

第二章 SSA基础

2.1 基础算法

2.1.1 算法描述

2.1.1.1 第一阶段：分解

2.1.1.2 第二阶段：重构

2.1.2 对SSA基础算法中四个步骤的分析

2.1.2.1 嵌入

2.1.2.2 SVD分解

2.1.2.3 分组

2.1.2.4 对角平均

2.2 SSA的潜力

2.2.1 提取趋势分量和平滑

2.2.1.1 不同解的趋势

2.3 时间序列模型和SSA的目标

2.3.1 SSA和时间序列模型

2.3.1.1 时间序列模型和周期图

2.3.1.2 时间序列名模型和秩

2.3.1.3 加法和乘法模型

2.3.1.4 非参数与参数模型

2.3.2 SSA任务的分类

2.3.3 时间序列成分的可分性

2.4 基础SSA算法参数的选取

2.4.1 一般问题

2.4.1.1 奇异向量的形式

2.4.1.2 预测被重构的成分的形状

2.4.1.3 特征值

2.4.1.4 基本重构成分和 w 相关矩阵

2.4.2 对于给定窗口长度的分组

2.4.2.1 基础分析

2.4.2.2 如何分组

2.4.2.3 如何检查重构结果

2.4.2.4 周期估计的方法

2.4.3 窗口长度

2.4.3.1 一般原则

2.4.3.2 趋势提取和平滑的窗口长度

2.4.3.3 提取周期成分的窗口长度

2.4.3.4 精致的结构

2.4.4 信号提取

2.4.4.1 信号提取的细节

2.4.4.2 信号的秩的估计方法

2.4.5 SSA成分的自动确定

2.4.5.1 基于 w 相关矩阵的分组

2.4.5.2 趋势的确定

2.4.5.3 谐波的确定

2.5 基础SSA算法的一些变体

2.5.1 预处理

2.5.2 在SSA中中心化

2.5.3 平稳序列和Toeplitz SSA

2.5.4 可分离性旋转：SSA-ICA

2.5.5 SSA序列

2.5.6 SSA实现

2.5.7 使用其他过程替换SVD分解

第二章 SSA基础 目录

2.1 基础算法 目录

2.1.1 算法描述 目录

考虑一个长度为 N 的时间序列 $\mathbb{X} = \mathbb{X}_N = (x_1, \dots, x_N)$ 。假设 $N > 2$ 且 \mathbb{X} 是一个非零序列。用 $L(1 < L < N)$ 表示窗口长度，并且 $K = N - L + 1$ 。下面将讲述SSA的基础算法，其中包含两个阶段：分解和重构。

2.1.1.1 第一阶段：分解 目录

第一步：嵌入

我们将原始序列映射为长度为 L 的 $K = N - L + 1$ 个滞后向量，称其为 L -滞后向量。

$$X_i = (x_i, \dots, x_{i+L-1})^T \quad (1 \leq i \leq K)$$

由 L -滞后向量组成 L -轨迹矩阵：

$$\mathbf{X} = [X_1 : \dots : X_K] = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_K \\ x_2 & x_3 & x_4 & \dots & x_{K+1} \\ x_3 & x_4 & x_5 & \dots & x_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \\ x_L & x_{L+1} & x_{L+2} & \dots & x_N \end{pmatrix} \quad (2.1)$$

滞后向量 X_i 是**轨迹矩阵** \mathbf{X} 的列。 \mathbf{X} 的行和列都是原序列的子序列。 \mathbf{X} 中的位于 (i, j) 的元素 $x_{ij} = x_{i+j-1}$ ，因此 \mathbf{X} 中反对角线的元素相等。（因此轨迹矩阵也被称称作**汉克尔矩阵**）公式(2.1)定义了一个从时间序列到轨迹矩阵的一对一的映射。

第二步：奇异值分解

在这一步，我们对 \mathbf{X} 进行SVD分解。令 $\mathbf{S} = \mathbf{X}\mathbf{X}^T$ 并且用 $\lambda_1, \dots, \lambda_L$ 表示 \mathbf{S} 的特征值，其中这些特征值按照从大到小的顺序排列。用 U_1, \dots, U_L 表示对应的特征向量。用 d 表示矩阵 \mathbf{X} 的秩，一般情况下，如果时间序列是真实世界中获取到的， $d = L^* = \min L, K$ 。记 $V_i = \mathbf{X}^T U_i / \sqrt{\lambda_i} (i = 1, \dots, d)$ 。轨迹矩阵 \mathbf{X} 可以被分解为：

$$\mathbf{X} = \mathbf{X}_1 + \dots + \mathbf{X}_d \quad (2.2)$$

其中 $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^T$ 。矩阵 \mathbf{X}_i 的秩为1，这样的矩阵被称作初等矩阵（elementary matrices），集合 $(\sqrt{\lambda_i}, U_i, V_i)$ 被称作三特征（eigentriple）简记作ET。

2.1.1.2 第二阶段：重构 目录

第三步：三特征分组

一旦得到(2.2)的分解结果，就把 d 个分解结果分配到 m 个不相交的子集中 I_1, \dots, I_m 。令 $I = \{i_1, \dots, i_p\}$ 。分组 I 对应的分组结果 $\mathbf{X}_I = \mathbf{X}_{i_1} + \dots + \mathbf{X}_{i_p}$ 。最终分组后的结果记为

$$\mathbf{X} = \mathbf{X}_{I_1} + \dots + \mathbf{X}_{I_m} \quad (2.3)$$

挑选集合 I_1, \dots, I_m 的过程称作三特征分组，如果 $m = d$ ，即 $I_j = \{j\}, j = 1, \dots, d$ ，则该分组成初等分组。

第四步：对角平均

在这一步，我们需要把矩阵 \mathbf{X}_{I_j} 还原为时间序列。令 \mathbf{Y} 表示 $L * K$ 的一个矩阵，记 $L^* = \min(L, K), K^* = \max(L, K), N = L + K - 1$ 。令 $y_{ij}^* = y_{ij}$ 如果 $L < K$ ，否则 $y_{ij}^* = y_{ji}$ 。使用以下公式将矩阵 \mathbf{Y} 还原为时间序列：

$$y_k = \begin{cases} \frac{1}{k} \sum_{m=1}^k y_{m,k-m+1}^* & \text{for } 1 \leq k < L^*, \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m,k-m+1}^* & \text{for } L^* \leq k \leq K^*, \\ \frac{1}{N-k+1} \sum_{m=k-K^*+1}^{N-K^*+1} y_{m,k-m+1}^* & \text{for } K^* < k \leq N. \end{cases} \quad (2.4)$$

该公式对应着求取矩阵反对角线元素的平均值。

对 \mathbf{X}_{I_k} 使用对角平均可以得到重构序列 $\tilde{\mathbf{X}}^{(k)} = (\tilde{x}_1^{(k)}, \dots, \tilde{x}_N^{(k)})$ 。最终，初始序列被分解为 m 成分：

$$x_n = \sum_{k=1}^m \tilde{x}_n^{(k)} \quad (n = 1, 2, \dots, N) \quad (2.5)$$

由初等分组得到的序列重构结果被称作初等重构序列 (elementary reconstructed series)。

Remark 2.1 基础SSA算法可以自然扩展到复数时间序列：唯一的区别在于矩阵的转置要替换为复数的共轭转置。

2.1.2 对SSA基础算法中四个步骤的分析 目录

SSA基础算法中的步骤需要一些说明。在这个部分我们简要讨论所涉及的程序的含义。

2.1.2.1 嵌入 目录

嵌入是一个把一维时间序列 $\mathbb{X} = (x_1, \dots, x_N)$ 使用向量 $\mathbf{X}_i = (x_i, \dots, x_{i+L-1})$ 转换为多维序列 $\mathbf{X}_1, \dots, \mathbf{X}_K$ 的过程，其中 $K = N - L + 1$ 。嵌入的参数是窗口长度 L 。注意轨迹矩阵(2.1)具有明显的对称性。 \mathbb{X}^T 相当于是窗口长度为 K 的多维序列。

对于动力系统专家来说，一种常用的技术是获得所有成对的滞后向量 \mathbf{X}_i 和 \mathbf{X}_j 之间的经验，然后计算时间序列相关的维度。该维数与产生时间序列的动力系统的吸引子的分形维数有关。需要注意的是，在这种方法中， L 必须相当小而 K 相当大。同样，在具有汉克尔矩阵结构的结构总最小二乘 (Structural Total Least Squares) 中，通常的做法是选择 $L = r + 1$ ，其中 r 是近似矩阵的猜测秩。

在SSA中，窗口长度 L 应该足够大。特别是， L 必须足够大，以便每个 L 滞后向量包含初始序列 $\mathbb{X} = (x_1, \dots, x_N)$ 的一个基础部分。 L 必须足够大时，才有可能将每个 L 滞后向量 \mathbf{X}_i 视为一个单独的序列。

