. 1.22 in addition to both principal directions of the SVD of the trajectory matrix of F. Since the principal directions are determined by the eigenvectors, they are proportional to the vectors $(1, 1.05)^{\mathrm{T}}$ and $(1.05, -1)^{\mathrm{T}}$. The projection of the vectors X_n on the first of these directions fully determines the series $F^{(1)}$ and 'annihilates' the series $F^{(2)}$, while the projection of these vectors on the second principal direction has the opposite effect: the series $F^{(2)}$ is left untouched and $F^{(1)}$ disappears.

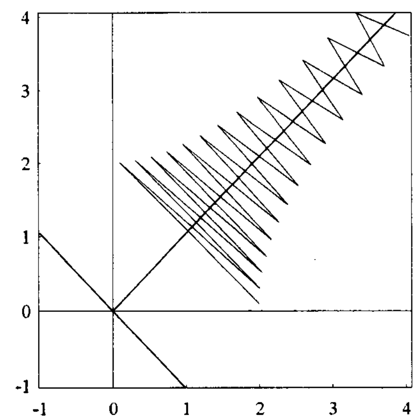

Figure 1.22 *Separability: two-dimensional phase diagram of the series.*

Since a similar phenomenon holds also for the K-dimensional phase diagram of the series ($K = 27$), Fig. 1.22 provides us with a simple geometrical interpretation of separability.

Let us give one more separability condition, which is only a necessary condition (it is not sufficient). This condition is very clear and easy to check.

Let $L^* = \min(L, K)$ and $K^* = \max(L, K)$. Introduce the weights

$$w_i = \begin{cases} i+1 & \text{for } 0 \le i \le L^* - 1, \\ L^* & \text{for } L^* \le i < K^*, \\ N-i & \text{for } K^* \le i \le N-1. \end{cases} \tag{1.22}$$

Define the inner product of series $F^{(1)}$ and $F^{(2)}$ of length N as

$$\left(F^{(1)}, F^{(2)}\right)_w \stackrel{\text{def}}{=} \sum_{i=0}^{N-1} w_i f_i^{(1)} f_i^{(2)} \tag{1.23}$$

and call the series $F^{(1)}$ and $F^{(2)}$ w-*orthogonal* if

$$\left(F^{(1)}, F^{(2)}\right)_w = 0.$$

It can be shown (see Section 6.2) that separability implies w-orthogonality. Therefore, if the series F is split into a sum of separable series $F^{(1)}, \ldots, F^{(m)}$, then this sum can be interpreted as an expansion of the series F with respect to a certain w-orthonormal basis, generated by the original series itself. Expansions of this kind are typical in linear algebra and analysis.

The window length L enters the definition of w-orthogonality; see (1.22). The weights in the inner product (1.23) have the form of a trapezium. If L is small relative to N, then almost all the weights are equal, but for $L \approx N/2$ the influence of the central terms in the series is much higher than of those close to the end-points in the time interval.

1.5.2 *Approximate and asymptotic separability*

Exact separability does not happen for real-life series and in practice we can talk only about approximate separability. Let us discuss the characteristics that reflect the degree of separability, leaving for the moment questions relating to the singular values of the trajectory matrices.

In the case of exact separability, the orthogonality of rows and columns of the trajectory matrices $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ means that all pairwise inner products of their rows and columns are zero. In statistical language, this means that the noncentral covariances (and therefore, noncentral correlations — the cosines of the angles between the corresponding vectors) are all zero. (Below, for brevity, when talking about covariances and correlations, we shall drop the word 'noncentral'.)

This implies that we can consider as a characteristic of separability of two series $F^{(1)}$ and $F^{(2)}$ the *maximum correlation coefficient* $\rho^{(L,K)}$, that is the maximum of the absolute value of the correlations between the rows and between the columns of the trajectory matrices of these two series (as usual, $K = N - L + 1$).

We shall say that two series $F^{(1)}$ and $F^{(2)}$ are *approximately separable* if all the correlations between the rows and the columns of the trajectory matrices $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ are close to zero.

Let us consider other characteristics of the quality of separability. The following quantity (called the *weighted correlation* or w-*correlation*) is a natural measure of deviation of two series $F^{(1)}$ and $F^{(2)}$ from w-orthogonality:

$$\rho_{12}^{(w)} = \frac{\left(F^{(1)}, F^{(2)}\right)_w}{\|F^{(1)}\|_w \|F^{(2)}\|_w}, \tag{1.24}$$

where $\|F^{(i)}\|_w = \sqrt{\left(F^{(i)}, F^{(i)}\right)_w}$, $i = 1, 2$.

If the absolute value of the w-correlation is small, then the two series are almost w-orthogonal, but, if it is large, then the series are far from being w-orthogonal and are therefore badly separable.

For long (formally, for infinitely long) series it is convenient to introduce the notion of asymptotic separability. Consider two infinite series $F^{(1)}$ and $F^{(2)}$ and denote by $F_N^{(1)}$ and $F_N^{(2)}$ the finite series consisting of the first N elements of the series $F^{(1)}$ and $F^{(2)}$. Assume also that the window length L is a function of the series length N.

We shall say that the series $F^{(1)}$ and $F^{(2)}$ are *asymptotically separable* if the maximum $\rho^{(L,K)}$ of the absolute values of the correlation coefficients between the rows/columns of the trajectory matrices of the series $F_N^{(1)}$ and $F_N^{(2)}$ tends to zero, as $N \to \infty$. The standard behaviour of the window length $L = L(N)$ in the definition of the asymptotic separability is such that $L, K \to \infty$.

From the practical viewpoint, the effect of the asymptotic separability becomes apparent in the analysis of long series and means that two asymptotically separable series are approximately separable for large N.

Section 6.1 contains several analytical examples of both exact and asymptotic separability. These examples show that the class of asymptotically separable series is much wider than the class of series that are exactly separable, and the conditions on the choice of the window length L are much weaker in the case of asymptotic separability.

For instance, exact separability of two harmonics with different periods can be achieved when both periods are divisors of both the window length L and $K = N - L + 1$. This requires, in particular, that the quotient of the periods is rational.

Another example of exact separability is provided by the series $f_n^{(1)} = \exp(\alpha n)$ and $f_n^{(2)} = \exp(-\alpha)\cos(2\pi n/T)$ with an integer T. (We thus deal with the exponential trend and an amplitude-modulated harmonic.) Here separability holds if the window length L and K are proportional to T.

Conditions for asymptotic separability are much weaker. In particular, two harmonics with arbitrary different frequencies are asymptotically separable as soon as L and K tend to infinity. Moreover, under the same conditions the periodic components are asymptotically separable from the trends of a general form (for example, from exponentials and polynomials).

1.5.3 Separability and Fourier expansions

Since separability of series is described in terms of orthogonality of their subseries, and the inner product of series can be expressed in terms of the coefficients in the Fourier expansion (1.19), new separability characteristics related to these expansions can be introduced.

In the expansion (1.19), the terms d_k defined in (1.18) have the meaning of the weights of the frequencies k/N in the inner product (1.19). If all the d_k are zero, then the series are orthogonal and this can be interpreted in terms of the Fourier expansion: each frequency k/N makes a zero input into the inner product.

On the other hand, if $\Pi_{f_1}^N(k/N)\,\Pi_{f_2}^N(k/N) = 0$ for all k, then, since

$$\frac{N}{2}\,|d_k| \le \sqrt{\Pi_{f_1}^N(k/N)\,\Pi_{f_2}^N(k/N)}, \tag{1.25}$$

all the d_k are zero and the orthogonality of the series has the following explanation in terms of the periodograms: in this case the supports of the periodograms of the series $F^{(1)}$ and $F^{(2)}$ do not intersect.

Thus, we can formulate the following sufficient separability condition in terms of periodograms: if for each subseries of length L (and K as well) of the series $F^{(1)}$ the frequency range of its periodogram is disjoint from the frequency range of the periodogram of each subseries of the same length of the series $F^{(2)}$, then the two series are exactly separable.

In this language, the above (sufficient) conditions for separability of two finite harmonic series with different integer periods T_1 and T_2 become obvious: if

$$L = k_1 T_1 = k_2 T_2, \qquad K = m_1 T_1 = m_2 T_2$$

with integer k_1, k_2, m_1, m_2, then for any subseries of length L of the first series its periodogram must contain just one frequency $\omega_1^{(L)} = 1/k_1$, and at the same time the corresponding frequency for the second series must be equal $\omega_2^{(L)} = 1/k_2$. For the subseries of length K, the analogous condition must hold for the frequencies $\omega_1^{(K)} = 1/m_1$ and $\omega_2^{(K)} = 1/m_2$.

For stationary series, the analogous conditions for asymptotic separability are: if the supports of the spectral measures of stationary series are disjoint, then these series are asymptotically separable as $L \to \infty$ and $K \to \infty$ (see Section 6.4.4).

The simplest example of this situation is provided by the sum of two harmonics with different frequencies.

When we deal with an approximate orthogonality of the series rather than with the exact one, we can use the characteristics describing the degree of disjointness of the supports of the periodograms, such that their smallness guarantees the smallness of the correlation coefficient between the series. Indeed, the formulae (1.18), (1.20) and (1.25) yield

$$\begin{aligned} \left|\left(F^{(1)}, F^{(2)}\right)\right| &\le \Phi^{(N)}(F^{(1)}, F^{(2)}) \\ &\stackrel{\text{def}}{=} \sum_{k=0}^{[N/2]} \sqrt{\Pi_{f_1}^N(k/N)\, \Pi_{f_2}^N(k/N)} \le \|F^{(1)}\|\, \|F^{(2)}\|. \end{aligned}$$

Therefore,

$$\rho_{12}^{(\Pi)} \stackrel{\text{def}}{=} \frac{\Phi^{(N)}(F^{(1)}, F^{(2)})}{\sqrt{\sum\limits_{k=0}^{[N/2]} \Pi_{f_1}^N(k/N)}\sqrt{\sum\limits_{k=0}^{[N/2]} \Pi_{f_2}^N(k/N)}} \tag{1.26}$$

can be taken as a natural measure of the spectral orthogonality of the series F_1 and F_2. We shall call this characteristic the *spectral correlation coefficient*. Obviously, the value of the spectral correlation coefficient is between 0 and 1, and the absolute value of the standard correlation coefficient between the series F_1 and F_2 does not exceed $\rho_{12}^{(\Pi)}$.

Therefore, smallness of the spectral correlation coefficients between all the subseries of length L (and K as well) of the series $F^{(1)}$ and $F^{(2)}$ is a sufficient condition for approximate separability of these series.

For this condition to hold, it is not necessary that the series $F^{(1)}$ and $F^{(2)}$ be subseries of stationary series. For example, if one of the series is a sum of a slowly varying monotone trend and a noise, while the other series is a high-frequency oscillatory series, then for a sufficiently large N and $L \lesssim N/2$ we have every reason to expect approximate separability of these two series.

Indeed, when L and K are sufficiently large, the periodogram of any subseries of the first series is mostly supported at the low-frequency range, while the main support of the periodogram of the subseries of the second series is in the range of high frequencies. This implies smallness of the spectral correlation coefficients.

We consider another similar example. Assume that the average values of all the subseries of the series $F^{(1)}$ are large, the average values of the corresponding subseries of $F^{(2)}$ are close to zero, and the amplitudes of all the harmonic components of the series $F^{(1)}$ and $F^{(2)}$ are small. Then it is easy to see that all the spectral correlations (1.26) are small.

This example explains the approximate separation of 'large' signals $F^{(1)}$ from series $F^{(2)}$ that oscillate rapidly around zero (a phenomenon that is regularly observed in practice). Note that the conditions imposed on $F^{(2)}$ in this example are typical for finite subseries of aperiodic series.