2.1.2.2 SVD分解 目录

SVD分解可以使用不同的术语被描述，并且可以被用于不同的目的。让我们从SVD一般的特点开始讲解，这些特点对于理解SSA十分重要。

正如我们已经提到过的，SVD分解可以对任意的维度为 $L * K$ 的矩阵 $\mathbf{X} = [\mathbf{X}_1 : \dots : \mathbf{X}_K]$ 进行分解，分解形式如下：

$$\mathbf{X} = \sum_{i=1}^d \sqrt{\lambda_i} \mathbf{U}_i \mathbf{V}_i^T \quad (2.6)$$

其中 $\lambda_i (i = 1, \dots, L)$ 是矩阵 $\mathbf{S} = \mathbf{X}\mathbf{X}^T$ 的特征值，并且按照递减的顺序进行排列。 $d = \max\{i, \text{such that } \lambda_i > 0\} = \text{rank}(\mathbf{X})$ ， $\{\mathbf{U}_1, \dots, \mathbf{U}_d\}$ 是 \mathbf{S} 的政教特征向量，并且 $\mathbf{V}_i = \mathbf{X}^T \mathbf{U}_i / \sqrt{\lambda_i}$ 。

标准的SVD术语称 $\sqrt{\lambda_i}$ 为奇异值； \mathbf{U}_i 和 \mathbf{V}_i 为矩阵 \mathbf{X} 的左奇异向量和右奇异向量。如果我们定义 $\mathbf{X}_i = \sqrt{\lambda_i} \mathbf{U}_i \mathbf{V}_i^T$ ，那么公式(2.6)可以被写作公式(2.2)的形式。

如果所有特征向量的重数为1，那么展开式(2.2)是唯一确定的。否则，如果至少有一个特征值的重数大于1，那么在选择相应的特征向量时就有了自由度。我们应该假设以某种方式选择特征向量并且选择是固定的。

公式(2.6)说明SVD分解具有以下的对称性： $\mathbf{V}_1, \dots, \mathbf{V}_d$ 构成了矩阵 $\mathbf{X}^T \mathbf{X}$ 的特征正交系统。注意到轨迹矩阵的行和列都是原时间序列的子序列，因此左右奇异向量都具有时间结构，所以这些奇异向量也可以看做时间序列。

SVD分解具有很多特征，其中一个如下：在所有秩 $r < d$ 的矩阵 $\mathbf{X}^{(r)}$ 中，矩阵 $\sum_{i=1}^r \mathbf{X}_i$ 给出了对轨迹矩阵 \mathbf{X} 的最优估计，即 $\|\mathbf{X} - \mathbf{X}^{(r)}\|_F$ 最小。其中 F 表示矩阵的Frobenius范数，一个矩阵 \mathbf{Y} 的Frobenius范数为 $\|\mathbf{Y}\|_F = \sqrt{\langle \mathbf{Y}, \mathbf{Y} \rangle_F}$ 。其中两个矩阵 $\mathbf{Y} = \{y_{ij}\}_{i,j=1}^{q,s}$ 和 $\mathbf{Z} = \{z_{ij}\}_{i,j=1}^{q,s}$ 的内积定义为：

$$\langle \mathbf{Y}, \mathbf{Z} \rangle_F = \sum_{i,j=1}^{q,s} y_{ij} z_{ij}$$

对于向量而言，Frobenius范数等同于传统的欧几里得范数。

注意到 $\|\mathbf{X}\|_F^2 = \sum_{i=1}^d \lambda_i$ 并且 $\lambda_i = \|\mathbf{X}_i\|_F^2$ ，因此我们考虑将 $\lambda_i / \|\mathbf{X}\|_F^2$ 作为度量矩阵 \mathbf{X}_i 在整个轨迹矩阵 \mathbf{X} 中所占的比重。最终可以使用 $\sum_{i=1}^r \lambda_i / \|\mathbf{X}\|_F^2$ 作为轨迹矩阵的最优近似的特征指标。而且，如果 $\lambda_r \neq \lambda_{r+1}$ ，可以把 $\sum_{i=r+1}^d \lambda_i$ 作为轨迹矩阵与其最优估计之间的误差距离。

现在我们考虑把轨迹矩阵 \mathbf{X} 看做是一系列L延时的向量。用 $\mathcal{X}^{(L)} \subset \mathbb{R}^L$ 表示向量 X_1, \dots, X_K 组成的线性空间。我们把这个空间称作时间序列 \mathbb{X} 的L-轨迹空间。为了强调序列 \mathbb{X} 的作用，我们把这个空间记作 $\mathcal{X}^{(L)}(\mathbb{X})$ 。公式(2.6)表明 $\mathcal{U} = (U_1, \dots, U_d)$ 是这个轨迹空间的一组正交基向量。

令 $Z_i = \sqrt{\lambda_i} V_i, i = 1, \dots, d$ ，公式(2.6)可以被写作 $\mathbf{X} = \sum_{i=1}^d U_i Z_i^T$ ，并且对于延时向量 X_j 我们有 $X_j = \sum_{i=1}^d z_{ji} U_i$ ，其中 z_{ji} 是向量 Z_i 中的元素。这意味着向量 Z_i 是由向量 X_j 在基向量 \mathcal{U} 下的第 i 个成分组成的。

让我们考虑转置的轨迹矩阵 \mathbf{X}^T 。引入 $Y_i = \sqrt{\lambda_i} U_i$ ，我们可以得到展开式 $\mathbf{X}^T = \sum_{i=1}^d V_i Y_i^T$ ，该公式对应着K-延时向量在正交基 V_1, \dots, V_d 下的表示。至此，SVD分解引出了关于轨迹矩阵 \mathbf{X} 的两种几何描述。

上述关于SVD特性的描述可以用L延时向量的多变量几何语言进行描述，如下：令 $r < d$ ，那么在整个 r 维子空间 \mathcal{L}_r 中，由 U_1, \dots, U_r 张成的子空间可以最好得估计原空间。也就是说，最小化 $\sum_{i=1}^K \text{dist}^2(X_i, \mathcal{L}_r)$ 得到结果是 $\mathcal{L}_r = \text{span}(U_1, \dots, U_r)$ ，而 $\sum_{i=1}^r \lambda_i / \sum_{i=1}^d \lambda_i$ 是这些延时向量的 r 维最佳估计的特征。

SVD的另一个特征与特征向量 U_1, \dots, U_d 有关。具体地，第一个特征向量 U_1 决定了延时向量的投影变化最大的方向。后续的每一个特征向量都与之前的特征向量正交，并且延时向量在该方向上的投影变化也是最大的。因此，很自然地将第 i 个特征向量所表示的方向称作第 i 个主方向。注意到，矩阵 $\mathbf{X}_i = U_i Z_i^T$ 是延时向量在第 i 个主方向投影后构成的。

这种对于SVD分解的理解角度引出了以下的术语。我们称向量 U_i 为第 i 个主特征向量，向量 V_i 和 $Z_i = \sqrt{\lambda_i} V_i$ 被称作第 i 个音字向量和第 i 个主成分。

Remark2.2: SSA基础算法中所使用的SVD分解类似于经典的多变量分析中的主成分分析（PCA）和平稳时间序列分析中的KL变换。然而，SSA中的SVD方法利用了轨迹矩阵具有汉克尔结构的特点。事实上，轨迹矩阵的行和列都是原始序列的子序列并具有相同的时域意义，而PCA和KL变换并非如此。

Remark 2.3: 一般地，轨迹空间中的任何一组正交基 P_1, \dots, P_d 都可以替代通过SVD分解得到的正交基 U_1, \dots, U_d 。在这种情况下，公式(2.2)可以被替换为 $\mathbf{X}_i = P_i Q_i^T$ ，其中 $Q_i = \mathbf{X}^T P_i$ 。一种可以替换正交基的例子是，在拓普利兹SSA中的自相关矩阵中的正交基，见章节2.5.3。另一个例子可以在独立成分分析（ICA）和因子分析中看到，见章节2.5.4。

关于SVD进一步的讨论以及使用见章节2.5.7。

2.1.2.3 分组 目录

分组这一步骤是对公式(2.2)中的 $\mathbf{X}_{i,j}$ 进行分组分配。假设 $m = 2$, $I_1 = I = \{i_1, \dots, i_r\}$ 并且 $I_2 = \{1, \dots, d\} \setminus I$, 其中 $1 \leq i_1 < \dots < i_r \leq d$ 。

分组的目的是对时间序列的附加成分进行划分。假设时间序列 \mathbb{X} 是两个时间序列 $\mathbb{X}^{(1)}$ 和 $\mathbb{X}^{(2)}$ 的和；也就是说 $x_i = x_i^{(1)} + x_i^{(2)}, i = 1, \dots, N$ 。固定窗口长度 L 并且用 $\mathbf{X}, \mathbf{X}^{(1)}, \mathbf{X}^{(2)}$ 分别表示时间序列 $\mathbb{X}, \mathbb{X}^{(1)}, \mathbb{X}^{(2)}$ 的 L 的轨迹矩阵。考虑轨迹矩阵 \mathbb{X} 的SVD分解。如果存在一组下标 $I \subset \{1, \dots, d\}$, 使得 $\mathbf{X}^{(1)} = \sum_{i \in I} \mathbf{X}_i$ 并且 $\mathbf{X}^{(2)} = \sum_{i \notin I} \mathbf{X}_i$, 那么我们称 $\mathbb{X}^{(1)}$ 和 $\mathbb{X}^{(2)}$ 是（弱）可分的。在这个例子中， $\mathbb{X}^{(1)}$ 在整个时间序列中的贡献值为对应的特征值的比重 $\sum_{i \in I} \lambda_i / \sum_{i=1}^d \lambda_i$ 。

(这里由两个小段没有翻译，感觉不是很必要)

分组的目的是找到一组分组方案 I_1, \dots, I_m ，它们对应的轨迹矩阵 $\mathbf{X}_{I_1}, \dots, \mathbf{X}_{I_m}$ 具有汉克尔矩阵的结构。（主要注意公式(2.2)中对轨迹矩阵分解以后，分解后的成分可能不具有汉克尔矩阵的结构，而分组这一步骤正是为了让各个成分具有汉克尔矩阵的结构）

现在让我们从多变量几何的角度看分组这一步骤。令 $\mathbf{X} = [X_1 : \dots : X_K]$ 表示时间序列 \mathbb{X} 的轨迹矩阵，序列 $\mathbb{X}^{(1)}$ 和 $\mathbb{X}^{(2)}$ 是一组分解方式并且各自对应下标集合 I 和 $\{1, \dots, d\} \setminus I$ 。这意味着由基向量 U_1, \dots, U_d 张成的轨迹空间 $\mathcal{X}^{(L)}$ 被分为了两个部分，即 $\mathcal{X}^{(L,1)} = \text{span}(U_i, i \in I)$, $\mathcal{X}^{(L,2)} = \text{span}(U_i, i \notin I)$ 。两个序列的可分离性意味着矩阵 \mathbf{X}_I 的列是延时向量 X_1, \dots, X_K 在特征子空间 $\mathcal{X}^{(L,2)}$ 上的投影，恰好是序列 $\mathbb{X}^{(1)}$ 的轨迹矩阵。章节2.4会介绍如何把SVD分解的特征向量划分到不同的组中。

(有两个段落没有翻译，感觉没什么用)

2.1.2.4 对角平均

如果时间序列的成分可分，并且被划分到合适的组中，那么公式(2.3)中的矩阵都具备汉克尔矩阵的结构特点。因此我们马上就能得到公式(2.5)的分解结果：对于所有的 k 和 n , $\tilde{\mathcal{X}}_n^{(k)}$ 表示矩阵 \mathbf{X}_{I_k} 的所有反对角元素 $x_{ij}^{(k)}$ （例如 $i + k = n + 1$ ）。

然而，实际中可能不存在这样的情况，即反对角线的元素并不相等。因此需要一个将汉克尔矩阵还原回时间序列的方法。因此我们考虑对反对角线进行平均。矩阵 \mathbf{X}_{I_k} 对应的反对角线平均以后的值记作 $\tilde{\mathbf{X}}^{(k)}$ 。

为了方便，我们使用汉克尔矩阵的算子 \mathcal{H} 表示对角线平均这个过程。这个算子对任意维度的矩阵进行操作，假设矩阵 \mathbf{Y} ，维度为 $L * K$ 。矩阵 $\mathcal{H}\mathbf{Y}$ 的元素记作 \tilde{y}_{ij} ，则

$$A_s = \{(l, k) : l + k = s, 1 \leq l \leq L, 1 \leq k \leq K\}$$

$$\tilde{y}_{ij} = \sum_{(l,k) \in A_s} y_{lk} / |A_s|$$

其中 $|A_s|$ 表示该集合的元素数量。

汉克尔矩阵 $\mathcal{H}\mathbf{Y}$ 在某种意义上是最接近矩阵 \mathbf{Y} 的（定理6.3）。反过来，汉克尔矩阵 $\mathcal{H}\mathbf{Y}$ 通过将反对角线中的值与级数中的值相关联，可以唯一地确定时间序列。通过使用汉克尔算子，我们可以将公式(2.3)改写为：

$$\mathbf{X} = \tilde{\mathbf{X}}_{I_1} + \dots + \tilde{\mathbf{X}}_{I_m}$$

$$\tilde{\mathbf{X}}_{I_l} = \mathcal{H}\mathbf{X}_{I_l}$$

一个有意义的分组策略能够使得矩阵 \mathbf{X}_{I_k} 几乎是汉克尔矩阵，这对应着时间序列的近似可分性并且暗示着不同矩阵 $\tilde{\mathbf{X}}_{I_l}$ 之间的内积非常小。

注意到 \mathcal{H} 是一个线性算子，因此前边的分组方法在经过对角线平均以后仍然适用。

计算时间序列 $\tilde{X}^{(k)}$ 的过程被称作序列成分 $\tilde{X}^{(k)}$ 的重构过程。

2.2 SSA的潜力

在这一章，我们开始讨论SSA的一些主要功能，例如提取趋势分量，平滑，去噪等。

2.2.1 提取趋势分量和平滑

2.2.1.1 不同解的趋势

<https://covid19.who.int/data>

时间序列模型和SSA的目标

上一章节我们介绍了SSA在时间序列中提取趋势分量，周期分量，平滑，求取信号包络和去噪等方面的应用。在这一章节，我们将对这些相关的概念进行定义并进行讨论。这一章的主要任务仍然基于SSA的基础算法

2.3.1 SSA和时间序列模型

2.3.1.1 时间序列模型和周期图

理论上，SSA可以被应用于任意的时间序列。然而，实际上对于不同类型的时间序列，需要对于SSA进行不同的特殊处理。而且不同类型的时间序列对于参数的选择和待求的结果也都不同。因此本章开始回西安对不同的时间序列进行讲解，并讨论一些重要的概念。

由于SSA的目标是把一个可分解的时间序列分解为一组可加的分量，因此我们假设这个时间序列是由一些简单的时间序列构成的。这些简单的时间序列就是接下来我们要讨论的内容。需要注意的是，我们只讨论确定性时间序列，其中包括噪声序列。

(a)周期图(Periodogram)

对于时间序列 $\mathbb{X}_N = (x_1, \dots, x_N)$ ，可以用傅里叶级数分解为：

$$x_n = c_0 + \sum_{k=1}^{\lfloor N/2 \rfloor} (c_k \cos(2\pi nk/N) + s_k \sin(2\pi nk/N)) \quad (2.9)$$

其中 N 表示这个序列的长度， $1 \leq n \leq N$ ，并当 N 为偶数的时候， $s_{N/2} = 0$ 。 c_0 等同于这个时间序列的平均值，因此如果该序列被中心化，则 $c_0 = 0$ 。令 $A_k^2 = c_k^2 + s_k^2$ 。则公式(2.9)可以被写作：

$$x_n = c_0 + \sum_{k=1}^{\lfloor N/2 \rfloor} A_k \cos(2\pi nk/N + \varphi_k).$$

我们定义周期图为：

$$\Pi_x^N(k/N) = \begin{cases} c_0^2 & \text{for } k = 0 \\ (c_k^2 + s_k^2) / 2 & \text{for } 0 < k < N/2 \\ c_{N/2}^2 & \text{for } k = N/2 \end{cases} \quad (2.10)$$

公式(2.10)的归一化如下：

$$\|\mathbb{X}_N\|_F^2 / N = \sum_{k=0}^{\lfloor N/2 \rfloor} \Pi_x^N(k/N) \quad (2.11)$$

公式(2.11)表明公式(2.10)在点 k/N 的值描述了频率为 $\omega = k/N$ 的谐波成分在整个时间序列中所占的比重。

具有正功率的频率的集合被称作周期图的支撑点。如果某个周期图的支撑点属于某个区间 $[a, b]$ ，则这个区间称为时间序列的频率范围。理论上，序列的周期图是平稳序列的频谱测量的模拟。当这个序列渐进平稳时，对应的周期图也越接近真是的频谱。周期图对于描述一个任意的时间序列非常有用。例如，趋势分量可以被认为是一个平稳时间序列的低频子序列。

周期图分析的缺点在于它的分辨率很低。事实上，周期图不能分辨出频率间隔小于 $1/N$ 的频率成分。对于较短的序列，例如 $\{j/N, j = 0, \dots, \lfloor N/2 \rfloor\}$ ，其很难估计整个频率范围 $[0, 1/2]$ ，并且周期图难以揭示时间序列的所有成分的周期性结构。