Thus, in many cases a qualitative analysis of the periodograms of the series provides an insight into their separability features. At the same time, smallness of the spectral correlation coefficients provides only a sufficient condition for approximate separability and it is not difficult to construct examples where exact separability takes place, but the spectral correlation coefficients are large.

Example 1.3 *Separability and spectral correlation*
Let $N = 399$, $a = 1.005$, $T = 200$ and the terms of the series $F^{(1)}$ and $F^{(2)}$ are

$$f_n^{(1)} = a^{-n}, \quad f_n^{(2)} = a^n \cos(2\pi n/T).$$

It can be shown that the choice of the window length $L = 200$ (hence, $K = 200$) leads to the exact separability of the series.

At the same time, the periodogram analysis does not suggest exact separability: the frequency ranges of the two series significantly intersect and the maximum spectral correlation coefficients between their subseries equals 0.43.

Thus, for a fixed window length L we have several characteristics of the quality of (weak) separability, related to the rows and columns of the trajectory matrices:

1. *Cross-correlation matrices between the rows (and columns) of the trajectory matrices.* If all the correlations are zero, then we have exact separability, while the smallness of their absolute values corresponds to approximate separability.
2. *Weighted correlation coefficient* $\rho_{12}^{(w)}$ *between time series* $F^{(1)}$ *and* $F^{(2)}$ *defined by* (1.24). The equality $\rho_{12}^{(w)} = 0$ is a necessary (but not sufficient) condition for separability. Irrespective of separability, the expansion of a time series onto w-uncorrelated or approximately w-uncorrelated components is a highly desirable property.
3. *Matrices of spectral correlations between the rows (and columns) of the trajectory matrices.* The vanishing of all the spectral correlations is a sufficient (but not necessary) condition for separability. Smallness of these correlations implies approximate separability. In the latter case, this separability has a useful interpretation in the language of periodograms: the frequency ranges (with

an account of their powers) of all the subseries of length L (or K) of the series are almost disjoint.

1.5.4 Strong separability

The criteria for (weak) separability of two series for a fixed window length give a solution to the problem that can be stated as follows: 'Does the sum of the SVDs of the trajectory matrices of the series $F^{(1)}$ and $F^{(2)}$ coincide with *one of* the SVDs of the trajectory matrix of the series $F = F^{(1)} + F^{(2)}$?'

Another question, closer to practical needs, can be stated as follows: 'Is it possible to group the matrix terms of *any* SVD of the trajectory matrix $\mathbf{X}$ of the series $F = F^{(1)} + F^{(2)}$, to obtain the trajectory matrices of the series $F^{(1)}$ and $F^{(2)}$?'

If the answer to this question is the affirmative, then we shall say that the series $F^{(1)}$ and $F^{(2)}$ are *strongly separable*. It is clear that if the series are weakly separable and all the singular values of the trajectory matrix $\mathbf{X}$ are different, then strong separability holds.

Moreover, strong separability of two series $F^{(1)}$ and $F^{(2)}$ is equivalent to the fulfillment of the following two conditions: (a) the series $F^{(1)}$ and $F^{(2)}$ are weakly separable, and (b) the collections of the singular values of the trajectory matrices $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ are disjoint.

Let us comment on this. Assume that

$$\mathbf{X}^{(1)} = \sum_k \mathbf{X}_k^{(1)}, \quad \mathbf{X}^{(2)} = \sum_m \mathbf{X}_m^{(2)}$$

are the SVDs of the trajectory matrices $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ of the series $F^{(1)}$ and $F^{(2)}$, respectively. If the series are weakly separable, then

$$\mathbf{X} = \sum_k \mathbf{X}_k^{(1)} + \sum_m \mathbf{X}_m^{(2)}$$

is the SVD of the trajectory matrix $\mathbf{X}$ of the series $F = F^{(1)} + F^{(2)}$.

Assume now that the singular values corresponding to the elementary matrices $\mathbf{X}_1^{(1)}$ and $\mathbf{X}_1^{(2)}$ coincide. This means that using the SVD of the matrix $\mathbf{X}$ we cannot uniquely identify the terms $\mathbf{X}_1^{(1)}$ and $\mathbf{X}_1^{(2)}$ in the sum $\mathbf{X}_1^{(1)} + \mathbf{X}_1^{(2)}$, since these two matrices correspond to the same eigenvalue of the matrix $\mathbf{X}\mathbf{X}^{\mathrm{T}}$.

To illustrate this discussion, let us consider a simple example.

Example 1.4 *Weak and strong separability*
Let $N = 3$, $L = K = 2$ and consider the series

$$F^{(1)} = (1, -a, a^2), \quad F^{(2)} = (1, a^{-1}, a^{-2}), \quad a \neq 0.$$

In this case, the matrices $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ are

$$\mathbf{X}^{(1)} = \begin{pmatrix} 1 & -a \\ -a & a^2 \end{pmatrix}, \quad \mathbf{X}^{(2)} = \begin{pmatrix} 1 & a^{-1} \\ a^{-1} & a^{-2} \end{pmatrix}.$$

Checking the weak separability of the series is easy. At the same time, the matrices $\mathbf{X}^{(1)}(\mathbf{X}^{(1)})^{\mathrm{T}}$ and $\mathbf{X}^{(2)}(\mathbf{X}^{(2)})^{\mathrm{T}}$ have only one positive eigenvalue each, and the singular values of the matrices $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ are

$$\sqrt{\lambda^{(1)}} = 1 + a^2, \quad \sqrt{\lambda^{(2)}} = 1 + a^{-2},$$

respectively. Thus, for any $a \neq 1$ the series $F^{(1)}$ and $F^{(2)}$ are strongly separable, but for $a = 1$ these series are only weakly separable; they loose the strong separability.

Indeed, if $a \neq 1$, then the SVD of the trajectory matrix $\mathbf{X}$ of the series $F^{(1)} + F^{(2)}$ is uniquely defined and has the form

$$\mathbf{X} = \begin{pmatrix} 2 & -a + a^{-1} \\ -a + a^{-1} & a^2 + a^{-2} \end{pmatrix}$$
$$= (1 + a^2)\, U_1 V_1^{\mathrm{T}} + (1 + a^{-2})\, U_2 V_2^{\mathrm{T}} = \mathbf{X}^{(1)} + \mathbf{X}^{(2)},$$

where $U_1 = V_1 = (1, -a)^{\mathrm{T}}/\sqrt{(1 + a^2)}$ and $U_2 = V_2 = (1, a^{-1})^{\mathrm{T}}/\sqrt{(1 + a^{-2})}$.

At the same time, when $a = 1$ there are infinitely many SVDs of the matrix $\mathbf{X}$:

$$\mathbf{X} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = 2\, U_1 V_1^{\mathrm{T}} + 2\, U_2 V_2^{\mathrm{T}},$$

where $\{U_1, U_2\}$ is an arbitrary orthonormal basis in $\mathbf{R}^2$, $V_1 = \mathbf{X}^{\mathrm{T}} U_1/2$ and $V_2 = \mathbf{X}^{\mathrm{T}} U_2/2$. From all these SVDs only one is acceptable, the one corresponding to $U_1 = (1, 1)^{\mathrm{T}}/\sqrt{2}$ and $U_2 = (1, -1)^{\mathrm{T}}/\sqrt{2}$. This SVD gives us the required decomposition $\mathbf{X} = \mathbf{X}^{(1)} + \mathbf{X}^{(2)}$.

In practice, the lack of strong separability (under the presence of the weak separability, perhaps, approximate) becomes essential when the matrix $\mathbf{X}\mathbf{X}^{\mathrm{T}}$ has two close eigenvalues. This leads to an instability of the SVD computations. Let us return to the example of the asymptotic separability of the harmonic series $F^{(1)}$ and $F^{(2)}$ with

$$f_n^{(1)} = \cos(2\pi\omega_1 n), \quad f_n^{(2)} = \cos(2\pi\omega_2 n),$$

where $\omega_1 \neq \omega_2$ and $L, K \to \infty$.

As demonstrated in Section 5.1, for all L and K the SVD of the trajectory matrix of each of the series $F^{(1)}, F^{(2)}$ consists of two terms so that the eigen and factor vectors are the harmonic series with the same frequency as the original series (however, the phase of the harmonics may differ from the phase in the original series). Also, the singular values asymptotically (when $L, K \to \infty$) coincide and do not depend on the frequencies of the harmonics.

Moreover, the series $F^{(1)}$ and $F^{(2)}$ are asymptotically separable, and, since these series have the same amplitude, asymptotically all four singular values are equal. Therefore, even when N, L, and K are large, we cannot as a rule separate the periodicities $F^{(1)}$ and $F^{(2)}$ out of the sum $F = F^{(1)} + F^{(2)}$ (if, of course, we do not use special rotations in the four-dimensional eigenspace of the trajectory matrix $\mathbf{X}$ of the series F).

Another way to deal with the case of equal singular values is described in Section 1.7.3. It is based on the simple fact that if the series $F^{(1)}$ are $F^{(2)}$ weakly separable, then we can always find a constant $c \neq 0$ such that the series $F^{(1)}$ and $cF^{(2)}$ are strongly separable.

The presence of close singular values is the reason why SSA often fails to decompose the component consisting of many harmonics with similar weights. If these weights are small, then it may be natural to consider such components as the noise components.

1.6 Choice of SSA parameters

In this section we discuss the role of the parameters in Basic SSA and the principles for their selection. As was mentioned in Section 1.4.1, we assume that the time series under consideration can be regarded as a sum of a slowly varying trend, different oscillatory components, and a noise. The time series analysis issues related to this assumption were discussed in Section 1.4.2.

Certainly, the choice of parameters depends on the data we have and the analysis we have to perform. We discuss the selection issues separately for all the main problems of time series analysis.

There are two parameters in Basic SSA: the first is an integer L, the window length, and the second parameter is structural; loosely speaking, it is the way of grouping.

1.6.1 Grouping effects

Assume that the window length L is fixed and we have already made the SVD of the trajectory matrix of the original time series. The next step is to group the SVD terms in order to solve one of the problems discussed in Section 1.4.2. We suppose that this problem has a solution; that is, the corresponding terms can be found in the SVD, and the result of the proper grouping would lead to the (approximate) separation of the time series components (see Section 1.5).

Therefore, we have to decide what the proper grouping is and how to find the proper groups of the eigentriples. In other words, we need to identify an eigentriple corresponding to the related time series component. Since each eigentriple consists of an eigenvector (left singular vector), a factor vector (right singular vector) and a singular value, this is to be achieved using only the information contained in these vectors (considered as time series) and the singular values.

(a) General issues

We start by mentioning several purely theoretical results about the eigentriples of several 'simple' time series (see Section 5.1).

Exponential-cosine sequences
Consider the series

$$f(n) = Ae^{\alpha n}\cos(2\pi\omega n + \phi), \tag{1.27}$$

$\omega \in [0, 1/2]$, $\phi \in [0, 2\pi)$ and denote $T = 1/\omega$.

Depending on the parameters, the exponential-cosine sequence produces the following eigentriples:

1. *Exponentially modulated harmonic time series with a frequency* $\omega \in (0, 1/2)$
If $\omega \in (0, 1/2)$, then for any L and N the SVD of the trajectory matrix has two terms. Both eigenvectors (and factor vectors) have the same form (1.27) with the same frequency ω and the exponential rate α. If $\alpha \le 0$ then for large N, L and $K = N - L + 1$, both singular values are close (formally they asymptotically coincide for $L, K \to \infty$). Practically, they are close enough when L and K are several times greater than $T = 1/\omega$.
2. *Exponentially modulated saw-tooth curve* ($\omega = 1/2$)
If $\omega = 1/2$ and $\sin(\phi) \neq 0$, then f_n is proportional to $(-e^{\alpha})^n$. In this case for any L the corresponding SVD has just one term. Both singular vectors have the same form as the initial series.
3. *Exponential sequence* ($\omega = 0$)
If $\omega = 0$ and $\cos(\phi) \neq 0$, then f_n is proportional to $e^{\alpha n}$ and we have an exponential series. For any N and window length L, the trajectory matrix of the exponential series has only one eigentriple. Both singular vectors of this eigentriple are exponential with the same parameter α.
4. *Harmonic series* ($\alpha = 0$)
If $\alpha = 0$ and $\omega \neq 0$, then the series is a pure harmonic one. The eigenvectors and factor vectors are harmonic series with the same ω. If $\omega \neq 1/2$ and $T = 1/\omega$ is a divisor of K and L, then both singular values coincide.

Polynomial series
Consider a polynomial series of the form

$$f_n = \sum_{k=0}^{m} a_k n^k, \quad a_m \neq 0.$$

1. *General case*
If f_n is a polynomial of degree m, then the order of the corresponding SVD does not exceed $m + 1$ and all the singular vectors are polynomials too; also their degrees do not exceed m.
2. *Linear series*
For a linear series

$$f_n = an + b, \quad a \neq 0,$$

with arbitrary N and L the SVD of the L-trajectory matrix consists of two terms. All singular vectors are also linear series with the same $|a|$.

Note that the exponential-cosine and linear series (in addition to the sum of two exponential series with different rates) are the only series that have at most two terms in the SVD of their trajectory matrices for any series length N and window length $L \geq 2$ (see the proofs in Section 5.1). This fact helps in their SSA identification as components of more complex series.

Let us now turn to the various different grouping problems and the corresponding grouping principles. We start with mentioning several general rules.

1. *If we reconstruct a component of a time series with the help of just one eigentriple and both singular vectors of this eigentriple have a similar form, then the reconstructed component will have approximately the same form.*

   This means that when dealing with a single eigentriple we can often predict the behaviour of the corresponding component of the series. For example, if both singular vectors of an eigentriple resemble linear series with similar slopes, then the corresponding component is also almost linear. If the singular vectors have the form of the same exponential series, then the trend has a similar form. Harmonic-like singular vectors produce harmonic-like components (compare this with the results for exponential-cosine series presented at the beginning of this section).

   The conservation law under discussion can be extended to incorporate monotonicity (monotone singular vectors generate monotone components of the series) as well as some other properties of time series.

2. *If $L \ll K$ then the factor vector in an eigentriple has a greater similarity with the component, reconstructed from this eigentriple, than the eigenvector.* Consequently we can approximately predict the result of reconstruction from a single eigentriple with the help of its factor vector.

3. *If we reconstruct a series with the help of several eigentriples, and the periodograms of their singular vectors are (approximately) supported on the same frequency interval $[a, b]$, then the frequency power of the reconstructed series will be mainly supported on $[a, b]$.* This feature is analogous to that in item 1 but concerns several eigentriples and is formulated in terms of the Fourier expansions.

4. *The larger the singular value of the eigentriple is, the bigger the weight of the corresponding component of the series.* Roughly speaking, this weight may be considered as being proportional to the singular value.

Now let us turn to the SSA problems.

(b) Grouping for extraction of trends and smoothing

1. *Trends*

According to our definition, trend is a slowly varying component of a time series which does not contain oscillatory components. Assume that the time series itself is such a component alone. Practice shows that in this case, one or more leading singular vectors will be slowly varying as well. Exponential and polynomial sequences are good examples of this situation.

For a general series F we typically assume that its trend component $F^{(1)}$ is (approximately) strongly separable from all the other components. This means that among the eigentriples of the series F, there are eigentriples that approximately correspond to the SVD components of the series $F^{(1)}$.

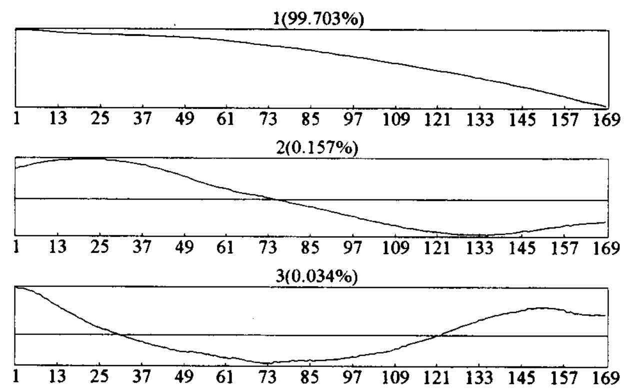

Figure 1.23 *Production: three factor vectors of the 'accurate trend'.*

Thus, *to extract a trend of a series, we have to collect all the elementary matrices related to slowly varying singular vectors.*

The ordinal numbers of these eigentriples depend not only on the trend $F^{(1)}$ itself, but on the 'residual series' $F^{(2)} = F - F^{(1)}$ also. Consider two different extremes. First, let the series F have a strong trend tendency $F^{(1)}$ with a relatively small oscillatory-and-noise component $F^{(2)}$. Then most of the trend eigentriples will have the leading positions in the SVD of the whole series F. Certainly, some of these eigentriples can have small singular values, especially if we are looking for a more or less refined trend.

For instance, in the 'Production' example (Section 1.3.1, Fig. 1.2) a reasonably accurate trend is described by the three leading eigentriples, and the singular value in the third eigentriple is five times smaller than the second one. The corresponding factor vectors are shown at Fig. 1.23.

The other extreme is the situation where we deal with high oscillations on the background of a small and slow general tendency. Here, the leading elementary matrices describe oscillations, while the trend eigentriples can have small singular values (and therefore can be far from the top in the ordered list of the eigentriples).

2. *Smoothing*

The problem of smoothing may seem similar to that of trend extraction but it has

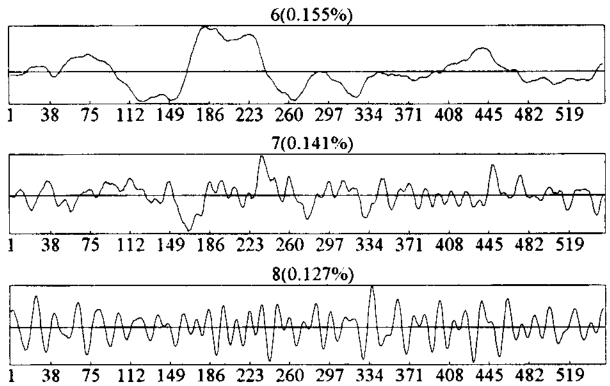

Figure 1.24 *Tree rings: factor vectors with ordinal numbers 6, 7, and 8.*

its own specifics. In particular, we can smooth any time series, even if it does not have an obvious trend component (the example 'Tree rings' in Section 1.3.2 is just one of this sort). That means that for the extraction of a trend, we collect all the eigentriples corresponding to a slowly varying (but not oscillatory) part of the series; at the same time a smoothed component of a series is composed of a collection of the eigentriples whose singular vectors do not oscillate rapidly.

In the 'Tree rings' example (Section 1.3.2, Fig. 1.3) seven leading eigentriples corresponding to low frequencies were chosen for smoothing. Fig. 1.24 demonstrates factor vectors with numbers 6, 7, and 8. The sixth factor vector is slowly varying, and therefore it is selected for smoothing, while the eighth one corresponds to rather high frequencies and is omitted. As for the 7th factor vector, it demonstrates mixing of rather low (0-0.05) and high (approximately 0.08) frequencies. To include all low frequencies, the seventh eigentriple was also selected for smoothing.

For relatively long series, periodogram analysis serves as a good description of the matter. Periodograms (see Figs. 1.25 and 1.26) confirm that smoothing in the 'Tree rings' example splits the frequencies into two parts rather well. The small intersection of the frequency ranges for the result of the smoothing and the residual is due to mixing of some of the frequencies in some of the eigentriples (see the 7th factor vector in Fig. 1.24).

(c) Grouping for oscillations

1. *Harmonic series*

Let us start with a pure harmonic with a frequency ω and a certain phase and amplitude. Since we assume that such a component $F^{(1)}$ in the original series is approximately strongly separable from $F^{(2)} = F - F^{(1)}$, we may hope that two (or one if $\omega = 1/2$) SVD eigentriples of the trajectory matrix generated by

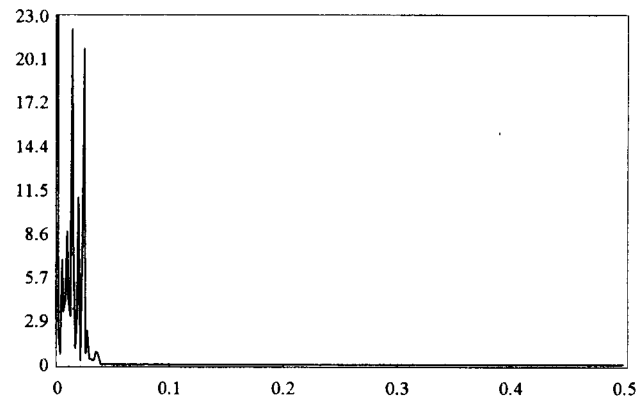

Figure 1.25 *Tree rings: periodogram of smoothed series.*

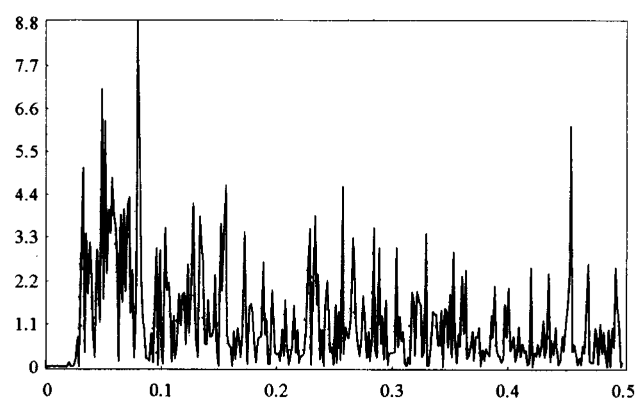

Figure 1.26 *Tree rings: periodogram of smoothing residuals.*

F correspond to $F^{(1)}$. The problem is, therefore, to identify these eigentriples among all other eigentriples generated by F.

Let $\omega \neq 1/2$. As was stated in the example of the exponential-cosine function, the pure harmonic corresponding to (1.27) with $\alpha = 0$ generates an SVD of order two with the singular vectors having the same harmonic form.

Consider the ideal situation where $T = 1/\omega$ is a divisor of the window length L and $K = N - L + 1$. Since T is an integer, it is a period of the harmonic.

Let us take, for definiteness, the left singular vectors (that is, the eigenvectors). In the ideal situation described above, the eigenvectors have the form of sine and cosine sequences with the same T and the same phases. The factor vectors are also of the same form.

Thus, the identification of the components that are generated by a harmonic is reduced to the determination of these pairs. Viewing the pairwise scatterplots of the eigenvectors (and factor vectors) simplifies the search for these pairs; indeed, the pure sine and cosine with equal frequencies, amplitudes, and phases create the scatterplot with the points lying on a circle.