(b)平稳时间序列

如果对于所有的非负整数 k ，一个无限的时间序列 $\mathbb{X}_\infty = (x_1, x_2, \dots, x_N, \dots)$ 具备以下特征，则称该时间序列是平稳的：

$$\frac{1}{N} \sum_{j=1}^N x_{j+k} x_{j+m} \xrightarrow{N \rightarrow \infty} R(k-m), \quad (2.12)$$

其中函数 $R()$ 被称作序列 \mathbb{X} 的协方差函数。也就是说如果一个无限时间序列的自相关函数只与时间间隔有关，那么该时间序列就是平稳的。后边，当我们讨论平稳性时，都假定 $\frac{1}{N} \sum_{j=1}^N x_{j+k} \rightarrow 0$ (as $N \rightarrow \infty$)，即原时间序列的均值为0。

协方差函数可以用频谱测量来表示，在这个过程中决定了平稳序列的一些特性。我们可以通过一个有限的时间序列 \mathbb{X}_N 来估计对应的无限序列 \mathbb{X}_∞ 的频谱密度。一个平稳的无限序列 \mathbb{X}_∞ 的频谱测量 m_x 可以被写作：

$$x_n \sim \sum_i a_k \cos(2\pi\omega_k n) + \sum_i b_k \sin(2\pi\omega_k n), \quad \omega_k \in (0, 1/2] \quad (2.13)$$

、
其中， $a_k = a(\omega_k)$, $b_k = b(\omega_k)$, $b(1/2) = 0$ ，并且 $\sum_k (a_k^2 + b_k^2)$ 收敛。（注意其中 $a(1/2) \neq 0$ 如果 ω_k 恰好是 $1/2$ 。）公式(2.13)意味着时间序列 \mathbb{X}_∞ 的频谱测量 m_x 集中在频率点 $\pm\omega_k$ ，其中 $\omega_k \in (0, 1/2)$ 并且对应的幅值为 $(a_k^2 + b_k^2) / 4$ 。点 $1/2$ 处的权重为 $a^2(1/2)$ 。

公式(2.13)所示的序列被认为是近似周期的。周期序列对应一个频谱测量 m_x ，其集中在点 $\pm j/T$ ($j = 1, \dots, \lfloor T/2 \rfloor$)，其中 T 为某些整数。公式(2.13)意味着这在这个信号的表示方法下，公式中的求和项的数量是有限的，并且所有的频率 ω_k 都是合理的。

非周期但是近似周期的序列称为准周期性。小盒子额序列的频谱测量是离散的，但是频率点并不会集中在任意一个 $\pm j/T$ 上。具有不合理频率 ω 的谐波序列 $x_n = \cos(2\pi\omega n)$ 就是一个准周期序列的例子。

非周期序列（或者称作混沌序列）的特点是频率测量没有atoms（这里我翻译为“基本频率”）。在这种情况下，通常假设存在频谱密度： $m_x(d\omega) = p_x(\omega)d\omega$ 。非周期模型通常被用于为噪声建模。如果一个时间序列的谱密度是常数，那么这个序列就被称作白噪声。需要注意，白噪声序列不一定是随机的。在许多情况下，现实生活中的平稳序列具有两个分量，即周期（或准周期）分量和噪声（或非周期）分量。在处理有限序列时，很难区分一个具有较大周期的周期序列和一个准周期序列。而且，在有限的时间隔上，非周期序列与具有宽频谱和小幅值的周期序列几乎是难以区分的。

(c) 周期序列的幅度调制

平稳性的定义是渐进的。这种渐进性质既有优点又有缺点。优点是可以通过严格的数学定义说明模型治理的主要概念。缺点是不可能仅使用有限的数据检验平稳性的假设。有许多非平稳序列，但是我们只考虑两类非平稳时间序列，并且从定性的层面上来进行说明。具体地，这两类非平稳序列分别是幅度调制的周期序列和趋势序列。选择这两类时间序列与SSA的实际意义和重要性有关。

在下一个子章节中我们会对趋势序列进行处理。这里我们讨论幅度调制的周期信号，其形式为 $x_n = A(n)y_n$ ，其中 y_n 为周期序列，并且 $A(n) \geq 0$ 。通常假设给定的幅度序列 $A(n)$ 的变化频率要远低于周期序列 y_n 的低频谐波频率。这类时间序列通常在经济领域可以见到，其中周期序列 y_n 与季节性有关，但是幅度调制取决于长期趋势。

让我使用周期图的方法来分析幅度调制的周期信号。

$$x_n = A(n) \cos(2\pi\omega n + \theta), \quad n = 1, \dots, N. \quad (2.14)$$

除非上述序列非常短，否则它的周期图在一个短的包含 ω 的频率区间。对于频率比较大并且接近的两个频率值 $\omega_1 \approx \omega_2$ ，对应的两个周期序列相加以后如下：

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

即两个周期信号相加可以表示为一个缓慢变化点额序列 $A(n)$ 与一个具有较高频率 $(\omega_1 + \omega_2)/2$ 的谐波序列相乘。如果时间序列的长度 N 太小，就很难看到序列 $A(n)$ 震荡的特性。

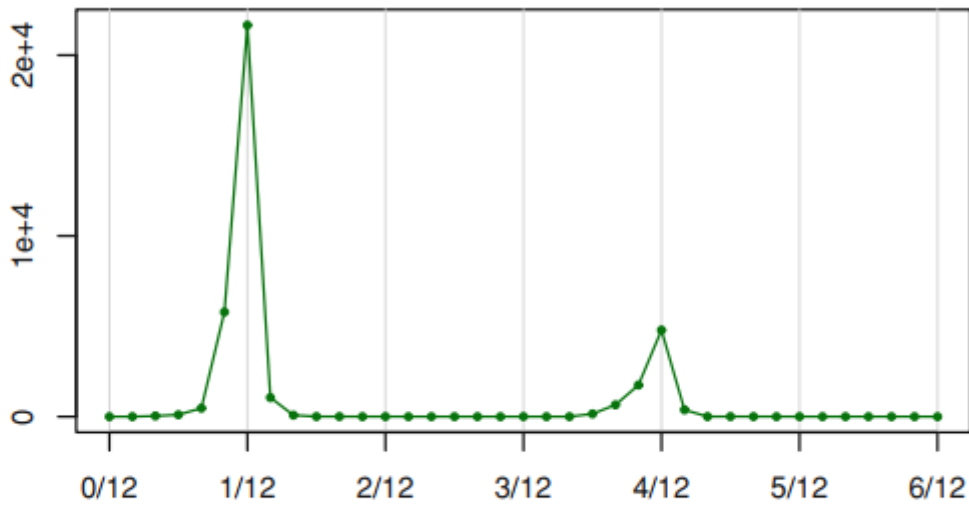


Fig. 2.10 War: periodogram of the main seasonality component

图2.10展示了一个时间序列的周期图，其中频谱由两个比较明显的频率支撑点对应着周期成分，而其他的频率支撑也有较高的幅值，这是由于受到了幅度调制的影响。

上述讨论暗示了完全相同的幅度调制可能是由完全不同的原因引起的。一种是真正的调制，另一种则是由于两个频率相近的周期序列相加形成了幅度调制的效应。

(有两个段落没有翻译，感觉没有必要)

(d) 趋势

(e) 时间序列的附加成分：案例研究

2.3.1.2 时间序列模型和秩 [目录](#)

在SSA的框架中，时间序列的结构与轨迹矩阵的秩相关，这里记 $d(L) = \text{rank} \mathbf{X}$ 。轨迹矩阵的秩就是其SVD分解中非零的特征值得个数。如果对于一些固定的 d ，当 L 足够大时，我们有 $d(L) = d$ ，那么我们称这个时间序列为一个有限秩时间序列（finite-rank time series）。对于这样的一个序列，我们有 $d(L) = \min(d, L)$ if $L \leq K$ 。对于任意有限长度的时间序列， $d \leq \min(L, K)$ 。如果 $d < \min(L, K)$ ，则时间序列就具有一个结构。 d 越小，序列的结构越简单，重构的时候分组也越简单。让我们讨论一些有限秩序列的例子，例如多项式乘积之和，指数和正弦曲线。

纯周期性。任何频率在 $(0, 0.5)$ 范围内的正弦序列的秩为2，频率为0.5的锯齿波序列的秩为1。因此任何接近公式2.13所描述的周期性的时间序列的秩都是有限的。非周期性序列的秩是无限的。注意到在SSA的框架下，正弦序列的简单性取决于观测周期的数量，而在 $(0, 0.5)$ 范围内的任何频率上，正弦序列的秩都是2。

调制周期。周期调制可能是序列的SSA结构变得复杂甚至破坏其结构。通常，对于任意调制，调制的正弦曲线不是有限秩的。公式2.14中定义的余弦调制是一个把秩从2增加到4的例子。唯一可能的不会改变序列的秩的调制方式就是指数调制，即 $A(n) = \exp(\alpha n) = \rho^n$ with $\rho = e^\alpha$ 。例如，指数阻尼正弦曲

线的秩与无阻尼正弦曲线的秩相同。这是SSA与经典的诸如傅里叶分析的本质区别，并且SSA允许在处理时间序列时不使用对数转换。还有就是这允许SSA处理形状在变化的周期序列。**这里我也不是很理解。**

趋势。趋势具有非结构性的行为，并且趋势对序列的非平稳性做出了主要贡献。一种典型的趋势可以通过一系列的有限秩进行估计。具有简单的SSA结构和较小的秩的缓慢变化的序列包括一个指数序列（秩为1），一个具有较大周期的正弦序列（秩为2），一个线性序列（秩为2）和高阶的多项式序列（rank>2）。

2.3.1.3 加法和乘法模型 [目录](#)

序列的加法模型就是一组成分的和；而乘法模型就是一组值为正数的成分的积。通过对数变换，任何乘法模型都可以被转换为加法模型。SSA使得时间序列可以被表示为一组成分的和。可能有人会认为SSA不能被用于通过乘法模型建模的时间序列。但是，在乘法模型中，一些时间序列可以被表示为加法模型并且无需额外的变换。例如，令 $x_n = t_n(1 + s_n)$ ，其中 (t_1, \dots, t_N) 是一个趋势并且 (s_1, \dots, s_N) 是一个幅值小于1的正弦序列（这是为了 $(1 + s_n)$ 可以为正数）。很容易得到 $x_n = t_n + t_n s_n$ ，也就是说，原始时间序列可以被看做是一个趋势序列和一个调制的正弦序列的和。因此，乘法模型可以被认为是一个具有调制震荡和噪声成分的加法模型。

因此SSA既可以被用于加法模型，又可以被用于乘法模型。如果对数变换可以使得信号更加简单（即具有更小的秩），那么对数模型可以增加分析的准确性或提升可分性。否则对数变换可能会带来负面的影响。例如对数变换会是的指数趋势的序列结构更加复杂。

2.3.1.4 非参数与参数模型 [目录](#)

为了使用基础SSA算法，我们无需对时间序列进行任何模型上的假设。因此，SSA属于一种非参数且模型自由的技术。然而，在可分性的假设下，可以基于SSA建立一个参数化的模型。接下来让我们来阐述这个想法。

令 $\mathbb{X}^{(1)}$ 为时间序列 $\mathbb{X} = \mathbb{X}^{(1)} + \mathbb{X}^{(2)}$ 的成分，其窗口长度为 L ，并且具有有限秩 $r < \min(L, K)$ 。用 $\mathbb{X}^{(1)}$ 的前 r 个特征向量进行重构，也就是说， $I_1 = [1, \dots, r]$ 。用 $\mathcal{X}^{(1)} = \text{span}(U_1, \dots, U_r)$ 表示对应的轨迹空间。如果第 L 个条件向量 $\mathbf{e}_L = (0, \dots, 0, 1)^T \notin \mathcal{X}^{(1)}$ ，那么 $\mathbb{X}^{(1)}$ 可以由一个线性循环关系（LRR）来表示：

$$x_n^{(1)} = \sum_{j=1}^r a_j x_{n-j}^{(1)}, \quad n = r+1, \dots, N$$

其中系数 a_j 可以由 r 维子空间 $\mathcal{X}^{(1)}$ 唯一确定。系数 a_j 决定了复数 μ_1, \dots, μ_r ，这些复数是LRR特征多项式的根，见章节3.2。（为了简单起见，我们假设所有的复数根都是不同的）。这时，时间序列 $x_n^{(1)}$ 可以被

表示为

$$x_n^{(1)} = \sum_{j=1} C_j \mu_j^n. \quad (2.15)$$

注意到原时间序列是一个实数序列，如果 $\mu_j \in \{\mu_1, \dots, \mu_r\}$ 并且 μ_j 是一个复数，那么存在共轭关系 $\mu_k = \mu_j^*$ 其中 $\mu_j \in \{\mu_1, \dots, \mu_r\}$ 。由于我们可以把 μ_j 写作 $\mu_j = \rho_j \exp(i2\pi\omega_j)$ ，因此集合 $\{\mu_j\}$ 包含了关于频率的所有信息。如果知道了 $\{\mu_j\}$ ，就可以得到系数 C_j 的值。

由于在实际情况中，在时间序列的成分之间没有准确的可分性，因此提出了许多方法来估计参数化模型的参数，见章节3.8获取更多信息。

2.3.2 SSA任务的分类 目录

基础SSA算法可以解决以下问题：平滑，趋势提取和震荡成分提取。SSA可以解决的最普遍的问题就是分析时间序列的结构；也就是把时间序列分割为一些“简单”并且“可解释”的成分以及噪声成分。接下来让我们分别讨论这些问题。

1. 趋势提取和平滑

在趋势提取和平滑之间没有明确的区分。有时平滑的结果就是趋势。这两者也没有明确的含义除非使用参数化模型进行定义。因此，大量的模型自由的方法都可以进行趋势提取和平滑。不过可以从定性的层面对趋势提取和平滑进行简单地区分，一般方法就是看仿真结果进行对比。注意到信号去噪与平滑非常相似。两者的区别在于对残差的假设不同：去噪要求残差中不包含任何信号的部分，但是平滑对应的残差可能包含信号中的高频周期成分。

2. 震荡成分的提取

这里的一般问题是识别和分离不构成趋势部分的振荡分量。在参数形式中（在零趋势、有限谐波数量和加性随机白噪声的假设下），这个问题在经典频谱分析理论中得到了广泛的研究。基础SSA是一种模型自由的方法。因此，基础SSA提取序列的单个谐波分量的结果通常不是纯谐波顺序。这与在实际中我们处理近似可分性而不是精确的可离性有关（参见第2.3.3节）。基础SSA不需要对谐波数量和谐波频率进行假设。然而，关于初试时间序列的辅助信息可以使问题更清晰，并且有助于SSA的参数选择。参考章节2.4.2.1。最后SSA允许对幅度调制的振荡成分进行分析。

3. 将时间序列分割为“简单”，“可解释”的成分和噪声

这个任务可以被认为是上面两个任务的组合；具体来说，就是趋势提取任务和周期分量提取任务。该任务的一个特定特征是，在完全分解中，残差应该仅由噪声组成。由于无模型技术往往会发现噪声中错误的可解释成分，强烈建议对于找到的每个信号分量有一个清晰的解释（基于时间序列数据的附加信息获得）。

2.3.3 时间序列成分的可分性 目录

正如我们上边讨论的，SSA的主要目标是将原时间序列分解为一些信号分量的和。这些信号分量包括趋势，周期或准周期分量。在这个过程中，信号分量的可分性扮演着很重要的角色。换句话说，只有当SSA信号分量是相互可分离的，SSA才能发挥自身的作用。

弱可分性和强可分性

考虑一个长度为 N 的时间序列 \mathbb{X} ，指定窗口长度为 L ，对应的 L -轨迹矩阵为 \mathbf{X} ，并且假设时间序列 \mathbb{X} 是由两个信号分量组成的，即 $\mathbb{X} = \mathbb{X}^{(1)} + \mathbb{X}^{(2)}$ 。在这种情况下，信号分量 $\mathbb{X}^{(1)}$ 和 $\mathbb{X}^{(2)}$ 的可分离性意味着我们可以将轨迹矩阵 \mathbf{X} 经过SVD分解以后的矩阵项分为两个不同的组，这样一来，每个组中的矩阵项相加以后就可以分别得到每个信号分量对应的轨迹矩阵 $\mathbf{X}^{(1)}$ 和 $\mathbf{X}^{(2)}$ 。

由于如果奇异值是多重的话，SVD分解的结果不唯一，因此可以考虑两种情况的可分离性。若可分离性是指存在一个轨迹矩阵 \mathbf{X} 的SVD分解，其求和的矩阵项可以被分为两组，并且每个组中的矩阵项相加以后就可以分别得到每个信号分量对应的轨迹矩阵。强可分离性是指对于轨迹矩阵的任意SVD分解都满足弱可分离性。

强可分离性要求SVD分解分组之后得到的 $\mathbf{X}^{(1)}$ 和 $\mathbf{X}^{(2)}$ 在各自特征值集合上没有交集。强可分离性暗示了若可分离性的存在，并在在实际中更可能出现若可分离性。对于SSA而言，缺乏强可分离性是一个严重的问题。在章节2.5.4中，我们会提出SSA-ICA，该方法可以在没有强可分离性的情况下提供可分性。在实际情况中，若可分离性比较容易研究并且有效。尽管准确的（弱）可分离性需要的条件相当严格，但是它们可以被扩展用于估计可分离性。