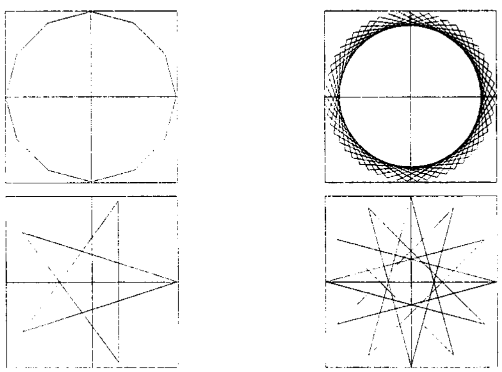

Figure 1.27 *Scatterplots of sines/cosines.*

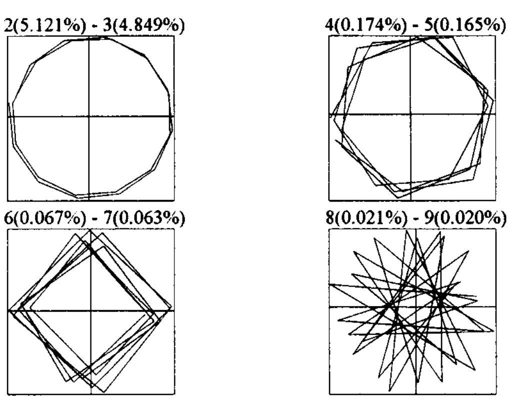

Figure 1.28 *Eggs: scatterplots of paired harmonic factor vectors.*

If $T = 1/\omega$ is an integer, then these points are the vertices of the regular T-vertex polygon. For the rational frequency $\omega = q/p < 1/2$ with relatively prime integers p and q, the points are the vertices of the regular p-vertex polygon. Fig. 1.27 depicts scatterplots of four pairs of sin/cosine sequences with zero phases, the same amplitudes and frequencies $1/12$, $10/53$, $2/5$, and $5/12$.

Small deviations from this ideal situation would imply that the points in the scatterplots are no longer exactly the vertices of the regular p-vertex polygon, although staying reasonably close to them. As an example, Fig. 1.28 provides scatterplots of paired factor vectors in the 'Eggs' example (Section 1.3.3), corresponding to the harmonics with the frequencies $1/12, 2/12, 3/12$ and $5/12 = 1/2.4$.

Therefore, an analysis of the pairwise scatterplots of the singular vectors allows one to visually identify those eigentriples that correspond to the harmonic components of the series, provided these components are separable from the residual component.

In practice, the singular values of the two eigentriples of a harmonic series are often close to each other, and this fact simplifies the visual identification of the harmonic components. (In this case, the corresponding eigentriples are, as a rule, consecutive in the SVD order.) Such a situation typically occurs when, say, both L and K are several times greater than $1/\omega$.

Alternatively, if the period of the harmonic is comparable to N, then the corresponding singular values may not be close and therefore the two eigentriples may not be consecutive. The same effect often happens when the two singular values of a harmonic component are small and comparable with the singular values of the components of the noise.

The series F may contain several purely harmonic components, and the frequency of each one should be identified using the corresponding pair of eigentriples. In easy cases it is a straightforward operation (see, for example, the scatterplot of period 12 at Fig. 1.28). In more complex cases, the periodogram analysis applied to the singular vectors often helps.

One more method of approximate identification of the frequency, which can be useful even for short series, is as follows. Consider two eigentriples, which approximately describe a harmonic component with frequency ω_0. Then the scatterplot of their eigenvectors can be expressed as a two-dimensional curve with Euclidean components of the form

$$\begin{aligned} x(n) &= r(n)\cos(2\pi\omega(n)n + \phi(n)) \\ y(n) &= r(n)\sin(2\pi\omega(n)n + \phi(n))\,, \end{aligned}$$

where the functions r, ω and ϕ are close to constants and $\omega(n) \approx \omega_0$. The polar coordinates of the curve vertices are $(r(n), \delta(n))$ with $\delta(n) = 2\pi\omega(n)n + \phi(n)$.

Since $\Delta_n \stackrel{\mathrm{def}}{=} \delta(n+1) - \delta(n) \approx 2\pi\omega_0$, one can estimate ω_0 by averaging polar angle increments Δ_n $(n = 0, \ldots, L-1)$. The same procedure can be applied to a pair of factor vectors.

If the period of the harmonic is equal to 2, that is $\omega = 1/2$, then the situation is simpler since in this case the SVD of this harmonic consists of only one eigentriple, and the corresponding eigenvector and factor vector have a saw-tooth form. Usually the identification of such vectors is easy.

2. *Grouping for identification of a general periodic component*

Consider now the more comprehensive case of extraction of a general periodic component $F^{(1)}$ out of the series F. If the integer T is the period of this component, then according to Section 1.4.1,

$$f_n^{(1)} = \sum_{k=1}^{[T/2]} a_k \cos(2\pi kn/T) + \sum_{k=1}^{[T/2]} b_k \sin(2\pi kn/T). \qquad (1.28)$$

Hence (see Section 5.1), there are at most $T-1$ matrix components in the SVD of the trajectory matrix of the series $F^{(1)}$, for any window length $L \geq T-1$. Moreover, for large L the harmonic components in the sum (1.28) are approximately strongly separable, assuming their powers $a_k^2 + b_k^2$ are all different.

In this case, each of these components produces either two (for $k \neq T/2$) or one (for $k = T/2$) eigentriples with singular vectors of the same harmonic kind.

Therefore, under the assumptions:

(a) the series $F^{(1)}$ is (approximately) strongly separable from $F^{(2)}$ in the sum $F = F^{(1)} + F^{(2)}$ for the window length L,

(b) all the nonzero powers $a_k^2 + b_k^2$ in the expansion (1.28) are different, and

(c) L is large enough,

we should be able to approximately separate all the eigentriples, corresponding to the periodic series $F^{(1)}$ in the SVD of the trajectory matrix of the whole series F.

To perform this separation, it is enough to identify the eigentriples that correspond to all the harmonics with frequencies k/T (this operation has been described above in the section 'Grouping for oscillations: Harmonic series') and group them.

For instance, if it is known that there is a seasonal component in the series F and the data is monthly, then one must look at periodicities with frequencies $1/12$ (annual), $1/6$ (half-year), $1/4$, $1/3$ (quarterly), $1/2.4$, and $1/2$. Of course, some of these periodicities may be missing.

In the example 'Eggs' (see Section 1.3.3), all the above-mentioned periodicities are present; they can all be approximately separated for the window length 12.

If some of the nonzero powers in (1.28) coincide, then the problem of identification of the eigentriples in the SVD of the series F has certain specifics. For instance, suppose that $a_1^2 + b_1^2 = a_2^2 + b_2^2$. If L is large, then four of the singular values in the SVD of the trajectory matrix of the series F are (approximately) the same, and the corresponding singular vectors are linear combinations of the harmonics with two frequencies: $\omega_1 = 1/T$ and $\omega_2 = 2/T$.

Therefore, the components with these frequencies are not strongly separable. In this case, the periodogram analysis of the singular vectors may help a lot; if their periodograms have sharp peaks at around the frequencies ω_1 and ω_2, then the corresponding eigentriples must be regarded as those related to the series $F^{(1)}$.

Figs. 1.29 and 1.30 depict the periodograms of the sixth and the eighth eigenvectors from the example 'Rosé wine' (Section 1.4.1, Fig. 1.17). Due to the closeness of the singular values in the eigentriples 6-9 (the eigenvalue shares for these eigentriples lie between 0.349% and 0.385%) the harmonics (mostly, with the fre-

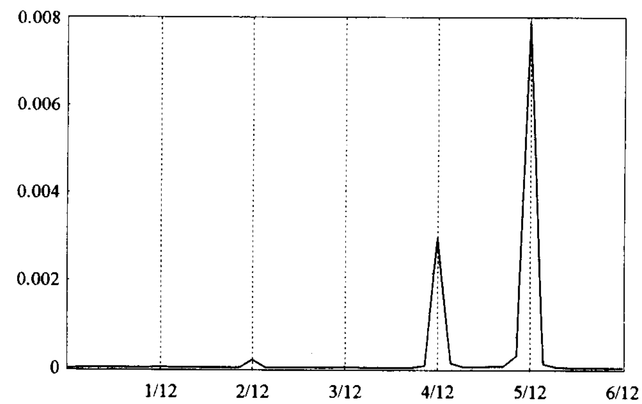

Figure 1.29 *Rosé wine: periodogram of the sixth eigenvector.*

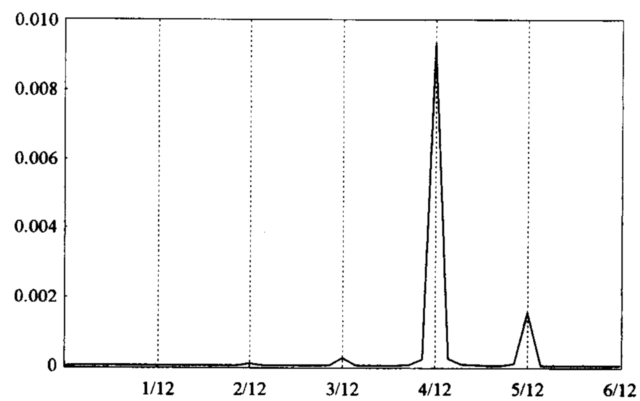

Figure 1.30 *Rosé wine: periodogram of the eighth eigenvector.*

quencies 4/12 and 5/12) are being mixed up, and this is perfectly reflected in the periodograms provided.

3. *Modulated periodicities*

The case of amplitude-modulated periodicity is much more complex since we do not assume the exact form of modulation. However, the example of the exponentially modulated harmonic (1.27) with $\alpha \neq 0$ shows that sometimes identification of the components of such signals can be performed. Let us start with this series. Being one of the simplest, the exponentially modulated harmonic can be considered as an additive component of some econometric series describing an

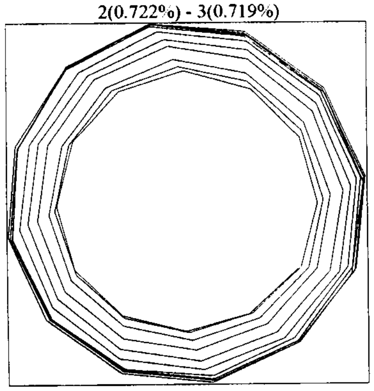

Figure 1.31 *Hotels: spiral in the scatterplot of factor vectors.*

exponential growth associated with an exponentially modulated seasonal oscillation.

On the whole, the situation here is analogous to the case of a pure harmonic series. If the frequency is not $1/2$ and the window length L is large, then we have two eigentriples with approximately equal singular values, both being characterized by the singular vectors of the same exponential-cosine shape. Therefore, the scatterplots of these pairs of eigen/factor vectors have the form of a spiral and visually are easily distinguishable.

For instance, Fig. 1.31 depicts the scatterplot of the second and third factor vectors corresponding to the annual periodicity with increasing amplitude in the series 'Hotels' for $L = 48$ (description of this series can be found in Section 1.7.1).

If the period of the harmonic is 2, then the series is $f_n = (-a)^n$, where $a = e^{\alpha}$. The singular vectors of the single eigentriple, created by this series, have exactly the same shape as the modulated saw-tooth curve. Of course, visual identification of this series is also not difficult.

If the series modulating the pure harmonic is not exponential, then the extraction of the corresponding components is much more difficult (see the theoretical results of Section 5.1 concerning the general form of infinite series that generate finite SVDs of their trajectory matrices). Let us, however, describe the situation when the identification of the components is possible.

As we already mentioned in Section 1.4.1, if the amplitude of the modulated harmonic $F^{(1)}$ varies slowly, then the range of frequencies is concentrated around the main frequency, which can clearly be seen in the periodogram of this modulated harmonic. If the window length L and $K = N - L + 1$ are large, then all the singular vectors of this series will have the same property.