下面的条件等价于若可分离性的定义：

1. $\mathbb{X}^{(1)}$ 与 $\mathbb{X}^{(2)}$ 正交；或者说 $\mathbf{X}^{(1)}(\mathbf{X}^{(2)})^T = \mathbf{0}_{LL}$ 且 $(\mathbf{X}^{(1)})^T \mathbf{X}^{(2)} = \mathbf{0}_{KK}$ 。
2. 由轨迹矩阵 $\mathbf{X}^{(1)}$ 的列向量张成的子空间 $\mathcal{X}^{(L,1)}$ 正交于轨迹矩阵 $\mathbf{X}^{(2)}$ 张成的子空间 $\mathcal{X}^{(L,2)}$ ，并且对于它们的行向量张成的子空间必须具有相同的正交性。

若可分性的特点

令 $L^* = \min(L, K)$ 并且 $K^* = \max(L, K)$ ，引入权重

$$w_i = \begin{cases} i & \text{for } 0 \leq i < L^* \\ L^* & \text{for } L^* \leq i \leq K^* \\ N - i + 1 & \text{for } K^* < i \leq N \end{cases} \quad (2.16)$$

上述公式中的 w_i 等于元素 x_i 在轨迹矩阵 \mathbb{X} 中出现的次数。定义两个信号分量的内积为：

$$\left(\mathbb{X}^{(1)}, \mathbb{X}^{(2)} \right)_w \stackrel{\text{def}}{=} \sum_{i=1}^N w_i x_i^{(1)} x_i^{(2)} \quad (2.17)$$

如果 $(\mathbb{X}^{(1)}, \mathbb{X}^{(2)})_w = 0$ ，称信号分量 $\mathbb{X}^{(1)}$ 和 $\mathbb{X}^{(2)}$ w-正交。可分性的条件意味着w-正交。为了衡量两个信号分量的近似可分离度，我们引入w-相关（w-correlation）：

$$\rho^{(w)}(\mathbb{X}^{(1)}, \mathbb{X}^{(2)}) \stackrel{\text{def}}{=} \frac{(\mathbb{X}^{(1)}, \mathbb{X}^{(2)})_w}{\|\mathbb{X}^{(1)}\|_w \|\mathbb{X}^{(2)}\|_w} \quad (2.18)$$

如果 $\rho^{(w)}(\mathbb{X}^{(1)}, \mathbb{X}^{(2)}) \stackrel{\text{def}}{\approx} 0$ ，我们可以自然地认为这两个信号分量是近似可分的。注意到窗口长度 L 参与到了w-正交和w-相关的定义中。

另一种可分性的定量度量与信号分量的频率结构有关。对于弱可分离性，两个信号分量的周期图没有交集是充分但非必要条件。如果周期图的支撑点交集在某种意义上很小，那么就近似可分离。请注意，频率的可分离性相当于弱平稳时间序列的可分性。

可分的时间序列

尽管对于有限秩的时间序列，有很多准确可分的结果，但是准确的可分离性只是纯粹的理论结果。实际上，近似可分离性才是经常发生的。在实际中，趋势，振荡和噪声分量是近似可分的对于足够长的时间序列和窗口长度是非常重要的。

为了解释可分离性的概念，我们考虑一个正弦信号的例子。

$$x_n^{(1)} = A_1 \cos(2\pi n\omega_1 + \varphi_1), \quad x_n^{(2)} = A_2 \cos(2\pi n\omega_2 + \varphi_2) \quad (2.19)$$

其中， $n = 1, \dots, N, 0 < \omega_i < 0.5$ 并且 $\omega_1 \neq \omega_2$ ，令窗口长度 $L \leq N/2$ 并且 $K = N - L + 1$ 。这些时间序列是弱可分离的，如果 $L\omega_i$ 和 $K\omega_i$ 是整数，即 L 和 K 是周期的整倍数。如果 $A_1 \neq A_2$ ，这两个时间序列是强可分离的。这是因为幅值会影响特征值的大小。

对于较大的 N 和 L ，如果 $\omega_1 \neq \omega_2$ ，这两个正弦序列是近似弱可分离的，上述 L 和 K 是周期的整倍数这个条件不是必须的。尽管这个条件可以提高可分离性。可分离性的质量（会影响重构的准确性）取决于 $|\omega_1 - \omega_2|$ 的幅度。频率越接近，就需要更长的时间序列来获取足够的可分离性。

在近似弱可分离的条件下，如果 A_1 和 A_2 太接近，会导致强可分离性下降。需要注意，正弦曲线可分离性的频率解释是足够的，因为对于较大的 L ，正弦曲线周期图上的泄漏很小。

2.4 基础SSA算法参数的选取 目录

在这一章，我们讨论基础SSA算法的参数选取以及要遵守的原则。在SSA中有两个参数，一个是窗口长度，另一个是分组的方法。根据章节2.3.1.1和2.3.2，我们假设时间序列是一组信号分量的和，包括：缓慢变化的趋势，不同的振荡分量和一个噪声分量。

2.4.1 一般问题 目录

2.4.1.1 奇异向量的形式 目录

我们从一些简单时间序列的特征三元组的理论结果谈起。

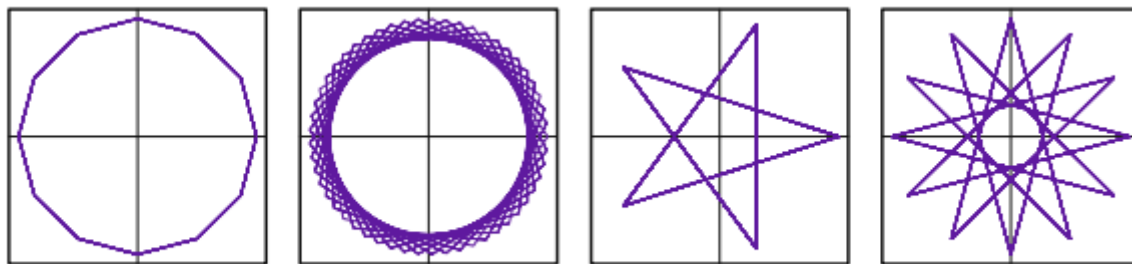


Fig. 2.18 Scatterplots of sines/cosines

振荡：指数余弦序列

考虑时间序列

$$x_n = Ae^{\alpha n} \cos(2\pi\omega n + \varphi) \quad (2.20)$$

其中 $\omega \in (0, 1/2]$ 并且 $\phi \in [0, 2\pi)$ 。依赖于这些参数的值，指数余弦序列产生以下的非零特征三元组。

1. 频率 $\omega \in (0, 1/2)$ 的指数调制谐波序列。
如果 $\omega \in (0, 1/2)$ ，那么任何 L 和 N ，轨迹矩阵的SVD分解由两个非零项。这两个特征向量与公式2.20具有相同的形式并且 ω 和 α 的值相同。
2. 指数调制锯齿波曲线($\omega = 1/2$)。
如果 $\sin(\phi) \neq 0$ ，那么 x_n 与 $(-e^\alpha)^n$ 成正比例。如果 $\alpha = 0$ ，那么 $x_n = A(-1)^n = A\cos(\pi n)$ 。在这种情况下，对于任何 L ，对应的SVD分解只有一项。奇异向量和原始序列具有相同的形式。

令 $\omega \neq 1/2$ 并且 $\alpha = 0$ 。那么通过公式2.20我们可以得到纯谐波序列。这会产生一个二阶的SVD奇异向量，并且这些奇异向量都具有相同的谐波形式。

为了确定性，让我们考虑左奇异向量（即特征向量）并假设一个理想情况，其中 $L\omega$ 是整数。在这种情况下，特征向量具有正弦和余弦序列的形式并且具有相同频率和相位。

图2.18描绘了四对正弦/余弦序列的成对散点图（零相位），他们都具有相同的幅度，频率分别为 $1/12$ 、 $10/53$ 、 $2/5$ 和 $5/12$ 。可以看到，所有的点都分布在一个单位圆上。如果周期 T 是一个整数，那么这些点是规则 T 顶点多边形的顶点。对于有理频率 $\omega = q/p < 1/2$ 相对素数 p 和 q ，这些点是规则 p 顶点多边形的定点。

趋势：指数序列和多项式序列

1. 指数序列 $x_n = e^{\alpha n}$

对于任意 N 和窗口长度 L ，指数序列的轨迹矩阵只有一个特征三元组。特征三元组里的奇异向量都是指数形式的，并且具有相同的参数 α 。

2. 一个一般的多项式序列。

考虑一个度为 m 的多项式序列：

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

对于这个序列，对应的SVD分解的阶数为 $m + 1$ 并且所有的奇异向量都具有多项式序列的形式并且度不超过 m 。

3. 线性序列

对于一个线性序列 $x_n = an + b$, $a \neq 0$ ，对于任意的 N 和 L 。L-轨迹矩阵的SVD分解有两个非零项组成。对应的奇异向量也是线性的。

请注意，对于任何长度 N 和窗口长度的序列，指数余弦和线性序列（除了具有不同速率的两个指数序列的总和之外）是唯一在其轨迹矩阵的 SVD 中最多具有两个非零项的序列 并且 $L \geq 2$ 。这一事实有助于将它们的 SSA 识别为更复杂系列的组件。

2.4.1.2 预测被重构的成分的形状 目录

特征三元组的形状有助于帮助我们预测将要基于这些特征三元组进行重构的信号分量的形状。

1. 如果我们只是用一个特征三元组进行信号分量的重构，并且这个特征三元组的两个奇异向量具有相似的形式，那么被重构的信号分量将具有近似相同的形式。

这意味着当处理一个奇异特征三元组时，我们经常可以预测到对应的信号分量的行为模式。例如，如果一个特征三元组的两个奇异向量都类似于线性序列，那么对应的信号分量几乎也是线性的。如果奇异向量是指数形式的，那么趋势分量也具有相似的形状。类似谐波形式的奇异向量会产生类似谐波的分量。这个一般性规律也可以用于其他特性的时间序列，包括单调性（单调的奇异向量产生的信号分量也是单调的）。

2. 如果 $L \ll K$ ，那么在被选中的特征三元组中的因子向量比奇异向量更像信号分量。

最终我们可以只根据因子向量大致预测对应的特征三元组重构得到的信号分量的形式。

3. 如果我们在一些多个特征三元组的帮助下重构时间序列，并且他们的奇异向量的周期图在频率范围 $[a, b]$ 中具有相同的频率支撑点，那么重构序列的频谱功率主要在频率范围 $[a, b]$ 。这个特点与第一个类似，但是该特点重点在于使用了多个特征三元组，并根据傅立叶展开式表示。

2.4.1.3 特征值 目录

让我们列举几个轨迹矩阵奇异向量的特点：

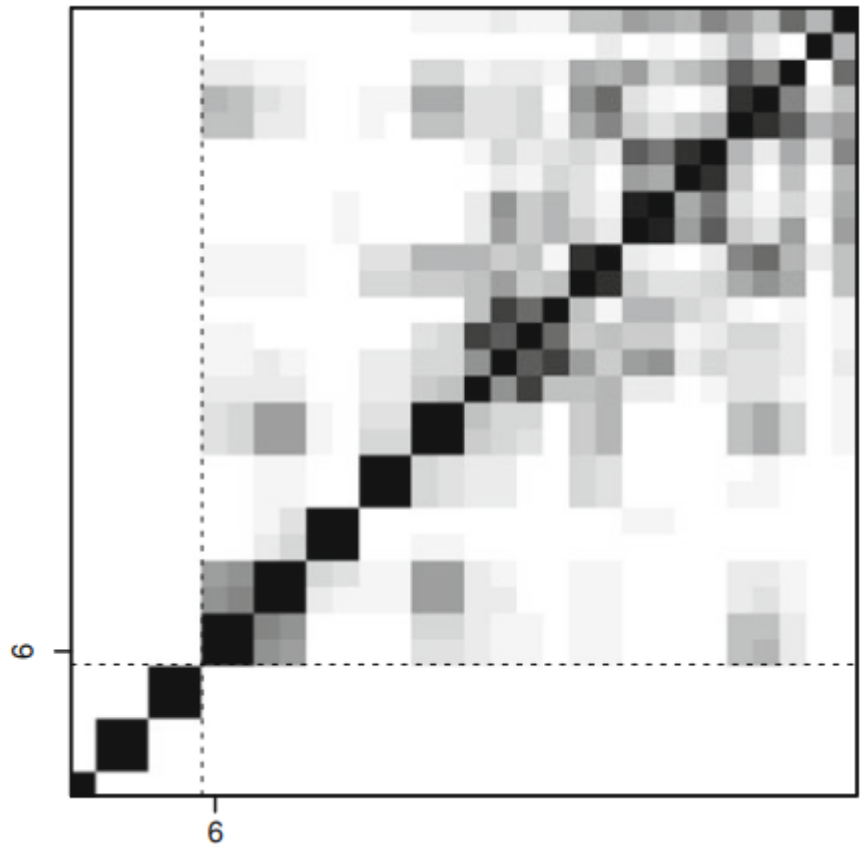
1. 特征三元组的奇异值越大，对应的信号分量所占的权重越大。粗略地说，这个权重可以被认为与奇异值成正比。
2. 通过与主成分分析 (PCA) 类比，主特征值的占比反映了相应特征三元组逼近原始序列的质量。但是，基础 SSA 和 PCA 之间存在显著差异，因为 PCA 执行变量的中心化。由于基础 SSA 不执行中心化，因此特征值的占比作为近似度量可能没有什么意义。
3. 对于时间序列 $x_n = A \exp(\alpha n) \cos(2\pi\omega n)$, $\omega \in (0, 0.5)$ ，如果 $L\omega$ 是整数，那么两个奇异值是重合的。如果 $\alpha < 0$ ，那么对于较大的 N ， L 和 $K = N - L + 1$ ，两个奇异值非常接近（当 $L, K \rightarrow \infty$ 时这两个奇异值渐进重合。）。在实际中，当 L 和 K 大于周期 T 的几倍时，这两个奇异值非常接近。

2.4.1.4 基本重构成分和w相关矩阵 目录

基本重构序列反应了重构的最终结果。如果我们将两个特征三元组分为一组，即第 i 个和第 j 个，那么重构的时间序列等于第 i 个和第 j 个的基本重构序列的和。

让我们使用章节 2.3.3 中的 w -相关作为基本重构分量之间可分离性的度量。当一个谐波序列产生的连个奇异向量是正交的并且相位评议近似等于 $\pi/2$ ，则对应的两个基本重构序列的相位偏移近似为 0 并且具有较强的 w -相关性。如果两个时间序列分量是强可分离的，那么他们产生的基本重构序列是 w -正交的。因此 w -相关矩阵 $\{\rho_{ij}^{(w)}\}$ 可以反映时间序列的结构。

Fig. 2.19 Series (2.21):
matrix of w-correlations



以下面的时间序列为例，

$$x_n = e^{n/400} + \sin(2\pi n/17) + 0.5 \sin(2\pi n/10) + \varepsilon_n, \quad n = 1, \dots, 3402.21) \quad (())$$

其中， ε_n 为高斯白噪声， $L = 85$ 。图2.16展示了前30个基本重构分量对应的w-相关矩阵，其中从白色到黑色，表示值从0到1变化。主特征三元组描述了指数量趋势，随后的两对特征三元组对应谐波分量，大的闪闪发光的正方形表示白噪声分量。请注意，这与（渐近）可分性理论完全一致。

2.4.2 对于给给定窗口长度的分组 [目录](#)

假设窗口长度固定为 L ，并且已经完成了轨迹矩阵的SVD分解。下一步就是对SVD的矩阵项进行分组，从而解决章节2.3.2中讨论的问题。我们假定这个问题有一个解决方案，也就是说，可以在SVD分解的结果中找到各种信号分量对应的矩阵项，并且合适的分组结果可以使得对应的信号分量近似可分。

因此我们需要确定合适的分组应该是什么样子的，并且如何构造这样的分组。换句话说，我们需要确定我们感兴趣的信号分量对应的特征三元组。由于每一个特征三元组由一个特征向量（左奇异向量），一个因子向量（右奇异向量）和一个奇异值组成，因此分组需要依靠这些向量和奇异值中蕴含的信息。

2.4.2.1 基础分析 [目录](#)

时间序列的基础分析并不是必须的，但是这有助于我们更简单地解释SSA处理的结果。可以进行如下的步骤：

1. 观察整个时间序列

- 可以观察趋势的一般形状、其复杂性以及在SVD扩展中可以预料趋势成分数量。
- 根据时间序列的形式及其性质，可以预期一些振荡及其周期。例如，对于季节性月度数据，周期12是很自然的。如果某个周期 T 是预期的，那么它的整数除数（结果应该 ≥ 2 ）很可能在SSA分解中找到。对于每月的季节性数据，它们是 12 、 $6 = 12/2$ 、 $4 = 12/3$ 、 $3 = 12/4$ 、 $2.4 = 12/5$ 和 $2 = 12/6$ 。