Therefore, if the series $F^{(1)}$ is (approximately) strongly separable from a series $F^{(2)}$ in the sum $F = F^{(1)} + F^{(2)}$, then one can expect that the frequency interval of $F^{(2)}$ has a small (in terms of powers) intersection with the frequency interval of the modulated harmonic $F^{(1)}$. Thus, by analyzing periodograms of the singular vectors in all the eigentriples of the series F, one can hope to identify the majority of those that (approximately) describe $F^{(1)}$.

The situation is similar when an arbitrary periodic (not necessarily harmonic) signal is being modulated. In this case, each term in the sum (1.28) is multiplied by the function modulating the amplitude, and every frequency k/T gives rise to a group of neighbouring frequencies. Under the condition that $F^{(1)}$ and $F^{(2)}$ are strongly separable, one should look for singular vectors in the SVD of the series F such that their periodograms are concentrated around the frequencies k/T.

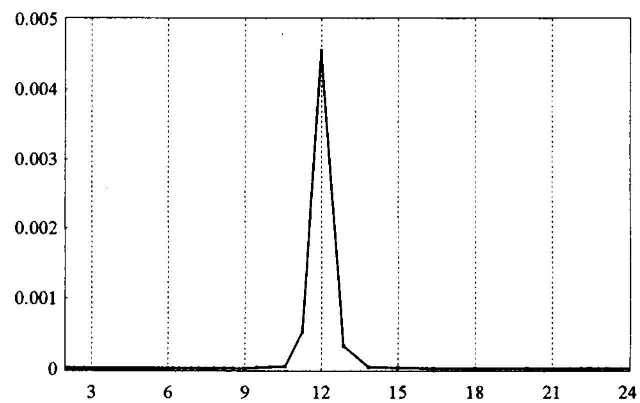

Figure 1.32 *Unemployment: periodogram of the 4th eigenvector (in periods).*

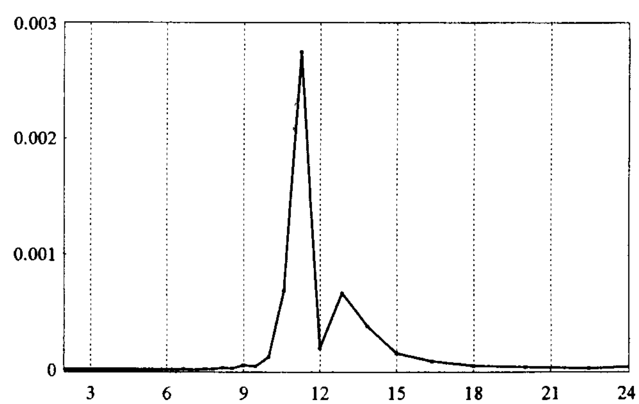

Figure 1.33 *Unemployment: periodogram of the 12th eigenvector (in periods).*

In the example 'Unemployment' (Section 1.3.6) the modulated annual periodicity generates, in particular, the pairs of eigentriples 3-4 and 12-13. Figs. 1.32 and 1.33 depict the periodograms of the fourth and the twelfth eigenvectors, describing the main frequency 1/12 and close frequencies, respectively. These figures demonstrate the typical shape of the periodograms of the singular vectors generated by a modulated harmonic.

(d) Grouping for finding a refined decomposition of a series

The problem of finding a refined structure of a series by Basic SSA is equivalent to the identification of the eigentriples of the SVD of the trajectory matrix of this series, which correspond to the trend, various oscillatory components, and noise. The principles of grouping for identification of the trend and oscillatory components have been described above.

As regards noise, we should always bear in mind the intrinsic uncertainty of this concept under the lack of a rigorous mathematical model for noise. From the practical point of view, a natural way of noise extraction is the grouping of the eigentriples, which do not seemingly contain elements of trend and oscillations. In doing that, one should be careful about the following.

1. If the frequency range of the noise contains the frequency of a harmonic component of the signal, then the harmonic component reconstructed from the related SVD eigentriples will also include the part of the noise corresponding to this frequency (compare Marple, 1987, Chapter 13.3). Analogously, the extracted trend 'grasps' the low-frequency parts of the noise, if there are any. If the frequency ranges of the signal and noise do not intersect, then this effect does not appear.

2. If the amplitude of a harmonic component of the signal is small and the noise is large, then the singular values corresponding to the harmonic and the noise may be close. That would imply the impossibility of separating the harmonic from the noise on the basis of the analysis of the eigentriples for the whole series. Speaking more formally, the harmonic and noise would not be strongly separable. This effect disappears asymptotically, when $N \to \infty$.

3. The SVD of the trajectory matrix of the pure noise (that is, of an aperiodic stationary sequence) for large N, L and K should be expected to contain at least some (leading) eigentriples looking like harmonics (see Section 6.4.3). The components of the original series, reconstructed from these eigentriples, will look similar. This necessitates a profound control of the interpretability of the reconstructed components.

Certainly, the above discussion concerns also the problem of noise reduction, which is a particular case of the problem under consideration.

(e) Grouping hints

A number of characteristics of the eigentriples of the SVD of the trajectory matrix of the original series may very much help in making the proper grouping for extraction of the components from the series. Let us discuss two of these characteristics.

1. *Singular values*
As mentioned above, if N, L and K are sufficiently large, then each harmonic different from the saw-tooth one produces two eigentriples with close singular values. Moreover, a similar situation occurs if we have a sum of several different harmonics with (approximately) the same amplitudes; though the corresponding singular vectors do not necessarily correspond to a pure harmonic (the frequencies can be mixed), they can still form pairs with close singular values and similar shapes.

Therefore, explicit plateaux in the eigenvalue spectra prompts the ordinal numbers of the paired eigentriples. Fig. 1.34 depicts the plot of leading singular val-

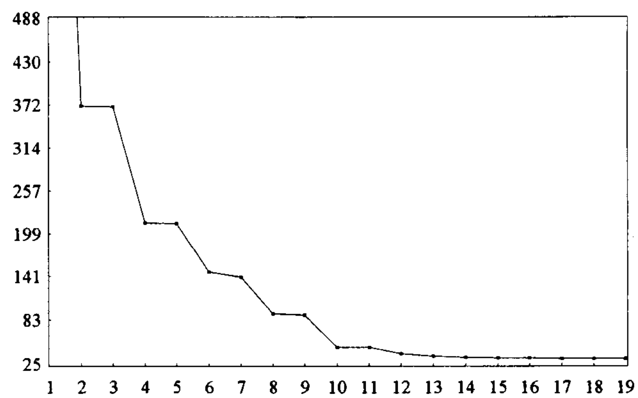

Figure 1.34 *Births: leading singular values.*

ues for the example 'Births' (Section 1.3.4). Five evident pairs with almost equal leading singular values correspond to five (almost) harmonic components of the 'Births' series: eigentriple pairs 2-3, 4-5 and 10-11 are related to a one-week periodicity with frequencies $1/7$, $2/7$ and $3/7$, while pairs 6-7 and 8-9 describe the annual birth cycle (frequencies $\approx 1/365$ and $\approx 2/365$). Note that the first singular value, equal to 4772.5, corresponds to the trend component of the series and is omitted in Fig. 1.34.

Another useful insight is provided by checking breaks in the eigenvalue spectra. As a rule, a pure noise series produces a slowly decreasing sequence of singular values. If such a noise is added to a signal, described by a few eigentriples with large singular values, then a break in the eigenvalue spectrum can distinguish signal eigentriples from the noise ones.

Note that in general there are no formal procedures enabling one to find such a break. Moreover, for complex signals and large noise, the signal and noise eigentriples can be mixed up with respect to the order of their singular values.

At any rate, singular values give important but supplementary information for grouping; the structure of the singular vectors is more essential.

2. **w**-*Correlations*

As discussed earlier, a necessary condition for the (approximate) separability of two series is the (approximate) zero **w**-correlation of the reconstructed components. On the other hand, the eigentriples entering the same group can correspond to highly correlated components of the series.

Thus, a natural hint for grouping is the matrix of the absolute values of the **w**-correlations, corresponding to the full decomposition (in this decomposition each group corresponds to only one matrix component of the SVD). This matrix for an artificial series F with

$$f_n = e^{n/400} + \sin(2\pi n/17) + 0.5\sin(2\pi n/10) + \varepsilon_n, \quad n = 0, \ldots, 339, \quad (1.29)$$

standard Gaussian white noise ε_n, and $L = 85$, is depicted in Fig. 1.35 (**w**-correlations for the first 30 reconstructed components are shown in 20-colour scale from white to black corresponding to the absolute values of correlations from 0 to 1).

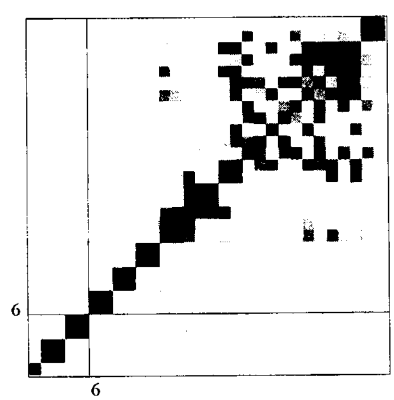

Figure 1.35 *Series* (1.29)*: matrix of* **w**-*correlations.*

The form of this matrix gives an indication of how to make the proper grouping: the leading eigentriple describes the exponential trend, the two pairs of the subsequent eigentriples correspond to the harmonics, and the large sparkling square indicates the white noise components. Note that theoretical results of Section 6.1.3 tell us that such a separation can be indeed (asymptotically) valid.

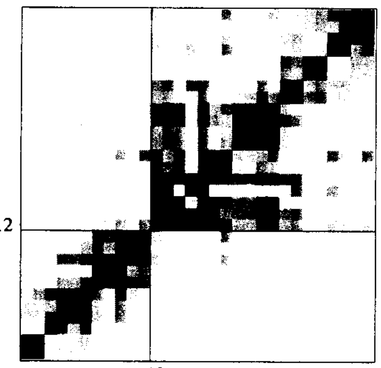

Figure 1.36 *White dwarf: matrix of* **w***-correlations.*

For the example 'White dwarf' (Section 1.3.2) with $L = 100$, the matrix of the absolute values of the w-correlations of the reconstructed components produced from the leading 30 eigentriples is depicted in Fig. 1.36 in the manner of Fig. 1.35.

It is clearly seen that the splitting of all the eigentriples into two groups — from the first to the 11th and the rest — gives rise to a decomposition of the trajectory matrix into two almost orthogonal blocks, with the first block corresponding to the smoothed version of the original series and the second block corresponding to the residual, see Figs. 1.5 and 1.6 in Section 1.3.2. Note that despite the fact that some w-correlations between the eigentriples in different blocks exceed 0.2, the reconstructed components are almost w-uncorrelated: $|\rho^{(w)}| = 0.004$.

The similarity of Figs. 1.35 and 1.36 gives us an additional argument in favour of the assertion that in the 'White dwarf' example smoothing leads to noise reduction.

1.6.2 Window length effects

Window length is the main parameter of Basic SSA, in the sense that its improper choice would imply that no grouping activities will help to obtain a good SSA decomposition. Moreover, it is the single parameter of the decomposition.

Selection of the proper window length depends on the problem in hand, and on preliminary information about the time series. In the general case no universal rules and unambiguous recommendations can be given for the selection of the window length. The main difficulty here is caused by the fact that variations in the window length may influence both weak and strong separability features of SSA, i.e., both the orthogonality of the appropriate time series intervals and the closeness of the singular values.

However, there are several general principles for the selection of the window length L that have certain theoretical and practical grounds. Let us discuss these principles.

(a) General effects

1. The SVDs of the trajectory matrices, corresponding to the window lengths L and $K = N - L + 1$, are equivalent (up to the symmetry: left singular vectors $\leftrightarrow$ right singular vectors). Therefore, *for the analysis of structure of time series by Basic SSA it is meaningless to take the window length larger than half of the time series length.*
2. Bearing in mind the previous remark, *the larger the window length is, the more detailed is the decomposition of the time series.* The most detailed decomposition is achieved when the window length is approximately equal to half of time series length, that is when $L \sim N/2$. The exceptions are the so-called series of finite rank, where for any L larger than d and $N > 2d - 1$ (d is the rank of the series) the number of nonzero components in the SVD of the series is equal to d and does not depend on the window length (see Section 1.6.1 for examples and Section 5.1 for general results).
3. The effects of weak separability.
   - Since the results concerning weak separability of time series components are mostly asymptotic (when $L, K \to \infty$), in the majority of examples to achieve a better (weak) separation one has to choose large window lengths. In other words, *a small window length could mix up interpretable components.*
   - If the window length L is relatively large (say, it is equal to several dozen), then the (weak) *separation results are stable with respect to small perturbations in L.*
   - On the other hand, for *specific series and tasks, there are concrete recommendations for the window length selection,* which can work for a relatively small N (see section 'Window length for periodicities' below).
4. The effects of closeness of singular values.
   The negative effects due to the closeness of the singular values related to different components of the original series (that is, the absence of strong separability in the situation where (approximate) weak separability does hold), are not easily formalized in terms of the window length. These effects are often difficult to overcome by means of selection of L alone.

   Let us mention two other issues related to the closeness of the singular values.

   - *For the series with a complex structure, too large window length L can produce an undesirable decomposition of the series components of interest,* which may lead, in particular, to their mixing with other series components.

This is an unpleasant possibility, especially since a significant reduction of L can lead to a poor quality of the (weak) separation.

- Alternatively, sometimes in these circumstances *even a small variation in the value of L can reduce mixing and lead to a better separation of the components*, i.e., provide a transition from weak to strong separability. At any rate, it is always worthwhile trying several window lengths.

(b) Window length for extraction of trends and smoothing

1. Trends

In the problem of trend extraction, a possible contradiction between the requirements for weak and strong separability emerges most frequently. Since trend is a relatively smooth curve, its separability from noise and oscillations requires large values of L.

On the other hand, if trend has a complex structure, then for values of L that are too large, it can be described only by a large number of eigentriples with relatively small singular values. Some of these singular values could be close to those generated by oscillations and/or noise time series components.

This happens in the example 'Births', where the window length of order 1000 and more (the series length is 5113) leads to the situation where the components of the trend are mixed up with the components of the annual and half-year periodicities (a short description of the series 'Births' is provided in Section 1.3.4; other aspects relating to the choice of the window length in this example are discussed below).

The problem is complex; there is a large variety of situations. We briefly consider, on a quantitative level, two extreme cases: when the trend can be extracted relatively easily, and when the selection of the window length for extraction of trend is difficult or even impossible.

(i) *Trends: reliable separation*

Let $F = F^{(1)} + F^{(2)}$ where $F^{(1)}$ is a trend and $F^{(2)}$ is the residual. We assume the following.

1. The series $F^{(1)}$ is 'simple'. The notion of 'simplicity' can be understood as follows:

   - From the theoretical viewpoint, the series $F^{(1)}$ is well approximated by a series with finite and small rank d (for example, if it looks like an exponential, $d = 1$, a linear function, $d = 2$, a quadratic function, $d = 3$, etc.). See Section 5.1 for a description of the series of finite rank.
   - We are interested in the extraction of the general tendency of the series rather than of the refined trend.
   - In terms of frequencies, the periodogram of the series $F^{(1)}$ is concentrated in the domain of rather small frequencies.

- In terms of the SSA decomposition, the few first eigentriples of the decomposition of the trajectory matrix of the series $F^{(1)}$ are enough for a reasonably good approximation of it, even for large L.

2. Assume also that the series $F^{(1)}$ is much 'larger' than the series $F^{(2)}$ (for instance, the inequality $||F^{(1)}|| \gg ||F^{(2)}||$ is valid).

Suppose that these assumptions hold and the window length L provides a certain (weak, approximate) separation of the time series $F^{(1)}$ and $F^{(2)}$. We can expect that in the SVD of the trajectory matrix of the series F, the leading eigentriples will correspond to the trend $F^{(1)}$; i.e., they will have larger singular values than the eigentriples corresponding to $F^{(2)}$. In other words, strong separation occurs. Moreover, the window length L, sufficient for the separation, should not be very large in this case in view of the 'simplicity' of the trend.

This situation is illustrated by the example 'Production' (Section 1.3.1, Figs. 1.1 and 1.2), where both trend versions are described by the leading eigentriples. Analogously, if we are interested only in extracting the main tendency of the series 'Unemployment' (Section 1.3.6), then, according to Fig. 1.37, taking just one leading eigentriple will be a perfectly satisfactory decision for $L = 12$.

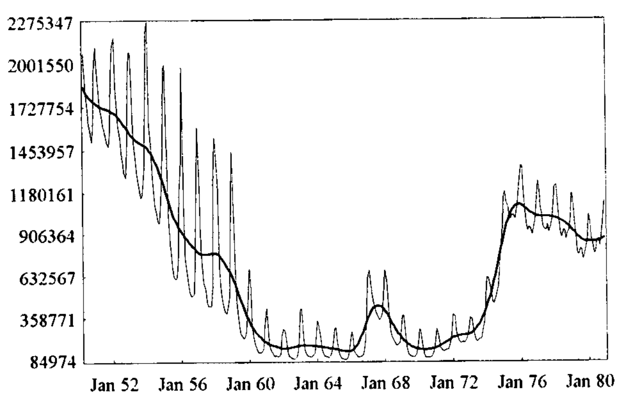

Figure 1.37 *Unemployment: $L = 12$ for extraction of the main tendency.*

(ii) *Trends: difficult case*
Much more difficult situations arise if we want to extract a refined trend $F^{(1)}$, when the residual $F^{(2)}$ has a complex structure (for example, it includes a large noise component) with $||F^{(2)}||$ being large. Then large L can cause not only mixing of the ordinal numbers of the eigentriples corresponding to $F^{(1)}$ and $F^{(2)}$ (this is the case of the 'Unemployment' example), but also closeness of the corresponding singular values, and therefore a lack of strong separability.

Certainly, there are many intermediate situations between these two extremes. Consider, for instance, the 'England temperatures' example (average annual tem-

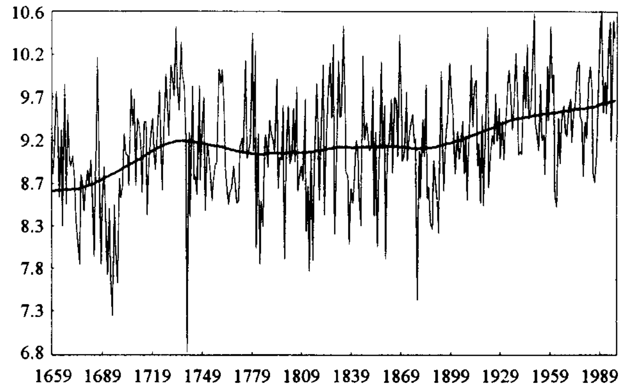

Figure 1.38 *England temperatures:* $L = 48$ *for extraction of the main tendency.*

peratures, Central England, from 1659 to 1998). Here the problem of extraction of a smooth trend can easily be solved: under the choice of L equal to several dozen, the first eigentriple always describes the general tendency; see Fig. 1.38 for the choice $L = 48$.

This happens because relatively small values of L are enough to provide (weak) separability, the trend has a simple form, and it thus corresponds to only one eigentriple; this eigentriple is leading due to a relatively large mean value of the series.

At the same time, if we wish to centre the series (which may seem a natural operation since in this kind of problem the deviation from the mean is often the main interest), then small values of L, say $L < 30$, do not provide weak separation. Large values of L, say $L > 60$, mix up the trend eigentriple with some other eigentriple of the series; this is a consequence of the complexity of the series structure.

2. Smoothing

Generally, the recommendations concerning the selection of the window length for the problem of smoothing are similar to the case of the trend extraction. This is because these two problems are closely related. Let us describe the effects of the window length in the language of frequencies.

Treating smoothing as removing of the high-frequency part of a series, we have to take the window length L large enough to provide separation of this low-frequency part from the high-frequency one. If the powers of all low frequencies of interest are significantly larger than those of the high ones, then the smoothing problem is not difficult, and the only job is collecting several leading eigentriples. This is the case for the 'Tree rings' and 'White dwarf' examples of Section 1.3.2. Here, the larger we take L, the narrower the interval of low frequencies we can extract.

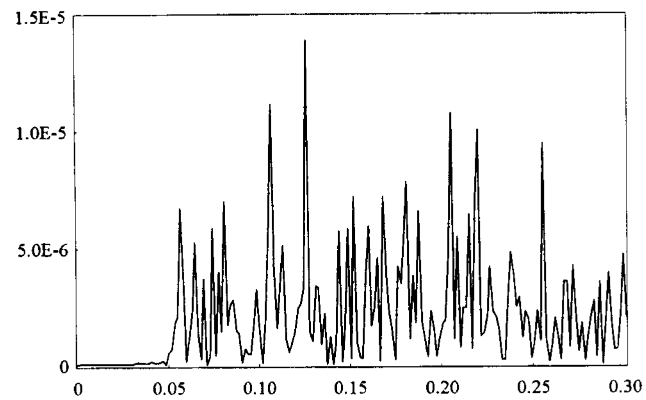

Figure 1.39 *White dwarf:* $L = 100$, *periodogram of residuals.*

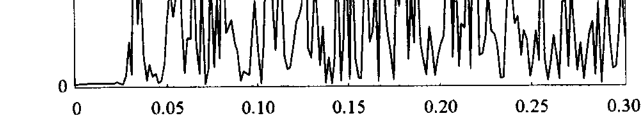

Figure 1.40 *White dwarf:* $L = 200$, *periodogram of residuals.*

For instance, in Section 1.3.2, the smoothing of the series 'White dwarf' has been done with $L = 100$, with the result of the smoothing being described by the leading 11 eigentriples. In the periodogram of the residuals (see Fig. 1.39) we can see that for this window length the powers of the frequencies in the interval $[0, 0.05]$ are practically zero.

If we take $L = 200$ and 16 leading eigentriples for the smoothing, then this frequency interval is reduced to $[0, 0.03]$ (see Fig. 1.40). At the same time, for $L = 10$ and two leading eigentriples, the result of smoothing contains the frequencies from the interval $[0, 0.09]$.

Visual inspection shows that all smoothing results look similar. Also, their eigenvalue shares are equal to $95.9\% \pm 0.1\%$. Certainly, this effect can be explained by the specificity of the series: its frequency power is highly concentrated in the narrow low-frequency region.

More difficult problems with smoothing occur when powers of low and high frequencies are not separable from each other by their values.

(c) Window length for periodicities

The problem of choosing the window length L for extraction of a periodic component $F^{(1)}$ out of the sum $F = F^{(1)} + F^{(2)}$ has certain peculiarities related to the correspondence between the window length and the period. In general, these peculiarities are the same for the pure harmonics and for complex periodicities, and even for modulated periodicities. Thus, we do not consider these cases separately.