2. 探索时间序列的周期图

- 周期图峰值反映了可以在SSA分解中找到的预期周期。
- 峰值处的相等或接近的值表示缺乏强分离性的潜在问题。

2.4.2.2 如何分组 目录

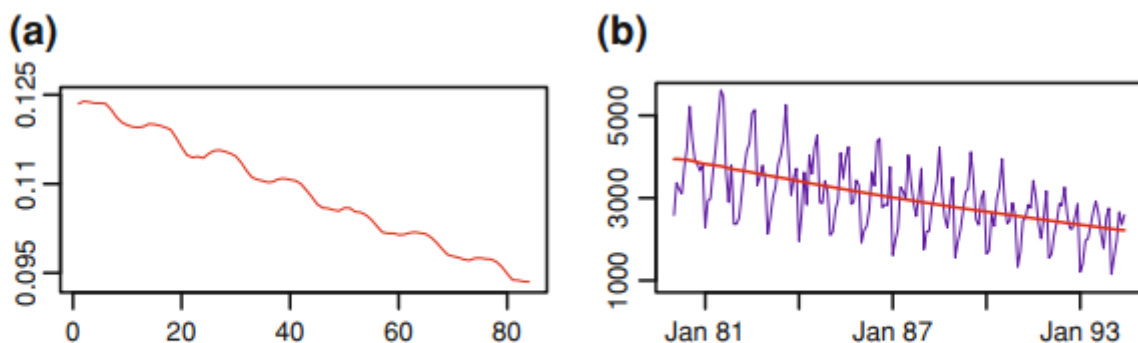


Fig. 2.20 Fortified wine: trend component. **a** Trend eigenvector. **b** Trend reconstruction and the initial series

一般建议

1. 观察特征向量、因子向量或者基本重构分量的一维图像。找到变化缓慢的成分。请注意，如果趋势分量和振荡分量没有分开，任何缓慢变化的分量都可能被振荡分量破坏。基本重构分量展示了振荡分量是否被反对角线平均抑制。最有可能的是，分量之间存在混淆是由于缺乏强可分离性造成的。窗口长度的变化和不同预处理程序的应用可以提高强可分离性。所有慢慢变化的分量都应归入趋势组。图2.20显示了趋势特征向量和趋势重构分量。
2. 考虑连续特征向量的二维图。找到正 p 多边形，可能是螺旋形的。对找到的三元组进行分组。周期为2的谐波产生1个特征三元组，因此可以在一维图中找到，例如，特征向量作为锯齿图。
3. 如果有一个基本周期 T ，那么应该努力找出周期为 T 作为除数的谐波。此外，为了重构具有给定周期 T 的整个周期分量，周期为 T/k 、 $k = 1, \dots, T/2$ 的分量应该被分为一组。
4. 如果信号分量没有被很好地分离并且以上的方法效果不好，那么可以考虑使用 w -相关矩阵。 w -相关矩阵中的一个块反映了两个相关的谐波分量；4个小块组成的一个块反映了两个混合的谐波的相关

性。由于噪声在某种意义上是许多不可分离成分的混合，因此w-相关矩阵可以帮助确定要识别的成分数量。

2.4.2.3 如何检查重构结果 目录

1. 任何统计学方法都只在某些假设成立的时候有效。他可能是信号和噪声的一个参数化模型。非参数化模型通常需要样本服从某些分布。SSA把它自己定位在一种模型自由的技术，因此结果的证明是复杂的。这样一来，最终结果得到的信号分量的可解释性就非常重要。例如当提取处的信号分量的周期为7个月通常比周期为6个月更值得令人怀疑。
2. 虽然信号可能具有非常不同的形式和结构，但噪声通常看起来像白噪声或更稀有的红噪声。如果有理由假设噪声模型，则可以定期测试相应的假设以确认结果。**在任何情况下，残差的周期图或其自相关函数都可以显示残差中是否存在信号的一部分。**
3. 为了检验该序列是红噪声的特定假设（具有正相关性的 AR(1) 模型），可以使用 Monte Carlo SSA。该测试的公开优势在于它对于替代被短时间序列信号破坏的红噪声的能力。

2.4.2.4 周期估计的方法 目录

由于周期估计在处理确定性周期信号分量的时候非常有用，因此我们列举一些可以在SSA框架下使用的估计方法。

1. 一种便捷的频率估计的方法就是周期图分析。我们可以使用这一方法来估计特征向量。因子向量还有重构信号分量的频率成分。这一方法对于长时间序列非常高效。如果时间序列比较短，那么周期图分析的分辨率也会比较低。
2. 我们可以使用一个谐波产生的两个特征向量(或者因子向量)来估计周期。如果已经计算出了特征向量，这个方法会非常地快。考虑两个特征三元组，他们今似地描述了一个频率在 $[0, 0.5]$ 的谐波分量。那么他们的特征向量的散点是一个二维曲线，并且欧几里得分量的形式为：

$$x(n) = r(n) \cos(2\pi\omega(n)n + \varphi(n)), \quad y(n) = r(n) \sin(2\pi\omega(n)n + \varphi(n))$$

其中函数 r ， ω 和 ϕ 都接近常数并且 $\omega(n) \approx \omega_0$ 。这个曲线的极坐标为 $(r(n), \delta(n))$ 并且 $\delta(n) = 2\pi\omega(n)n + \varphi(n)$ 。由于 $\Delta_n \stackrel{\text{def}}{=} \delta(n+1) - \delta(n) \approx 2\pi\omega_0$ ，我们可以通过平均极角增量 $\Delta_n (n = 1, \dots, L)$ 来估计 ω_0 。

3. 我们也可以使用基于子空间的方法包括：ESPRIT，MUSIC等。见章节3.8.这些方法有着较高的分辨率并且可以被应用于较短的时间序列如果我们准确地把信号和噪声进行分离。这些方法的一个共同点就是他们都不需要正弦分量之间相互可分离。

2.4.3 窗口长度 目录

窗口长度 L 是基础SSA算法的重要参数：该参数的不正确选择可能会导致分组难以是的SSA得到令人满意的分解结果。窗口长度的选择没有通用的规则。这里的主要困难是由于 L 的变化可能影响SSA的弱和

强可分性特征，即原始序列对应子序列的正交性和奇异值的接近性（封闭性）。然而，窗长 L 的选择有几个一般原则，这些原则有一定的理论和实践依据。让我们讨论一下这些原则。

2.4.3.1 一般原则 目录

1. 轨迹矩阵的SVD，对应的窗口长度 L 和 $K = N - L + 1$ 是等价的。因此我们总是假定 $L \leq N/2$ 。
2. 假设 $L \leq N/2$ ， L 越大，这个信号的分解就越细致。当 $L \cong N/2$ 时，分解结果是最细致的，除非这个时间序列是有限秩的，记为 d ，见章节2.3.1.2。在这种情况下，当 $d \leq L \leq N + 1 - d$ 时，SSA的分解结果都是一致的。
3. 窗口长度较小时，SSA的行为类似于宽度为 $2L - 1$ 的线性平滑滤波器。对于较小的 L ，这个滤波器产生的主特征三元组接近Barlett滤波器。该滤波器具有三角形的系数。见章节3.9.3。
4. 以下是弱可分离性的影响：
 - 由于关于时间序列分量弱可分性的结果大多是渐近的(当 $L, K \rightarrow \infty$ 时)，在大多数例子中，为了实现更好的(弱)分离，必须选择较大的窗口长度。换句话说，使用小 L 可能会导致信号分量之间的混淆，否则这些分量是可以解释的。除非两个时间序列是确定性的且完全可分的，否则当 L 固定且 $K \rightarrow \infty$ 时，重构误差不收敛于零(详见[13])。
 - 如果窗长 L 相对较大，则(弱)分离相对于 L 中的小扰动是稳定的。
 - 另一方面，对于具体的序列和任务，可以给出一些具体的窗口长度选择建议;这些建议对于相对较小的 N 可能非常有用(参见下面的章节2.4.3.3)。
5. 很难(仅通过改变 L)克服与奇异值的接近性有关的困难；即在没有强可分性的情况下，存在一个近似的弱可分性。让我们提一下与奇异值的接近性有关的两点。
 - 对于结构复杂的序列， L 值过大会导致感兴趣的信号分量被分解，进而可能导致它们与其他信号分量混合。这是一种令人不快的可能性，特别是由于 L 的显著降低可能导致分离质量差(弱)。
 - 或者，有时在这些情况下，即使 L 值的微小变化也可以减少混合，导致更好的分量分离，从而提供从弱可分性到强可分性的过渡。
6. 无论在什么情况下，使用不同的 L 值多次重复SSA分析总是一个好主意。

2.4.3.2 趋势提取和平滑的窗口长度 目录

1. 趋势

在

2.4.3.3 提取周期成分的窗口长度 目录

2.4.3.4 精致的结构 目录

2.4.4 信号提取 目录

2.4.4.1 信号提取的细节 [目录](#)

2.4.4.2 信号的秩的估计方法 [目录](#)

2.4.5 SSA成分的自动确定 [目录](#)

2.4.5.1 基于w相关矩阵的分组 [目录](#)

2.4.5.2 趋势的确定 [目录](#)

2.4.5.3 谐波的确定 [目录](#)

2.5 基础SSA算法的一些变体 [目录](#)

2.5.1 预处理 [目录](#)

2.5.2 在SSA中中心化 [目录](#)

2.5.3 平稳序列和Toeplitz SSA [目录](#)

2.5.4 可分离性旋转: SSA-ICA [目录](#)

2.5.5 SSA序列 [目录](#)

2.5.6 SSA实现 [目录](#)

2.5.7 使用其他过程替换SVD分解 [目录](#)