1. For the problem of extraction of a periodic component with period T, it is natural to measure the length of the series in terms of the number of periods. Specifically, if $F^{(1)}$ is asymptotically separable from $F^{(2)}$, then to achieve the separation we must have, as a rule, the length of the series N such that the ratio N/T is at least several units.
2. For relatively short series, it is preferable to take into account the conditions for pure (nonasymptotic) separability (see Section 1.5); if one knows that the time series has a periodic component with an integer period T (for example, if this component is a seasonal component), then it is better to take the window length L proportional to that period. Note that from the theoretical viewpoint, $N-1$ must also be proportional to T.
3. In the case of a long series, the demand that L/T and $(N-1)/T$ be integers is not that important. In this case, it is recommended that the window length be chosen as large as possible (for instance, close to $N/2$, if the computer facilities allow one to do this). Nevertheless, even in the case of long series it is recommended that L be chosen such that L/T is an integer.
4. If the series $F^{(2)}$ contains a periodic component with period $T_1 \approx T$, then to extract $F^{(1)}$ we generally need a larger window length than for the case when such a component is absent (see Section 6.1.2 for the theory).
5. Since two harmonic components with equal amplitudes produce equal singular values, asymptotically, when L and K tend to infinity, a large window length can cause a lack of strong separability and therefore a mixing up of the components.

   If in addition the frequencies of the two harmonics are (almost) equal, then a contradiction between the demands for the weak and strong separability can occur; close frequencies demand large window lengths, but the large window lengths lead to approximately equal singular values.

To demonstrate the effect of divisibility of L by T, let us return to the 'Eggs' example (Section 1.3.3). Figs. 1.41 and 1.42 depict the matrices of w-correlations for the full decomposition of the series with $L = 12$ and $L = 18$. It is clearly seen that for $L = 12$ the matrix is essentially diagonal, which means that the eigentriples related to the trend and different seasonal harmonics are almost w-uncorrelated. This means that the choice $L = 12$ allows us to extract all the harmonic components of the series.

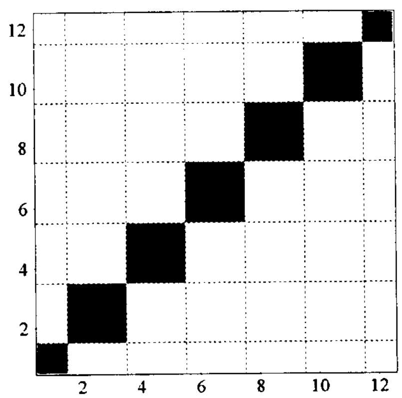

Figure 1.41 *Eggs:* $L = 12$, w*-correlations.*

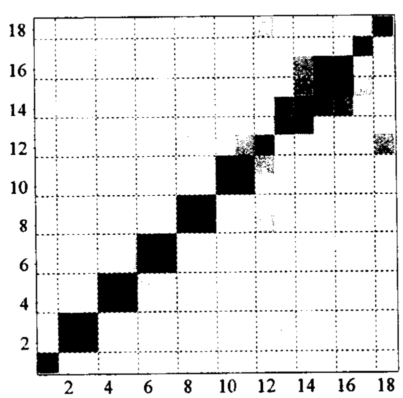

Figure 1.42 *Eggs:* $L = 18$, w*-correlations.*

For $L = 18$ (that is, when the period 12 does not divide L), only the leading seasonality harmonics can be extracted in a proper way; the other components have relatively large w-correlations.

The choice $L = 13$ would give results that are slightly worse than for $L = 12$, but much better than for $L = 18$. This confirms the robustness of the method with respect to small variations in L.

(d) Refined structure

In doing a simultaneous extraction of different components from the whole series, all the aspects discussed above should be taken into account. Thus, in basically all the examples of Section 1.3, where the periodicities were the main interest, the window length was a multiple of the periods. At the same time, if in addition the trends were to be extracted, L was reasonably large.

For the short series and/or series with a complex structure, these simple recommendations may not suffice and the choice of the window length becomes a more difficult problem.

For instance, in the example 'War' (Section 1.3.7), the choice $L = 18$ is dictated by the specific amplitude modulation of the harmonic components of the series, which is reflected in the shape of the trend. When the window length is reduced to 12, the amplitude-modulated harmonics are mixed up with the trend (the effect of the small window length). On the other hand, the decompositions with $L = 24$ and $L = 36$ lead to a more detailed decomposition of the sum of the trend and the annual periodicity (six eigentriples instead of four for $L = 12$), where the components are again mixed up (the effect of a too large L).

Note that if we wish to solve the problem of extracting the sum of the trend and the annual periodicity, then the choice of $L = 36$ is preferable to $L = 18$.

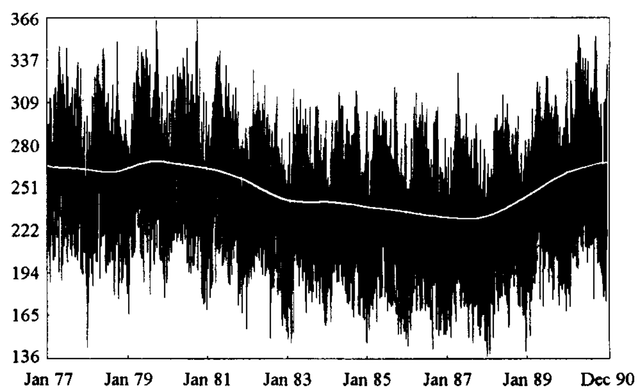

Figure 1.43 *Births: trend.*

To demonstrate the influence of the window length on the result of the decomposition, let us consider another more complex example, namely 'Births' (Section 1.3.4).

In the series 'Births' (daily data for about 14 years, $N = 5113$) there is a one-week periodicity ($T_1 = 7$) and an annual periodicity ($T_2 = 365$). Since $T_2 \gg T_1$, it is natural to take the window length as a multiple of T_2.

The choice $L = T_2$, as was shown in Section 1.3.4, guarantees the simultaneous extraction of both weekly and annual periodicities. Moreover, this window length allows also the extraction of the trend of the series (see Fig. 1.43) using the single leading eigentriple. Note that these results are essentially the same as for the cases $L = 364$ and $L = 366$.

At the same time, if we would choose $L = 3T_2 = 1095$, then the components of the trend will be mixed up with the components of the annual and half-year periodicities; this is a consequence of the complex shape of the trend and the closeness of the corresponding eigenvalues. Thus, large values of the window length lead to violation of strong separability.

If the problem of separation of the trend from the annual periodicity is not important, then values of L larger than 365 work well. If the window length is large, we can separate the global tendency of the series (trend + annual periodicity) from the weekly periodicity + noise even better than for $L = 365$ (for $L = 1095$ this component is described by several dozen eigentriples rather than by 5 eigentriples for $L = 365$). In this case, the weekly periodicity itself is perfectly separable from the noise as well.

Note also that if we were to take a small (in comparison to the annual period) window length (for example, $L = 28$), then the global behaviour of the series would be described by just one leading eigentriple. The weekly periodicity could also be separated, but a little worse than for large L.

In even more complex cases, better results are often achieved by the application of Sequential SSA, which after extraction of a component with a certain L requires a repeated application of SSA to the residuals with different L. An example of sequential SSA is described in Section 1.7.3.

(e) Hints

If we are not using the knowledge about the subject area and the series, then the only source of information helping in the window length selection is the series itself. Often the shape of the graph of the series (leading, for instance, to a visual identification of either a trend or a strong harmonic) is an effective indicator. At the same time, it is impossible to describe all possible manipulations with the series that may help in the selection of L, especially if we bear in mind that the corresponding algorithms should be fast (faster than the numerical calculation of the SVD of a large matrix) and should use the specifics of the problem.

We consider just one way of getting recommendations for selecting the window length, namely, the method based on the periodogram analysis of the original series and parts of the series.

1. If the resolution of the periodogram of the series F is good enough (that is, the series is sufficiently long), then the periodogram can help in determining the periods of the harmonic components of the series and thus in selecting the window length for their separation. Moreover, the presence of distinct powerful frequency ranges in the periodogram indicates possible natural components of the series related to these frequency ranges.

2. One of the sufficient conditions for approximate weak separability of two series is the smallness of all the spectral correlations for all the subseries of length L (and also $K = N - L + 1$) of these two series (see Section 1.5.3). Assume that:

   - periodograms of all the subseries of length L and K of the series F have the same structure,
   - this structure is characterized by the presence of distinct and distant powerful frequency ranges.

   In this case, the choice of a window length equal to L would most probably lead to a splitting of the series F into the components corresponding to these frequency ranges. This suggests that a preliminary periodogram analysis of at least several subseries might be useful.

A control of the correct choice of the window length is made at the grouping stage; the possibility of a successful grouping of the eigentriples means that the window length has been properly selected.

1.7 Supplementary SSA techniques

Supplementary SSA techniques may often improve Basic SSA for many specific classes of time series and for series of a complex structure. In this section, we consider several classes of this kind and describe the corresponding techniques. More precisely, we deal with the following series and problems:

1. The time series is oscillating around a linear function, and we want to extract this linear function.
2. The time series is stationary-like, and we want to extract several harmonic components from the series.
3. The time series has a complex structure (for example, its trend has a complex form, or several of its harmonic components have almost equal amplitudes), and therefore for any window length a mixing of the components of interest occurs.
4. The time series is an amplitude-modulated harmonic and we require its envelope.

The last problem is rather specific. Its solution is based on the simple idea that squared amplitude-modulated harmonic is a sum of low and high-frequency series that can be separated by Basic SSA.

The technique for the first two problems is in a way similar; having information about the time series, we use a certain decomposition of the trajectory matrix, which is different from the straightforward SVD and adapted to the series structure.

A lack of strong separability (in the presence of weak separability) is one of the main difficulties in Basic SSA. One of possible ways to overcome this difficulty is to enlarge the singular value of a series component of interest by adding a series similar to this component.

Alternatively, we can use Sequential SSA. This means that we extract some components of the initial series by the standard Basic SSA and then extract other components of interest from the residuals.

Suppose, for example, that the trend of the series has a complex form. If we choose a large window length L, then certain trend components would be mixed with other components of the series. For small L, we would extract the trend but obtain mixing of the other series components which are to be extracted.

A way to solve the problem, and this is a typical application of Sequential SSA, is to choose a relatively small L to extract the trend or its part, and then use a large window length to separate components of interest in the residual series.

Let us describe these approaches in detail and illustrate them with examples.

1.7.1 Centring in SSA

Consider the following extension of Basic SSA. Assume that we have selected the window length L. For $K = N - L + 1$, consider a matrix $\mathbf{A}$ of dimension $L \times K$ and pass from the trajectory matrix $\mathbf{X}$ of the series F to the matrix $\mathbf{X}^* = \mathbf{X} - \mathbf{A}$. Let $\mathbf{S}^* = \mathbf{X}^*(\mathbf{X}^*)^{\mathrm{T}}$, and denote by λ_i and U_i $(i = 1, \ldots, d)$ the nonzero eigenvalues and the corresponding orthonormal eigenvectors of the matrix $\mathbf{S}^*$. Setting $V_i = (\mathbf{X}^*)^{\mathrm{T}} U_i / \sqrt{\lambda_i}$ we obtain the decomposition

$$\mathbf{X} = \mathbf{A} + \sum_{i=1}^{d} \mathbf{X}_i^* \tag{1.30}$$

with $\mathbf{X}_i^* = \sqrt{\lambda_i} U_i V_i^{\mathrm{T}}$, instead of the standard SVD (1.2). At the grouping stage the matrix $\mathbf{A}$ will enter one of the resultant matrices as an addend. In particular, it can produce a separate time series component after the application of diagonal averaging.

If the matrix $\mathbf{A}$ is orthogonal to all the $\mathbf{X}_i^*$ (see Section 4.4), then the matrix decomposition (1.30) yields the decomposition

$$||\mathbf{X}||_{\mathcal{M}}^2 = ||\mathbf{A}||_{\mathcal{M}}^2 + \sum_{i=1}^{d} ||\mathbf{X}_i^*||_{\mathcal{M}}^2$$

of the squared norms of the corresponding matrices.

Here we consider two ways of choosing the matrix $\mathbf{A}$, thoroughly investigated in Sections 4.4 and 6.3. We follow the terminology and results from these sections.

(a) Single and Double centring

Single centring is the row centring of the trajectory matrix. Here

$$\mathbf{A} = \mathcal{A}(\mathbf{X}) = [\mathcal{E}_1(\mathbf{X}) : \ldots : \mathcal{E}_1(\mathbf{X})],$$

where each ith component of the vector $\mathcal{E}_1(\mathbf{X})$ $(i = 1, \ldots, L)$ is equal to the average of the ith components of the lagged vectors $X_1, \ldots, X_K$.

Thus, under Single centring we consider the space $\mathrm{span}(X_1^{(c)}, \ldots, X_K^{(c)})$ with $X_i^{(c)} = X_i - \mathcal{E}_1(\mathbf{X})$ rather than $\mathrm{span}(X_1, \ldots, X_K)$. In other words, we shift the origin to the centre of gravity of the lagged vectors and then use the SVD of the obtained matrix. Of course, Single centring is a standard procedure in the principal component analysis of multidimensional data.

For the *Double centring*, SVD is applied to the matrix, computed from the trajectory matrix by subtracting from each of its elements the corresponding row and column averages and by adding the total matrix average. In other words, in this case we have

$$\mathbf{A} = \mathcal{A}(\mathbf{X}) + \mathcal{B}(\mathbf{X}) \tag{1.31}$$

with $\mathcal{B}(\mathbf{X}) = [\mathcal{E}_{12}(\mathbf{X}) : \ldots : \mathcal{E}_{12}(\mathbf{X})]^{\mathrm{T}}$, where the jth component of the vector $\mathcal{E}_{12}(\mathbf{X})$ $(j = 1, \ldots, K)$ is equal to the average of all the components of the vector $X_j^{(c)}$.

Under Single centring the addend $\mathbf{A}$ has the same form as the other components of the decomposition (1.30), provided we have included the normalized vector of averages $U_{0(1)} = \mathcal{E}_1(\mathbf{X})/||\mathcal{E}_1(\mathbf{X})||$ in the list of eigenvectors U_i. Indeed, $\mathbf{A} = U_{0(1)} Z_{0(1)}^{\mathrm{T}}$ with $Z_{0(1)} = ||\mathcal{E}_1(\mathbf{X})|| \mathbf{1}_{\mathrm{K}}$. (Each component of the vector $\mathbf{1}_{\mathrm{K}} \in \mathbf{R}^K$ is equal to 1.)

In the Double centring case, we add one more vector to the list of eigenvectors, the vector $U_{0(2)} = \mathbf{1}_{\mathrm{L}}/\sqrt{L}$. Here

$$\mathbf{A} = U_{0(1)} Z_{0(1)}^{\mathrm{T}} + U_{0(2)} Z_{0(2)}^{\mathrm{T}}$$

with $Z_{0(2)} = \sqrt{L}\mathcal{E}_{12}(\mathbf{X})$. We set

$$\lambda_{0(1)} = ||Z_{0(1)}|| = ||\mathcal{E}_1(\mathbf{X})||\sqrt{K} \text{ and } \lambda_{0(2)} = ||Z_{0(2)}|| = ||\mathcal{E}_{12}(\mathbf{X})||\sqrt{L}.$$

Moreover, let $V_{0(1)} = Z_{0(1)}/\sqrt{\lambda_{0(1)}}$ and $V_{0(2)} = Z_{0(2)}/\sqrt{\lambda_{0(2)}}$. Then we call $\left(U_{0(i)}), V_{0(i)}, \lambda_{0(i)}\right)$ $(i = 1, 2)$ *the first and the second average triples.*

Since $\mathcal{A}(\mathbf{X})$ and $\mathcal{B}(\mathbf{X})$ are orthogonal to each other and to all the other decomposition components (see Section 4.4), we have for the Double centring

$$||\mathbf{X}||_{\mathcal{M}}^2 = \lambda_{0(1)} + \lambda_{0(2)} + \sum_{i=1}^{d} \lambda_i$$

(for the Single centring the term $\lambda_{0(2)}$ is omitted). Therefore, the shares of the average triples and the eigentriples are equal to

$$\lambda_{0(1)}/||\mathbf{X}||_{\mathcal{M}}^2, \quad \lambda_{0(2)}/||\mathbf{X}||_{\mathcal{M}}^2 \quad \text{and} \quad \lambda_i/||\mathbf{X}||_{\mathcal{M}}^2.$$

Note that Basic SSA does not use any centring. Nevertheless, Single centring can have some advantage if the series F can be expressed in the form $F = F^{(1)} + F^{(2)}$, where $F^{(1)}$ is a constant series and $F^{(2)}$ oscillates around zero.

Certainly, if the series length N is large enough, its additive constant component will undoubtedly be extracted by Basic SSA (as well as with the averaging of all the components of the series), but, for the short series, Single centring SSA can work better. Since the analogous, but much brighter, effects are produced by Double centring SSA, we do not consider any Single centring example here.

The effect of Double centring can be explained as follows. If the initial series is a linear one, $\mathbf{X}$ is its trajectory matrix and $\mathbf{A}$ is defined by (1.31), then $\mathbf{A} = \mathbf{X}$. Therefore, for $F = F^{(1)} + F^{(2)}$ with linear $F^{(1)}$, the matrix $\mathbf{A}$ contains the entire linear part of the series F. Theoretically, Double centring leads to the asymptotic extraction of the linear component of the series from rather general oscillatory residuals (see Section 6.3.2).

As usual, nonasymptotic effects occur as well. For fixed N and L, let us consider the series F which is the sum of a linear series $F^{(1)}$ and a pure harmonic $F^{(2)}$ with an integer period T. If the window length L and $K = N - L + 1$ divide T, then the matrix $\mathbf{A}$ defined by (1.31) coincides with the trajectory matrix of the series $F^{(1)}$. The residual matrix $\mathbf{X}^* = \mathbf{X} - \mathbf{A}$ corresponds to the trajectory matrix of the harmonic series $F^{(2)}$. Therefore, here we obtain the theoretically precise linear trend extraction (see Section 6.3.2 for the theory).

(b) Double centring and linear regression

Comparison of the extraction of a linear component of a series by Double centring SSA and by linear regression can be instructive. Note that these two methods have different origins and therefore can produce very different results.

As for linear regression, it is a formal procedure for the linear approximation by the least-squares method and gives a linear function of time for any series, even if the series does not have any linear tendency at all. By contrast, Double centring SSA gives us (usually, approximately) a linear component only if the strong linear tendency is really present. On the other hand, Double centring does not produce a precise linear function but only a pointwise approximation of it. Roughly speaking, linear regression estimates the coefficients of a linear function, while Double centring SSA estimates the values of a linear function at each point.

If the time series has a linear tendency and its length is rather large, then both methods produce similar results. The difference appears for a relatively short time series. For these series, the objective of the linear regression can be in disagreement with the problem of searching for a linear tendency of the series.

Let us illustrate these statements by two examples.

Example 1.5 *'Investment': long time series with a linear tendency*
The theory tells us (see Section 6.3.2) that Double centring SSA extracts (perhaps, approximately) the linear component of the series if the series oscillates near this linear component. Since linear regression automatically approximates any series by a linear function, a correspondence between the results of both methods would indicate that the linear function obtained by the regression method describes the

actual tendency of the series, and thus it is not merely the result of a formal procedure.

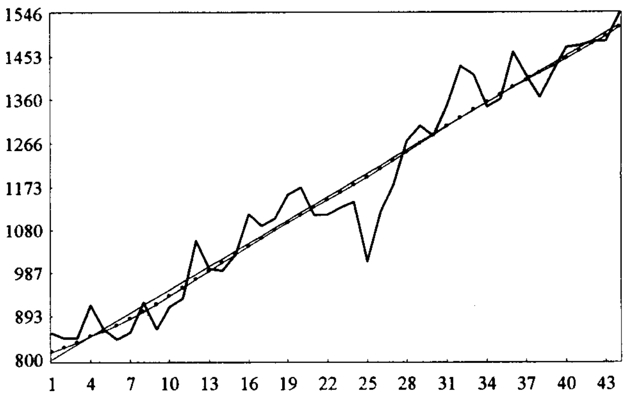

Figure 1.44 *Investment: linear regression and Double centring results.*

The example 'Investment' (the investment series of U.K. economic time series data, quarterly, successive observations, Prothero and Wallis, 1976), illustrates these considerations. The series is presented in Fig. 1.44, thick line.

Let us select the window length $L = 24$ in Double centring SSA, and take both average triples for the reconstruction. Then the reconstructed component (Fig. 1.44, thin line with black points) will resemble a linear function.

For comparison, the result of standard linear regression analysis (thin line) is placed on the same plot. Since both lines (SSA reconstruction curve and linear regression function) are very close, the general linear behaviour of the 'Investment' series can be considered as being confirmed. Note that the 'Investment' series can be regarded as a 'long' series since it oscillates rather rapidly around the regression line.

The second example demonstrates the difference between these two methods and shows the Double centring SSA capabilities for short series.

Example 1.6 *'Hotels': continuation of the extracted tendency*
The 'Hotels' series (hotel occupied room, monthly, from January 1963 to December 1976, O'Donovan, 1983) is a good example for discussing the difference between linear regression and Double centring SSA approaches to the extraction of a linear tendency in a series. Fig. 1.45 depicts the initial series and its linear regression approximation. Despite the fact that the series is not symmetric with respect to the regression line and has an increasing amplitude, the whole linear tendency seems to have been found in a proper way.

The second figure (Fig. 1.46) deals with the first 23 points of the series. Two lines intersect the plot of the series. The thin one is the linear regression line

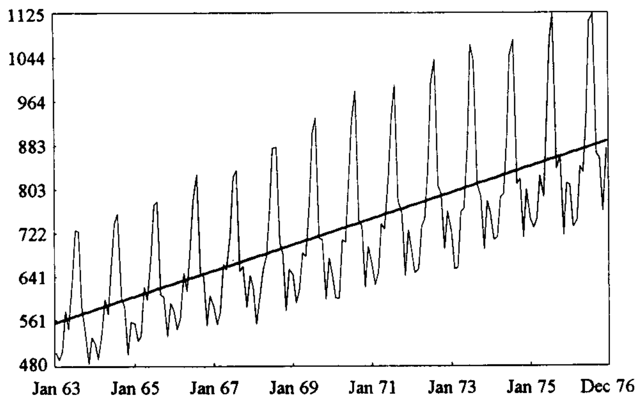

Figure 1.45 *Hotels: time series and its linear regression approximation.*

calculated from this subseries. The thick line is the reconstruction of the series produced by both average triples for the Double centring SSA with window length $L = 12$. This Double centring curve is almost linear but differs from that of the linear regression.

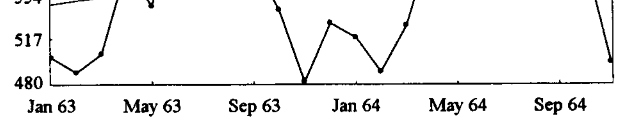

Figure 1.46 *Hotels: short interval. Regression and Double centring lines.*

Fig. 1.47 shows the continuation of both lines for the first 72 points of the 'Hotels' series. The upper linear function ($y = 543.6 + 3.2x$) is a continuation of the linear regression line of the Fig. 1.46. The lowest linear function ($y = 556.8 + 1.9x$) is the linear continuation of the Double centring line (Fig. 1.46). It is very close to the middle linear function ($y = 554.8 + 2.0x$) which is the part of the global linear regression line of Fig. 1.45.